LISBON PHILOSOPHICAL STUDIES

uses of language in interdisciplinary fields

António Marques & João Sàáguá (eds)

ESSAYS ON VALUES AND PRACTICAL RATIONALITY – ETHICAL AND AESTHETICAL DIMENSIONS

The essays presented here are the outcome of research carried out by members of IFILNOVA (Institute for Philosophy of New University of Lisbon) in 2016.

The IFILNOVA Permanent Seminar seeks to show how values are relevant to humans (both socially and individually). This seminar is the 'place' where different research will converge towards a unified viewpoint. This includes the discussion of the following questions: What is the philosophical contribution to current affairs and decisions that depend crucially on values? Can philosophy make a difference, namely by bringing practical reason to bear on these affairs and decision? And how to do it? Which are our scientific 'allies' in this enterprise; psychology, communication sciences, even sociology and history?

This volume shows the connection between practical rationality and values and covers the dimensions ethics, aesthetics and politics.

António Marques is Full Professor at the Department of Philosophy – Faculdade Ciências Socias e Humanas, Universidade Nova de Lisboa.

João Sàágua is Full Professor at the Department of Philosophy – Faculdade Ciências Socias e Humanas, Universidade Nova de Lisboa.

Essays on Values and Practical Rationality
–Ethical and Aesthetical Dimensions

LISBON PHILOSOPHICAL STUDIES
uses of language in interdisciplinary fields

A Publication from the Institute of Philosophy of Language at the New University of Lisbon

António Marques & João Sáàgua (eds)

Essays on Values and Practical Rationality – Ethical and Aesthetical Dimensions

PETER LANG
open

Foreword
Essays on Values and Practical Rationality – Ethical and Aesthetic Dimensions

Note on this volume:

The essays presented here are the outcome of research carried out by most of the members of IFILNOVA (Institute for Philosophy at the New University of Lisbon) in 2016.

The IFILNOVA Permanent Seminar has as its theme for the next three years "Values and Human Agency" and will seek to show how values are relevant to human beings (both socially and individually). This **seminar** is the 'place' where different research will converge towards a unified viewpoint. This does not mean that we take a single philosophical position, but rather that there is an overall questioning and analysis of the following issues such as: what is the philosophical contribution to current affairs and decisions that depend crucially on values? Can philosophy make a difference by bringing practical reason to bear on these affairs and decisions? How can one do this? Which disciplines might our scientific 'allies' be in this enterprise - psychology and communication sciences, and perhaps even sociology and history?

We wish to thank all the contributors of this present volume, where the link between practical rationality and values is worked under the **three** main dimensions of ethics, aesthetics and politics.

Editors:

António Marques
(Full Professor at New University of Lisbon)
João Sáàgua
(Full Professor at New University of Lisbon)

Table of Contents

II – Arguing about values

III – Ethics and Aesthetics: Implicit
and Explicit Conections

I – Practical Rationality at Work

Understanding Value 'Relativity'

Jamie Buckland

In this paper I want to question the feasibility of the distinction between agent-neutral and agent-relative values, and explore an alternative method of understanding value in terms of Alan Thomas's distinction between non-relational and relational intrinsic value. I shall then briefly suggest that these observations can lend support to *value primacy*: the idea that talk of value cannot be reduced to talk of the reason-giving force of other (perhaps non-evaluative) properties.[1] I shall begin by explaining how Thomas Nagel drew the distinction between the agent-neutral and agent-relative in terms of both reasons and values.

1. Agent-Relativity and Agent-Neutrality: A Value-Based Theory of Reasons

The clearest statement of the distinction between the agent-neutral and agent-relative can be found in Nagel's account of *reasons* for action in *The View from Nowhere*:

> If a reason can be given a general form which does not include an essential reference to the person who has it, it is an agent-neutral reason. For example, if it is a reason for anyone to do or want something that it would reduce the amount of wretchedness in the world, then that is a *neutral reason*. If on the other hand the general form of the reason does include an essential reference to the person who has it, it is an agent-relative reason. For example, if it is a reason for anyone to do or want something that it would be in his interest, then that is a *relative reason*. (Nagel 1986: 153–154)

---

1   I want to make it clear from the outset that I have no intention of settling anything in this paper. The idea is simply to outline a position I find plausible, and to demonstrate how such a theory lends support to the primacy of value.

In a nutshell, agent-neutral reasons are said to be reasons for anyone and everyone, for instance, the *general* reasons there are for us all to donate money to charity or look after the environment. Such reasons are said to 'depend on what everyone ought to value independently of its relation to oneself' (Nagel 1991). If some act, event, or circumstance is said to have agent-neutral value, then anyone has reason to promote its occurrence, or at least desire that it happen. Agent-relative reasons, on the other hand, are said to be reasons only for particular individuals or groups; the *special* reasons there are for each of us to look after our own interests, or the interests of our family and friends, etc. If some act, event, or circumstance is said to have agent-relative value, then only some particular agent or groups of agents have reason to want and pursue it because it is related to them in the right way – it is valuable for that individual. For instance, everyone's life is considered to possess *both* agent-relative and agent-neutral value. We each have agent-relative reasons to care about our own lives (because our lives are valuable *for us*) as well as having agent-neutral reasons to care about the lives of others (because the lives of all individuals have *intrinsic* or *absolute* value). In this sense *one-and-the-same* act, event, or circumstance can often be understood as having both agent-relative value *for* someone in particular, or as having agent-neutral value simpliciter, or *absolutely*.

2. Understanding Agent-Neutral Value

Christine Korsgaard noted there is some ambiguity surrounding this notion of agent-neutral value or good-absolutely. In her examination of Nagel's distinction, she considers two approaches to understanding agent-neutral value. The first approach she calls *'objective realism'*:

> An agent-neutral value might be a value that is not relative to what agents actually value. According to this interpretation, the goodness of, say, my happiness, has what G. E. Moore called an intrinsic value, a property that is independent either of my interest in promoting it or yours. It provides a reason for both of us the way the sun provides light for both of us: because it's out there, shining down. And

> just as the sun would exist in a world devoid of creatures who see and respond to light, so values would exist in a world devoid of creatures who see and respond to reasons. (Korsgaard 1996: 278)

On this Platonic or Moorean view, one comes to value something when one perceives or discovers its value empirically. Consequently, agent-neutral values are understood as fundamental; agent-relative values are generated by, or derived from agent-neutral ones (Nagel 1970: vii; Korsgaard 1996: 278). On a second approach, agent-neutral value is understood in terms of its *intersubjectivity:*

> [A]gent-neutrality does not mean independence of agents as such, but neutral with respect to the individual identities of agents. On this reading values are inter-subjective; they exist for all rational agents, but would not exist in a world without them. (Korsgaard 1996: 278)

Given this intersubjective account, agent-relative value is fundamental; agent-neutral values are *constructed* out of agent-relative values when agents recognise and come to share each other's ends.

Before considering where Nagel's position fits in with these approaches, it is worth looking at G. E. Moore's form of objective realism in some more detail, particularly his argument against what is now understood as agent-relative value. Moore infamously rejects the idea of agent-relative value qua some *sui generis* evaluative property:

> What, then, is meant by 'my own good'? In what sense can a thing be good *for me?* It is obvious, if we reflect, that the only thing which can belong to me, which can be *mine*, is something which is good, and not the fact that it is good. When, therefore, I talk of anything I get as 'my own good,' I must mean either that the thing I get is good, or that my possessing it is good. In both cases it is only the thing or the possession of it which is *mine*, and not *the goodness* of that thing or that possession. […] In short, when I talk of a thing as 'my own good' all that I can mean is that something which will be exclusively mine, as my own pleasure is mine (whatever be the various senses of this relation denoted by 'possession'), is also *good absolutely*; or rather that my possession of it is *good absolutely*. (Moore 1903: 99)

Now, some take this position to amount to the claim that talk of what is good-for or valuable-for some particular agent or another is, itself,

nonsensical (Kraut 2007: 70). However, the more plausible interpretation of Moore's argument is that the idea of some kind of *sui generis* agent-relative evaluative property is nonsensical, for goodness, itself, is not a relational property. Consequently, talk of agent-relative value, or 'good-for *x*' can only ever be talk of good-absolutely which stands in a certain relation to a particular agent. Goodness or value is a metaphysically simple, non-natural property which belongs to things which are good.

Moore's view aside, Korsgaard's favoured position is intersubjectivism. Nevertheless, she insists there is some ambiguity as to whether Nagel shares this position or is, rather, committed to some kind of objective realism (Korsgaard 1996: 279–282). However, when understood correctly, it is clear that Nagel's position is more akin to the intersubjectivist position, whereby values are understood as objective in the sense that they are *normatively real*.

As we saw in section 1, Nagel advances a value-based theory of reasons where the presence of a reason for action is explained by the relation in which an action stands to some valuable state of affairs, yet for Nagel the idea of some kind of Platonic or Moorean objective realism is misguided. For Nagel, evaluative judgements depend on judgements about the presence of particular reason-constitutive considerations, and the objectivity and reality of value is fundamentally grounded in the conception of a practical agent. Thus, values arise from and are analysed in terms of the conception of a good reason for action:

> *Values are judgments* from a standpoint external to ourselves about how to be and how to live. Because they are accepted from an impersonal standpoint, they apply not only to the point of view of the particular person I happen to be, but generally. They tell me how I should live because they tell me how anyone should live. (Nagel 1986: 135)

What appears odd, however, is that Nagel's objection to reductive accounts of normativity more generally is stated explicitly in terms of values rather than reasons:

> If values are objective, they must be so in their own right, and not through reducibility to some other kind of objective fact. They have to be objective *values*, not objective anything else. (Nagel 1986: 139)

Nevertheless, for Nagel this does not amount to a reduction of values to reasons; the relationship he envisages between reasons and value is one of direct correspondence or equivalence; there is an asymmetrical dependence relation between value and the grounding of reasons in the sense that valuable acts, events, or states of affairs are said to provide agents with a reason to promote them. It is in this sense that values are said to be *normatively objective*, for to objectify them under a non-normative criterion would be to reduce them to some other objective fact, for instance, a psychological fact such as a desire which is said to explain the presence of a reason.[2] Consequently:

> Normative realism is the view that propositions can be true or false independently of how things *appear to us*, and that we can hope to discover the truth by transcending the appearances and subjecting them to critical assessment. What we aim to discover is not an aspect of the external world, called value, but rather just the truth about what we and others should *do* and *want*. (Nagel 1986: 139 [emphasis added])

Agents, then, do not discover reasons for action that exist independently of their pre-existing subjective motivational states and interests, but rather, taking up an objective standpoint allows objectivity to bear on their will, which can alter and constrain those motives that are already present. These considerations are, as such, the seeds of our moral theorizing. Objectivity and realism in ethics, then, is not analogous to the sense found in theoretical reasoning or empirical Moorean realism. Nagel offers a practical account of objectivity and a normative realism disanalogous to theoretical reasoning in the sense that we do not arrive at new beliefs that include ourselves as components; rather we arrive at an extended set of values and normative judgements from a centreless and impersonal standpoint of objectivity.[3]

---

2 Though one may agree that values are not reducible to the non-normative without thinking that values = reasons.

3 Importantly, the distinction between agent-neutral and agent-relative *value* does not map onto the intrinsic/extrinsic distinction. Nagel stresses that there is a separate metaphysical question concerning the manner in which reasons vary in their externality or independence from human concerns. He acknowledges the fact that most of the apparent reasons visible from our personal standpoint are intimately connected with our first-personal interests and desires, yet it seems that many of

3. Three Kinds of Agent-Relativity?

Nagel's aim in *The Possibility of Altruism* was to show that all personal agent-relative reasons had to have an agent-neutral counterpart[4] (Nagel 1970: vii). Nevertheless, he eventually succumbed to the idea that there are agent-relative reasons (and values) that are too idiosyncratic to be subsumed under a suitable agent-neutral counterpart, so he allowed for the existence of certain agent-relative reasons (and values) so long as they were tolerable from an agent-neutral perspective. These agent-relative reasons were said to fall into three categories. Firstly: *reasons of special obligation*. These are reasons stemming from the value of the personal/familial relationships we have with others, for instance, the agent-relative reasons I have to look after my own family and friends. Secondly: *reasons of autonomy*. These reasons are grounded in the value of an agent's personal projects or goals, for instance, my ambition to climb Kilimanjaro. And, thirdly: *reasons of deontology*. These are reasons stemming from an agent's special concern with his or her own actions, for instance, reasons not to kill innocent people, tell lies, etc.

Now, although orthodox, the idea that deontological reasons are agent-relative is somewhat controversial (cf. McNaughton & Rawling 1991; Portmore 2013). Remember, the idea is that agent-relative reasons track agent-relative values, yet it is far from clear how the disvalue of, say, killing an innocent person is 'agent-relative' in the required

---

the objects of such interests and desires have an intrinsic value or a goodness in themselves, independently of how it is valued, or the satisfaction an agent may derive from valuing such objects, i.e. that some values are not reducible to their value *for* anyone. Admittedly, Nagel is unsure how to establish the existence of such values, but seems to regard the aesthetic value of the Frick Collection as a prime example of an object of external interest. Nevertheless, contra to Moore's beautiful world, which retains its value without sentient beings, Nagel wants to avoid the 'implausible consequence' that the Frick Collection somehow retains its 'practical importance', even if humanity were destroyed (Nagel 1986: 153).

4   Incidentally, Nagel regarded the argument of *The Possibility of Altruism* as intimately related to Moore's infamous argument against the egoist's notion of agent-relative value or personal goodness although, on Nagel's understanding, ethical egoism is a theory primarily concerned with reasons for action rather than the good (Nagel 1970: 86).

sense. Granted, there is a sense in which not killing an innocent person is good-for-me, but it is simply odd to say that this is what makes the killing of innocents bad. Nevertheless, it certainly seems fitting that the value of one's personal/familial relationships, personal projects, or goals can fittingly be described as 'agent-relative', i.e. *'good-for'* particular individuals. Though, again, as Korsgaard has noted, this seems to mischaracterise the phenomena.

4. Korsgaard on Ambitions and Special Obligations

Korsgaard refers to reasons of autonomy as *'ambitions'*. For Nagel, the agent-relative value of one's ambitions provide agents with agent-relative reasons to do things, yet the normative force of these reasons does not extend beyond the agent for whom they are ambitions: no one else has a reason to, say, help me climb Kilimanjaro, and I have no reason to help you achieve your ambitions (Nagel 1986). Korsgaard rightly insists this is mistaken:

> Suppose it is my ambition to write a book about Kant's ethics that will be required reading in all ethics classes … Following Nagel's analysis, we will say that this ambition is agent-relative … But this way of describing the situation implies a strange description of my own attitude. It suggests that my desire to have my book required is a product of raw vanity, and that if I want to write a good book, this is merely a means to getting it required [in all ethics classes.] … So the structure of this ambition is not:
>
> (1) I want my book to be required reading (where that is an agent- relative end);
>
> (2) therefore: I shall write a good book (as a means to that end); but rather:
>
> (1) Someone should write a book on Kant good enough that it will be required reading (where that is an agent-neutral end);
>
> (2) I want to be that someone (agent-relative motive). (Korsgaard 1993: 287–288)

The same observation extends to reasons of special obligation:

> The structure of reasons arising from love is similar to that of reasons of ambition. I think that someone should make my darling happy, and I very much want to be that someone. And others may have good reasons to encourage me in this. But if I try to prevent someone else from making my darling happy, or if I suppose that my darling's happiness has no value unless produced by me, that is no longer an expression of love. Again, it is a very familiar perversion of it. (Korsgaard 1991: 211)

The structure of ambitions and obligations is more complex than Nagel allows, then. Rather than a self-interested desire for an object, or a desire to realize something you think is good-for-you, an ambition is the desire to stand in a special relationship to what is good, agent-neutrally (intersubjectively). It may well be some type of 'agent-relative' element that has the motivating force, i.e. it is *my* ambition to climb Kilimanjaro, and *my* daughter whose care I must prioritize, but this desire is not the *source* of my reason. It is here that Korsgaard's *intersubjectivism* becomes apparent. In offering you my reasons I am offering you the 'familiar voice of humanity, not the voice of alien idiosyncrasies' (Korsgaard 1996: 290). Subsequently, the value of our ambitions and personal obligations is not, as the objective realist would have it, *intrinsic*, but rather an expression of the interest in other agents or humanity. Qua intersubjectivist, an agent first understands himself as simply an agent among many, and then attempts to understand and share the ends of others. The objective realist works in the opposite direction. He must first see if he can share another's ends, and then decide what relationship he wants to have with others. Korsgaard regards this as a mistake: 'We should promote the ends of others not because we recognize the value of those ends, but rather out of respect for the humanity of those who have them' (Korsgaard 1996: 279).

5. A False Dichotomy?

With regard to the nature of value, for Korsgaard there are only two possibilities: either all value is accounted for in terms of some mind-independent Moorean objective realism, placing it in no relation to the

subjective interests of agents whatsoever (objective realism), or it is explained in terms of intersubjectivism, where the interests of rational agents enter into the very *analysis* or *content* of value itself in the sense of rendering all values *relational*, i.e. the value of an object of moral concern obtains 'in its relation to the subject' (Korsgaard 1996: 279). However, this leaves Korsgaard committed to the implausible idea that all values are *extrinsic* values, as their existence depends solely on a process of construction by rational agents.

Following Alan Thomas, we can refer to this importation of relationality into the very content of value as *relationality in value* (Thomas 2006: 48). It is this peculiar notion which seems to fuel the idea of agent-relative value: that an agent's apprehension of value somehow constitutes it, i.e. the idea that the value of my ambition to climb Kilimanjaro is essentially *mine* is what constitutes it *being* valuable. This idea seems confused. Indeed, it is more than plausible that you (or anyone) can grasp the value of *my* ambition to climb Kilimanjaro in the same manner as you can grasp the value of your ambition to, say, play the piano. For Thomas, the very notion of agent-relative value points towards a more interesting distinction between *relational* and *non-relational* intrinsic values.

A value is both intrinsic and relational when it stands in an asymmetrical constitutive relation to a subject. In this sense, the value of my ambition to climb Kilimanjaro can only be accounted for by making reference to whom the project is valuable. However, unlike Korsgaard's intersubjectivism, these relational values can be *both* relational and intrinsic:

> [Relational intrinsic] values do not derive their value from the relations to any object outside their nature (extrinsically); it is, rather, that it is in their nature to be values or disvalues in so far as they stand in a certain constitutive relation. (Thomas, unpublished)

The key point to note is that this kind of relationality is not '*agent-relativity*' in the Nagelian sense. Of course, the value of my climbing Kilimanjaro is essentially relational, but this is not explained by the fact that it is valuable for *me, Jamie Buckland*. Analogously, take the disvalue of pain. Of course the pain I feel when I stub my toe on the coffee table is bad in so far as it is *my* pain, qua my mental state, but this

is not what constitutes its disvalue; 'mineness' cannot function as the metaphysical grounding of the value of pain qua mental state. The relationality concerns the metaphysical *relationality of value* (the value is constitutively the value *of* a subject), as opposed to Korsgaard's peculiar phenomena of relationality *in* value.

At this point it might be objected that Thomas's notion of relational intrinsic value is simply a truism. But this is avoided by the fact that there are also non-relationally valuable states of affairs whose value obtains simply by virtue of their intrinsic properties, i.e. the value of these non-relationally intrinsically valuable states of affairs can be explained without making reference to the relational states of a subject. For instance, Thomas is happy to identify the notion of a non-relational intrinsic value with Moore's idea that a beautiful world obtains without sentient subjects. However, unlike Moore's view, Thomas's idea is merely a micro-level claim because, as a whole, value itself is related to human interest via a doctrine of *presupposition*. Again though, this raises the concern that non-relational intrinsic values simply collapse back into relational intrinsic value or even extrinsic value.

The issue is complex, but a response ties into the wider sensibility theorist's position that values are anthropocentric, but, nonetheless, *real*, so the relation between value and human interest is one of *presupposition*: 'there is a sense in which value as a whole stands in relation to human interest. But this relation does not enter into the analysis of value itself' (Thomas 2006: 48). The relationality of inherently relational (but intrinsically valuable) values, then, must be distinguished from the deeper, presupposed perspectival relationality found within our evaluative judgements:

> Our relationship to value exhibits different kinds of relativity. One obvious case is that evaluative judgements deploy concepts that exhibit perspectivalness. Our peculiarities as a class of judgers enter into the possession conditions for the concepts that are deployed in our evaluative judgements. The metaphysical idea is that certain classes of judgeable contents are only available to judgers of a certain kind. But it is important to realise that in such an account this deep relativity to our metaphysical point of view does not enter in to the very content of the judgements we make, making them explicitly indexical in their very content. The relativity is rather presupposed, so that we can, from a standpoint of engagement with our conceptual scheme, make judgments that are, from our perspective, plainly true. (Thomas 2006: 47–48)

The idea that our relation to value is presupposed sheds further light on the distinction between metaphysical debates about the relationality *of* value, and the separate phenomenon of the relationality *in* value that besets Korsgaard. The problem with Korsgaard's analysis is that the relationality of value has become part of the very nature of value itself – it has become relationality *in* value. In keeping this distinction separate the door is opened for those who wish to defend a metaphysical account of certain intrinsic values as constitutively relational without them being either relative or extrinsic or even instrumental.

6. Value Primacy

How, then, does the foregoing lend support to the primacy of value; the idea that talk of value cannot be reduced to talk of the reason-giving force of non-evaluative properties?

As we saw above, for both Nagel and Korsgaard the idea of Moorean objective realism is misguided: evaluative judgements depend on judgements about the presence of particular reason-constitutive considerations. In this sense, the objectivity and reality of value is grounded fundamentally in the conception of a practical agent; values arise from and are analysed in terms of the conception of good reasons for action. This idea can be understood as a precursor to what has now become known as the 'buck-passing' theory of value, namely because Nagel's normative realism 'passes the buck' from evaluative talk to the grounding of practical reasons for action, i.e. goodness is not a reason-grounding property itself, but is reducible to the reason-giving force of other properties, to quote Scanlon:

> Goodness is not a single substantive property which gives us reason to promote or prefer the things that have it. Rather, to call something good is to claim that it has other properties (different ones in different cases) which provide such reasons. (Scanlon 1998: 11)

However, the sensibility theorist's account of value offered above explicitly rules out passing the buck because evaluative attitudes cannot

be characterized without '*ineliminable* reference to the values that these attitudes bring into view' (Brewer 2006: 157 [emphasis added]). In following Thomas by understanding the metaphysical status of value as both anthropocentric, yet real, *value* is cited '*irreducibly* and *ineliminably* in the best explanation of our formulation of moral beliefs' (Thomas 2006: 48 [emphasis added]; Wiggins 2002: Ch. IV, V). This is, essentially, a phenomenological appreciation, and irreducible to the recognition of reasons for action.

References

Brewer, T. (2006). *The Retrieval of Ethics*. Oxford: Oxford University Press.

Korsgaard, C. (1996). *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

Kraut, R. (2007). *What is Good and Why? The Ethics of Well Being.* Cambridge (MA): Harvard University Press.

McNaughton, D., & P. Rawling (1991). 'Agent-Relativity and the Doing-Happening Distinction'. *Philosophical Studies*, Vol. 63, No. 2: 167–185.

Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.

Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Oxford University Press.

(1986) *The View From Nowhere*. Oxford: Oxford University Press.

____(1991) *Equality and Partiality.* Oxford: Oxford University Press.

Portmore, D. (2013). 'Agent-Relative vs. Agent-Neutral', in LaFollette, H. (ed.), *International Encyclopedia of Ethics*. Oxford: Blackwell.

Scanlon, T. M. (1998). *What We Owe to Each Other.* Cambridge (MA): Harvard University Press.

Thomas, A. (2006). *Value and Context*. Oxford: Clarendon Press.

Wiggins, D. (2002). *Needs, Values, Truth* (3rd edition, amended). Oxford: Clarendon Press.

Intersubjectivity and Freedom in Merleau-Ponty's
Phenomenology of Perception

Luís Sousa

In this article, I intend to show that Merleau-Ponty's conception of freedom is essentially tied to his view of the human subject as socially situated: that is, that it presupposes his views on the intersubjective dimension of subjectivity. In other words, I will show that in his view sociality is not an impediment but rather a condition of freedom. In this regard, his position contrasts starkly not only with modern philosophy's prevailing conceptions of autonomy (perhaps with the exception of Hegel) but also, and in particular, with Sartre's view on freedom. Very roughly, Sartre equates being a subject with freedom and this, in turn, with self-determination in the sense of independence from others. Freedom is thus an essentially individual feature and other subjects are viewed as hindrances to it. For Merleau-Ponty, freedom is also tied to subjectivity – to the fact that we are in some sense subjects. It is only insofar as we are subjects that we are free, but our subjectivity is not opposed to our social nature for true subjectivity presupposes sociality.

For reasons of space, this investigation will focus on the *Phenomenology of Perception* although, as with many other topics in Merleau-Ponty's thought, the same ideas are developed in other works.[1] For this same reason, I will put to the side the problem of how far Merleau-Ponty develops his thought beyond this early articulation in subsequent works.

In the first section, I will briefly touch on the relevance of the problem of intersubjectivity not only to the way in which Merleau-Ponty conceives of the specificity of phenomenology but also to the problem

---

1   See, for example, the essay 'Cézanne's Doubt' from *Sense and Non-Sense,* which, despite what its title suggests, has freedom as its primary topic. On intersubjectivity, see for example the lecture 'The Child's Relation with Others' from *Child Psychology and Pedagogy: The Sorbonne Lectures 1949–1952*.

of subjectivity. In the second section, I will provide a close reading of Merleau-Ponty's phenomenological analysis of intersubjectivity. I will show that, contrary to what is sometimes claimed, the notion of body-subject does not solve the problem of other minds on its own; moreover, Merleau-Ponty does not claim this. He thinks there is a grain of truth in solipsism even if that position misunderstands and misstates the issue. This grain of truth corresponds to the privileging of the first-person perspective over the perspective of others. In the fourth section, I will show that, despite this privilege, subjectivity is not only socially situated but presupposes intersubjectivity. In the fifth section, I will introduce Merleau-Ponty's discussion of freedom. Much of his account of freedom is built on his critique of Sartre. I will show that, although Merleau-Ponty does not deny freedom, unlike Sartre he views it as crucially tied to the agent's factical and social situation. In the sixth section, I will show that Merleau-Ponty's account of class consciousness is a good example of how action can be socially propelled and at the same time free. Finally, in the seventh and final section, I will conclude by providing an overview of how Merleau-Ponty brings his accounts of subjectivity and intersubjectivity to bear on the topic of freedom.

1.

The *Phenomenology of Perception*, inasmuch as it is a *phenomenology*, is an attempt to present a conception of subjectivity that is opposed to all empiricist conceptions in a broad sense – that is, opposed to any attempt to explain away subjectivity in favour of third-person processes. At the same time (and this arguably constitutes the novelty of Merleau-Ponty's work in comparison to his predecessors), the *Phenomenology of Perception* is a radical critique of the Cartesian, and even the Kantian, view of subjectivity.

That Merleau-Ponty contrasts his position on subjectivity with the modern tradition's preferred account is particularly relevant here. One of Merleau-Ponty's main goals, indeed perhaps his central goal, is to show that the phenomenological point of view should be carefully

distinguished from the modern point of view with which it is often conflated. Merleau-Ponty views intellectualism and idealism as the main traits of modern philosophy. By contrast, he sees the 'phenomenological point of view', which he associates with his reading of Husserl's later philosophy and Heidegger's existential phenomenology, as overcoming the modern intellectualist-idealist trend in philosophy.

Merleau-Ponty understands the phenomenological point of view as an overcoming of the intellectualist-idealist point of view in part because of the status each grants to subjects other than oneself. In his view, Descartes' and Kant's theories of subjectivity, if brought to their ultimate conclusions, leave no place for 'other subjects'. If the Cartesian conception is true, if my nature is that of a thinking substance from which everything bodily and material is excluded, my only distinguishing feature is the bare 'I think'. But this 'I think', the cogito, cannot serve to individuate me, and thus on this assumption there is no point in talking about other souls beyond my own (except maybe God's). Kant, for his part, thinks of the subject as the ultimate centre of thought and conceptual activity. In his view, we cognize the subject not as it is in itself, as a noumenon, but only as it appears in time, as a phenomenon – that is, as an 'empirical subject'. Since the transcendental subject is the ultimate condition of all cognition and cannot itself be cognized, all other possible subjects are only objects for it. For this reason, Kant conceives of the plurality of subjects as a plurality of *empirical* subjects. The transcendental subject as such cannot be identified with someone in particular and contradicts the possibility of another transcendental subject, for there is, *ex hypothesi*, only one ultimate condition of thought, and this lies within me as the one and only transcendental subject. Thus, although Kant does not put it precisely this way, the transcendental subject, the subjective correlate of nature, is essentially one, transcending the plurality of empirical subjects that appear within nature.

In a way that is related to their views on the nature of subjectivity, empiricism and intellectualism present seemingly opposed ontological accounts of the world. Whereas empiricism tends to be a materialist theory of reality, intellectualism corresponds to an idealist point of view inasmuch as it traces reality back to some kind of subjective construct. For Merleau-Ponty, however, both approaches share the assumption that reality is ultimately intelligible, and the project of absolute knowledge is at least

ideally conceivable (even if it turns out, as with Kant, to be impossible for us humans). For this reason, the phenomenological point of view – which includes phenomenology's novel account of subjectivity – will bring with it a completely new way of looking at the world, a new ontology.[2]

As I said, Merleau-Ponty understands phenomenology, and in particular the work of Edmund Husserl, as presenting or at least strongly suggesting a view of subjectivity that is opposed to that favoured by the intellectualist-idealist tradition. It is true that Husserl's conception of subjectivity seems to have affinities with both Descartes' and Kant's conceptions. And it is perhaps no mistake that Husserl calls his own brand of idealism 'transcendental idealism' (although he tries to distinguish it from Kant's version). This notwithstanding, Merleau-Ponty sees implied in Husserl's work, especially in its final phase, a conception of the subject as temporal, embodied, finite and historical – a view that contrasts sharply with the intellectualist, idealist, Cartesian conception.[3]

2.

Merleau-Ponty's main discussion of intersubjectivity in the *Phenomenology of Perception* can be found in the chapter 'Others and the Human World'.[4] From the outset, when we take a look at the context

---

2   For this reason, I believe that Merleau-Ponty's project was, from its very beginnings, concerned with ontology. Ontological claims appear not only in his last, unfinished, work (*The Visible and the Invisible*) but also in the *Phenomenology of Perception*, however much they may have developed over time.

3   As is now widely known, much of Husserl's most relevant work was published after Merleau-Ponty's death. Although Merleau-Ponty was one of the first visitors to the Husserl Archives at Louvain, the material which is now at our disposal gives us a much more complete picture not only of Husserl's thought but also of its development throughout the years. However, I will not be concerned here with the accuracy of Merleau-Ponty's interpretation of Husserl. For our purposes, what is important is that Merleau-Ponty understands his own position as laying out what was already implied in what he viewed as Husserl's position.

4   Discussions on this chapter and the topic of intersubjectivity in Merleau-Ponty can be found in Barbaras (2004: 19ff., 239ff.); Carman (2008: 134ff.); Dillon (1988:

in which the chapter appears, we see that it is preceded by 'The Thing and the Natural World'. The sequence of chapters and Merleau-Ponty's introductory remarks in the chapter on intersubjectivity suggest that he may entertain the idea that there is a 'natural world' independent of the so-called 'cultural world'. However, although Merleau-Ponty deals with the natural world before and independently of the cultural world, it is likely that, for the most part, this is done only for the sake of presentation. For Merleau-Ponty, although there is a natural world, the world in which we find ourselves from the day we are born is already cultural, social and historical through and through, and it is only by abstraction that we can think of it as purely natural.[5] (It should be noted that I add the predicates 'cultural' and 'historical' to 'social' here because the former depend on sociality.)

The view that the natural world is already given to us as a social world is not even a particularly original position. Heidegger had already contended that the world is essentially a social world – that is, that others are always already there with me in the world, such that we never find ourselves in a private, solipsistic world.[6] Our world is already filled with others, even when they are not factically present.[7] They are there, for example, whenever we perceive a tool as a tool, or any kind of artefact as an artefact. (The reason for this is not only that we did not invent or make the tool ourselves but also that the tool was fashioned with reference to a normal-sized human body.) It is arguable that most of the ideas put forward by Heidegger in this regard can already be found, albeit in a different language, in Husserl. For example, we find in Husserl the idea that even natural things like trees are not given to us as private phenomena. Rather, we perceive them as public phenomena, as things that are in principle perceptible to others as well.[8] Merleau-Ponty, too, believes that we find the other embedded in things of culture, even when

---

113ff.); Hass (2008: 100ff.); Langer (1989: 97ff.); Matthews (2002: 89ff.); Priest (1998: 179ff.); Romdenh-Romluc (2011: 130ff.).

5   On the idea that for Merleau-Ponty the natural world is not independent of the intersubjective, cultural world, see Langer (1989: 97ff.).

6   See Heidegger (1962: §§25–7).

7   See Heidegger (1962: §26).

8   On this idea, see Zahavi (2003: 110).

factically there is no one beside us.[9] Whenever I find traces of human action, even from the distant past, I encounter the other.

These introductory remarks are enough to show that the phenomenological approach to intersubjectivity significantly transforms the philosophical problem of others as it is usually understood. From a phenomenological point of view, there is no sense in asking whether the world is my world only or also a world of others, a social world. The world is already given to me as a social world; it is therefore pointless to ask whether there are in fact others besides me, or whether others perceive and act in the world in which I find myself. Others are factically given to me along with my world, such that only by abstraction can we think of a solipsistic world. This is not to say that there is no phenomenological problem concerning others, for there is in fact such a problem. Since others are already given to me along with the world, phenomenology tries to make explicit the meaning of the other: what it means for me to encounter another I, and what the different forms of community and sociality mean.

According to Merleau-Ponty, although the other is given to me along with the world in the form of cultural artefacts, it is given through a 'veil of anonymity' (PhP 363 [405]).[10] In order to characterize the anonymous way in which the other is given to me through cultural objects, Merleau-Ponty relies on the French pronoun '*on*' (one). This is an obvious nod to Heidegger's notion of *das Man*,[11] which the latter describes as referring to everyone and no one in particular, to the anonymous other who is along with me in the world. For Merleau-Ponty, the anonymous givenness of the other presupposes the broader problem of knowing how the other, who is supposed to be an 'I' like myself, can be externalized in the form of cultural objects and behaviour. Merleau-Ponty even uses the Hegelian expression 'objective spirit' (PhP 363 [405]) to refer to the way human behaviour appears in the form of things.

---

9   See Romdenh-Romluc (2011: 131).
10   PhP refers to the *Phenomenology of Perception*. Page numbers are from the most recent English translation by Donald Landes. The original pagination, which is also reproduced in Landes' translation, appears in square brackets. Quotations are from Landes' translation, except where indicated.
11   See *Being and Time*, §27.

Since cultural things can always be traced back, in one way or another, to human behaviour, we are ultimately referred back to the question of how the other can be manifested through her body. This is what Merleau-Ponty means when he says that the body is the 'first cultural object' (PhP 406), not because it should be compared to a mere thing, even if a cultural one, but instead because it is the primary locus of behaviour, 'the place' where human behaviour shows itself. Thus the question becomes: how can a person, someone endowed with intelligence and freedom, manifest herself through her body?

Merleau-Ponty tries to deconstruct this problem by appealing to his theory of the body as laid out in the first part of the *Phenomenology of Perception*. In brief, Merleau-Ponty contends that the objective body – i.e. not only the body as seen from the third-person point of view, but in particular the body as a bundle of mechanisms and processes – is an abstraction. The body as it is immediately given to us, the body as we live it, the lived body – and also, as we will see, the body of the other as it is concretely given to us – is not an object among objects; it is not a mere thing that happens to be controlled by the mind. What phenomenological analysis of the way I live my body (for example by moving it in space) reveals is that my body has a special kind of intentionality; it partakes of the sense-giving activity usually ascribed to the mind, expressing meaning of itself not only through movement but also through perception and language.

The conception of the body as a lived body significantly transforms the problem of how I can access another through his body. For the objective view is, in this regard, exactly symmetrical to the intellectualist-idealist view mentioned above. Both the materialist and the idealist conceive of the body as a bundle of third-person processes. When we hold this view of the body, perception of the other becomes extremely problematic. Since the objective body displays no meaning, no intelligence – in sum, no subjectivity – we are allowed to question whether a mind lies behind the body of the other, that is, to question whether the other is in fact an *other* or just a meaningless body that deceives me into thinking it is a subject, an ego, just like myself. On the other hand, according to Merleau-Ponty's conception of the body as a lived, subjective body, there is no problem concerning the existence of selves other than mine for the body expresses meaning of itself, and it does

not make sense to ask about the existence of a mind or soul behind the body's meaningful behaviour. In Merleau-Ponty's approach, I have access to the other quite directly: I see him behave in meaningful ways, I read emotions off his face, I hear him convey thoughts and feelings through language, etc.:

> If I do not learn within myself to recognize the junction of the for-itself and the in-itself, then none of these mechanisms that we call "other bodies" will ever come to life; if I have no outside, then others have no inside. (PhP 391 [431])

One of the traditional philosophical arguments for explaining how I come to know about the existence of other minds (and thus other selves) beyond my own is the analogy argument. According to this argument, I implicitly ascribe mind and subjectivity to others in order to account for behaviour that I perceive in myself as a manifestation of mental states. Merleau-Ponty rejects this type of argument outright. First, he claims that it is empirically false. Children, for example, recognize others and their behaviour in a quite immediate way well before they are able to engage in this kind of reasoning. They perceive the body of the other and its intentions as being internally related to their bodies' intentions. (Merleau-Ponty observes that when a person pretends to bite a baby's hand, the baby immediately opens her mouth, recognizing without express comparison the relation between the adult's mouth as an 'organ for biting' and her own.)[12] In fact, contemporary neuroscience has even reinforced and confirmed Merleau-Ponty's insight into this matter. Since the discovery of mirror neurons, we know that our brains come equipped with inborn capacities to recognize and empathize with others.[13]

However, Merleau-Ponty also has a priori grounds for rejecting the analogy argument. Indeed, this rejection can be viewed as a corollary of his broader theory of subjectivity, mind, body and intentionality. Suffice it to say that all reasoning by analogy presupposes an explicit comparison between myself and the other, but for this comparison to take place

---

12   See PhP 368 [409].
13   On the possibility that Merleau-Ponty foreshadowed the discovery of mirror neurons, see Carman (2008: 138).

the other must somehow be previously given to me as an *other*, for otherwise I have no basis for drawing an analogy between us. In sum, the analogy argument already presupposes what it is trying to explain.[14]

The upshot of all this is that there is no absolute distance between myself and the other. Merleau-Ponty's main point in this respect is that if I were absolutely transparent to myself – that is, if I were nothing but a purely transcendental subject or a Cartesian cogito – the possibility of another subject like myself would remain quite unintelligible to me. However, before I come to regard myself as a cogito, before I am able to reflect and say to myself 'I think', I am already thrown into the world by way of my body. For Merleau-Ponty, I can never analyse and thoroughly rationalize this primitive relation to the world. Merleau-Ponty even calls this an 'ancient pact' (PhP 265 [302]) between my body and the world. It is this primitive relation to being, or to the world by virtue of being a body, of being *my* body, that Merleau-Ponty sometimes calls (making his own use of Heidegger's and, of course, Sartre's terminology) 'existence', 'transcendence' or 'being-in-the-world'.

Insofar as I am a perceptive body, I always find myself already thrown into the world. This bodily, perceptive, pre-personal self is in some sense other than myself considered as a personal subject capable of saying 'I'. According to Merleau-Ponty, there is no true correspondence between the reflective self, the self as subject in the Cartesian and Kantian sense, and the self as a pre-reflective body-subject in the world. Paradoxically, it is the lack of correspondence between these

---

14 Taylor Carman (2008: 138) claims that Merleau-Ponty's critique of the analogy argument can be read as directed at Husserl's account of intersubjectivity, at least as it is presented in the *Cartesian Meditations*. But Husserl's position is not as cut and dried as Carman would want us to believe. Husserl rejects the idea that what he calls the 'apperception' of the other is achieved by any kind of inference, and in my view this is exactly what Merleau-Ponty is trying to convey with his critique of the analogy argument. It is true that Husserl holds that the greatest difficulty concerning apperception of the other is that the other is never given to me as an ego; were he so given to me, he would not appear as an *other* – that is, as different from me. But as we will see, Merleau-Ponty agrees with this Husserlian position as well. For an interpretation of Husserl's analysis of intersubjectivity that presents him as much closer to Merleau-Ponty, see Zahavi (2003: 112ff.). On Husserl's influence on Merleau-Ponty's account of intersubjectivity, see also Dillon (1988: 114ff.).

selves that makes the other possible for me. If I am in some sense an other to myself, then the possibility of other subjects, and namely of other body-subjects, becomes intelligible. It is also for this reason that Merleau-Ponty speaks of my access to the other as a 'sort of reflection' (PhP 367 [409])[15] (an analogue to the relation between the reflecting ego and the pre-personal ego):

> Others can be evident because I am not transparent for myself, and because my subjectivity draws its body along behind itself. (PhP 368 [410])

Another way to put this is the following. If the world as a whole does not lie transparently before me, if I am nothing but a certain point of view on the world, then my perspective is not wholly discontinuous with other perspectives. Just as the different parts of my body display a pre-reflective unity among themselves (which Merleau-Ponty calls the *body schema*), a unity which is prior to them and determines their value and sense, the possibility of the point of view of the other is prefigured in my own. I form a sort of systematic unity with others that precedes all thought and rationality – a unity that ultimately corresponds to the world itself. In other words, my point of view is related to other possible points of view just as my different successive perspectives are related to each other. The possibility of a foreign perspective is already suggested by the fact that there is always more of the world for me to see:

> In fact, the other person is not enclosed in my perspective on the world because this perspective itself has no definite limits, because it spontaneously slips into the other's perspective, and because they are gathered together in a single world in which we all participate as anonymous subjects of perception. (PhP 369 [410–11])

There are issues, however, concerning how to interpret this unity. If it is such that I cannot distinguish between myself and others, it cannot be used to account for intersubjectivity; *ex hypothesi*, there would be no distance between us in such a unity, and therefore no intersubjectivity in the strict sense. Perhaps this is why Merleau-Ponty will say (see the next

---

15   That my original relation to the other is made possible and prefigured by the structure of my self-consciousness was already shown by Husserl. On this, see Zahavi (2003: 113).

section below) that his account of the lived body is not sufficient to account for intersubjectivity and to preclude every form of solipsism. Moreover, it can be argued (based in particular on the lecture 'The Child's Relation with Others' from *Child Psychology and Pedagogy: The Sorbonne Lectures 1949–1952*) that Merleau-Ponty holds that my primary relation to others is one of identity, and that only later do I develop a sense of being an individual self in opposition to them. This is also the core of Dillon's interpretation of Merleau-Ponty's early stance on intersubjectivity (1988: 120ff.). Nevertheless, I have doubts about whether this position can in fact be ascribed to Merleau-Ponty based only on what he says in the *Phenomenology of Perception*. In this text, it seems that, although I form a systematic unity with others in the sense that they complement my perspective on the world, there are also differences between us, such that I pre-reflectively know that my perspective differs from theirs. I will expand further on this pre-reflective sense of self in the next section.[16]

3.

As surprising as it may be, what we have said thus far is far from the whole story when it comes to Merleau-Ponty's take on the problem of intersubjectivity. First, just as it is my body that, through its 'ancient pact with the world', discloses the other, this other is not disclosed to me as a 'personal being', a 'subject' in the Cartesian or Kantian sense. Parallel to the anonymous character of myself as body-subject, the other is disclosed only as a body-subject, not as another 'I' considered as a personal being. This is what Merleau-Ponty's talk of the anonymous subject, of the 'one' (*on*), was meant to hint at. Thus, there remains a fundamental and inescapable asymmetry between myself and others. I can never live the point of view of the other as such; for that, I would have to stand to the other in the very relation in which he stands to himself, that is, to be him, in which case he would cease to be an *other*. What is more, for Merleau-Ponty there is even a grain of truth in solipsism:

---

16   For a discussion of this issue, see also Zahavi (2014: 78ff., 85ff., 203ff.).

> The difficulties of perceiving others are not all the result of objective thought, and they do not all cease with the discovery of behavior, or rather, objective thought and the resulting unicity of the cogito are not fictions, rather, they are well-founded phenomena, and we will have to seek their foundation. (PhP 373 [414])

Although it might seem that Merleau-Ponty is falling back into the aporiae of the intellectualist position, it is important to stress that this impression is mistaken. Intellectualism, if followed through in a manner consistent with its own principles, leads to the idea that there can ultimately only be one subject. We saw that, according to Merleau-Ponty, one of phenomenology's defining traits in contrast with the intellectualist tradition is precisely that at its roots it is open to the possibility of other subjects. We are confronted with others as a matter of fact; the problem concerns what this situation means: 'consciousnesses present the absurdity of a solipsism-shared-by-many, and such is the situation that must be understood' (PhP 376 [417]).

When he seems to endorse a form of solipsism, Merleau-Ponty wants to draw attention to the privilege of the first-person perspective. I am undeniably present to myself in a way that others can never be, at least for me. This self-presence is not necessarily that of the subject who is able to reflect and say of its representations that it is he who thinks them. For example, there is an undeniable self-presence even in unreflecting perception. I am pre-reflectively aware of my own perception and of its being my perspective on the world. This is what is at stake when Merleau-Ponty deals with what he calls the 'tacit cogito', in opposition to the Cartesian cogito, which he also calls the 'spoken cogito' (PhP 421ff. [461ff.]). This notion can easily be read as Merleau-Ponty's version of what Sartre calls the pre-reflective cogito. Sartre introduces this notion in *Transcendence of the Ego* and then, later, in *Being and Nothingness*, where it plays a pivotal role. With the notion of pre-reflective consciousness, Sartre showed that there is no contradiction between the fact that each of us is originally present to herself or himself in a pre-reflective way and the fact that this self-presence can be entirely pre-personal[17].

---

17   Sartre's conception of 'pre-reflective consciousness' was recently adopted by Dan Zahavi, who makes extensive use of it in his numerous writings. (See, for example, Zahavi 2005).

That we have a "pre-reflective" or "tacit" cogito through which we are always already pre-reflectively open to others – a cogito which is somehow prior to the "Cartesian cogito" in the sense that it is a condition of our being able to explicitly reflect on ourselves – is why we cannot read Merleau-Ponty's account of intersubjectivity as relying exclusively on the idea of a primordial identity with others (pace Dillon 1988). Indeed, appealing to this primordial identity does not solve the problem of intersubjectivity so much as prevent it from arising as a problem in the first place. Phenomenologically, we cannot escape the fact that our world is a social world. This means that others are always already here with me. My self-presence is always presence to the world, and others are disclosed along with that world regardless of how we are to account for this disclosedness. Merleau-Ponty expressly endorses Husserl's thesis that 'transcendental subjectivity is an intersubjectivity' (see PhP 378 [419]). For Merleau-Ponty, this means, roughly, that the other is a condition or presupposition of my ability to reflect and take myself as a transcendental subject. Here, the other refers not to a factically existing other but to the world's social horizon. Along with subjectivity and world, this horizon has been opened once and for all, and for this reason we cannot escape from the other in much the same sense that, for Merleau-Ponty, we cannot escape the world. We cannot cease to situate ourselves in relation to this horizon of otherness.

In other words, although solipsism not only exists as an abstract philosophical thesis but also has genuine phenomenological roots in the distinction between the way the other is given to us and the way each of us is given to ourselves, it does not understand itself adequately. Solipsism does not understand that what it says belies the situation it wants to express. Solipsism asserts that I am the only existent self. Not only does the solipsist's need to assert this contradict her own assertion – a well-known critique – but, Merleau-Ponty would add, she can also only pretend to take refuge in herself in opposition to others, because she finds herself already pre-reflectively surrounded by others in a social world.

The truth solipsism wants to express, although it does so in an inadequate way, is the fact that we are essentially free with regard to others. We will see below that, for Merleau-Ponty, freedom refers to the fact that I can never be totally pinned down by how others perceive me or by the situation in which I find myself. It is up to me to accept it,

make it my own, or transform and even escape from it. It is because I dispose of an unrestricted power to deny my situation and, correlatively, am always open to new ways of being that I can never be totally identified with my public ways of being. The same is valid for the other: even when we are sure of having experienced his innermost core, it is in principle always possible that we might be deceived in this respect. The other always manages to slip away beneath his most persistent character traits, such that the potential for misunderstanding him is unavoidable. Merleau-Ponty's point, however, is that the possibility that we may be fundamentally wrong about someone does not entail any kind of solipsism because it does not annul what I above refer to as the world's social horizon. Being wrong about others (and, Merleau-Ponty would add, being wrong about things) only means that I must replace a certain view of someone or something with another, but the fact that I am in a world – a world I share with others – cannot itself be escaped:

> As soon as existence gathers itself together and engages in a behavior, it appears to perception. And like every other perception, this one affirms more things than are grasped in it: when I say that I see the ashtray and that it is over there, I presuppose a complete unfolding of the experience that would have to go on indefinitely, and I open up an entire perceptual future. Likewise, when I say that I know someone or that I like him, I am aiming at an inexhaustible background beyond his qualities that indeed might one day shatter the image that I adopt of him. This is the price for there to be things and "others" for us, not through some illusion, but rather through a violent act that is perception itself. (PhP 378–9 [419–20])

At some point, Merleau-Ponty suggests the possibility that freedom from others may be realized by escaping our social world and choosing a secluded life in complete independence, treating them as mere natural things instead of other subjects (PhP 377 [418]). However, according to Merleau-Ponty, what this extreme possibility shows is that this kind of life always has the form of a flight from the world (in this case, a flight from the social word) and for that reason presupposes it. In general, solitude is only possible by reference to possible communication with others (PhP 376 [417]). I can only be or feel alone with reference to a possible being with others. If solipsism were actually true, the solipsist would never feel lonely, for loneliness already presupposes an original openness to others:

we must say about the experience of others what we have elsewhere said about reflection: that its object cannot absolutely escape it, since we only have a notion of the object through that experience.[18] Reflection must, in some way, present the unreflected, for otherwise we would have nothing to set against it, and it would not become a problem for us. Similarly, my experience must present others to me in some way, since if it did not do so I would not even speak of solitude, and I would not even declare others to be inaccessible. (PhP 376 [417]; trans. modified)

4.

I will now try to summarize Merleau-Ponty's understanding of sociality and the consequences of his view. What does it mean to say that sociality is an unavoidable horizon of our being in the world? We are not born into a private world without others, a world we must step out of in order somehow to reach them. We are always already with others in the world. However, this being with others is not the same as having a cognition or representation of them. We are already with others before we are able to turn them into objects of cognition. In the same sense that I am my body before I come to know my body as an object (as I move it around in space, for instance), intersubjectivity is a relation that precedes all cognition and knowledge. As we will see below (section 5), this does not entail that I can be reduced to a function of my social environment. If that were so, the privilege of my first-person perspective – the fact,

---

18   I have modified Landes' translation at this point. I replaced 'through reflection' with 'through that experience'. Landes took the French '*par elle*' in the original text to refer to 'reflection' instead of 'experience', and thus rendered the phrase 'through reflection'. Although the original French is ambiguous, in my view it only makes sense if '*elle*' refers to 'experience' for the sentence expresses the idea that reflection can only put into question or even revoke our pre-reflective notion of the other, because the latter is *in fact given* to us pre-reflectively. In other words, reflection by itself could never have come up with the idea of an *other* if it were not given to it beforehand in a pre-reflective way. (Incidentally, the older English translation of *Phénoménologie de la perception* by Colin Smith also uses 'through that experience'.)

that is, that there is only society, a social life and world for *someone* –
would be forever lost:

> Thus, we must rediscover the social world, after the natural world, not as an object
> or a sum of objects, but as the permanent field or dimension of existence: I can
> certainly turn away from the social world, but I cannot cease to be situated in
> relation to it. Our relation to the social, like our relation to the world, is deeper
> than every explicit perception and deeper than every judgment. It is just as false
> to place us within society like an object in the midst of other objects, as it is to
> put society in us as an object of thought, and the error on both sides consists in
> treating the social as an object. (PhP 379 [420])

Intersubjectivity is not a mere relation of being; at least by the time
of the *Phenomenology of Perception*, Merleau-Ponty maintained the
phenomenological assumption that all being must be being for me
(even if I am taken as a body). Rather, it can be characterized as the
kind of relation I entertain with what characterizes me as a fact, with
my 'facticity'. This relation stamps my relation to the world with
the mark of finitude but at the same time functions as a condition
of the possibility of my openness to that same world (the paramount
example of which is, of course, my body insofar as it opens me to
the world).

Relying on Husserl, Merleau-Ponty often uses the term 'operative
intentionality' (*fungierende Intentionalität*)[19] to characterize this pre-re-
flective, primordial relation to the other (but also to the body, time,
space, etc.) in contrast to the kind of intentionality that is typical of
cognition, that is, that posits its object explicitly before consciousness.
The 'object' of 'operative intentionality' dwells ambiguously in my per-
ception, simultaneously present and absent. Operative intentionality is
relevant in this context because it is the condition of the possibility of
act-intentionality – that is, of explicit cognitive acts.

Merleau-Ponty's middle path between objectivist and idealist ac-
counts of intersubjectivity can be summarized in the following idea: I
can only know sociality, culture and history because my point of view
of the world is itself social, cultural and historical. In other words, I only
have access to social, cultural and historical phenomena from my own

---

19   See, for example, PhP lxxxii [18], 440–1 [479–80], 453 [492], 472 [510].

particular social, cultural and historical perspective or point of view, whether I make this explicit to myself and others or not.

To the extent that it is an inescapable horizon of experience, the social world, much like the natural world, can never be adequately given to us in perception, and thus always remains on the fringe. It is essentially characterized as both a presence and an absence: a presence because I am in fact open to others and live in a social 'space', an absence because when I try to make clear to myself what the other is, how she is given to me, her innermost core always escapes me.

Intersubjectivity is thus an intrinsically paradoxical phenomenon. Merleau-Ponty does not shy away from this seemingly puzzling conclusion. In his view, phenomenology's task is not to solve the paradox but to make us aware of it. This is not to say that he does not try to ground the paradox of intersubjectivity in an ultimate condition of its possibility, but when he does so he is not so much attempting to solve the paradox as to link it to an ultimate fact about us. The latter is, according to Merleau-Ponty, our temporal being.

This is not the place to go into Merleau-Ponty's phenomenology of time. Suffice it to say, for him, following Heidegger, the sense of being must be read from that of time. Although time is always centred on an ever-changing present, its past and future dimensions reveal the same sort of presence-absence that characterizes the phenomenon of the other. If all being is ultimately temporal, if I am nothing but time, the fact not only that I am aware of others but also that my perspective is somehow informed by theirs is made intelligible. If to be temporal is essentially to exist outside oneself, if the present moment opens to a past before I was born and a future that will extend beyond my life, it is less puzzling that we are thrown into a social world we did not create – a world that, at the same time, presupposes a point of view from within it, from which it is possible to get a grip on it (much like every past and future presupposes a present from which they come into perspective, while at the same time the present presupposes all past moments and must make space for all subsequent, impending moments). In other words, what Merleau-Ponty will try to show in the chapter on temporality is that the temporal structure of the self lies at the root, among other things, of intersubjectivity and its paradoxical character.

5.

Most of the features that show up in Merleau-Ponty's analysis of inter-subjectivity are central to his account of freedom.[20] In the first place, Merleau-Ponty believes that we are free beings. The fact that we are free is, for him, closely connected to the fact that we are self-aware creatures, that is, subjects, and for this reason the precise nature of this freedom depends on the kind of subjectivity we have. This is why a book that is purportedly on 'perception' can close with a chapter on 'freedom': its main topic is actually the kind of subjects we really are.

Merleau-Ponty's endorsement of freedom is very much entailed by his rejection of empiricism. Empiricism is, among other things, the philosophical doctrine according to which perception can be explained away by third-person accounts of processes in the organism; it is therefore no surprise that Merleau-Ponty comes to view empiricism as being closely connected to determinism. According to this doctrine, everything we do is fully determined by its causal antecedents such that, having acted one way, we could not have acted differently (given these antecedents). Against determinists, Merleau-Ponty sides with the champions of freedom in the sense that he, too, holds that actions stem from self-aware creatures that can thus be held accountable for them. And yet, although Merleau-Ponty criticizes the empiricists, his real target in the chapter on freedom is a certain conception of freedom, of what it means for us to be free, which he sees as embodied in Sartre's *Being and Nothingness*.

Here I will not give a reading of the whole chapter on freedom; nor will I enter into detail on Sartre's own conception of freedom. My main aim is, rather, to show that one of Merleau-Ponty's main points of contention against Sartre's conception lies precisely in the role played by the intersubjective dimension of the self in his conception of freedom.

According to the intellectualist-idealist conception of freedom (which Merleau-Ponty seems to identify with Sartre), we are in principle

---

20   For discussions on Merleau-Ponty's account of freedom in the *Phenomenology of Perception*, see Langer (1989: 133ff.) and Priest (1998: 150ff.).

free whether we are explicitly aware of it or not, whatever we do and whatever our circumstances. Intellectualism pays no regard to our factical situation and views the subject as a pure, self-determining consciousness. In particular, for Sartre consciousness (or what he calls the 'for-itself') is a nothing, meaning that it cannot be determined as this or that by something outside itself. (The for-itself is, in Merleau-Ponty's view, another version of Descartes' cogito or Kant's transcendental subject, although Sartre characterizes it as pre-reflective.) Although the human being comes to regard herself as having qualities, permanent dispositions and character traits, she does this insofar as she does not live up to the nothing she is in her innermost core as consciousness. For example, whenever I regard myself as having determinate proprieties, I am looking at myself as if I were another; I objectify myself. In other words, according to Sartre's perspective, what I am for others, the for-others (*pour autrui*), negates my freedom.

Merleau-Ponty starts by pointing out that this conception of freedom is self-contradictory and self-undermining. First, since according to this conception each one of us is in his innermost core 'nothing' and thus free, we are all equally free; the slave who is resigned to his situation is as much a slave as he who breaks his shackles and sets himself free (PhP 461 [500]). Each one of our actions is free on this account; each manifests the same amount of freedom, and none is more deserving of this title than any other. For Merleau-Ponty, this conception of freedom causes us to lose sight of one particular feature of our natural concept of freedom, rendering it extremely counter-intuitive. There are some important cases where we feel that freedom, rather than being ready-made or innately part of us, is something for which we must strive, and thus some sort of achievement. In Sartre's view, '[t]he idea of an acquisition is rejected in the name of freedom, but then freedom becomes a primordial acquisition and something like our state of nature' (PhP 461 [500]). More than this, however, if freedom were actually how Sartre conceives it, freedom could not be made worldly; that is, it could not appear outwardly in our actions: 'Since we do not have to bring freedom about, it must be the gift granted us of having no gift, or that nature of consciousness that consists in not having a nature, and in no case can it be expressed on the outside or figure in our life' (PhP 461 [500]).

In Merleau-Ponty's view, Sartre's conception of freedom entails, furthermore, that I am at any moment able to pursue a different course of action to that in which I am presently engaged – that, at bottom, both my past and present circumstances and actions should be a matter of indifference when it comes to my absolute freedom since outside of myself nothing can determine me to action. This view disregards the fact that my past and present circumstances influence what I go on to do, that intentional activity stretches across time. Because of this, actions are only possible if time is not a mere sequence of instants, if my present action lives off the past, if what I did or decided moments ago lingers in what I go on to do (even if this ultimately entails giving up on a given project). The fact, for example, that I started writing this article early in the day and have been writing ever since might serve as a motive for me to carry on writing to the extent that it strengthens my self-confidence and my will to finish it. But it could equally serve as a motive for stopping; my early start explains my now being tired and supports the notion that I have surely done enough for the day. In either case, my earlier activity will impact my future choice. Had I instead decided to go out this morning and enjoy the beautiful outdoors, this would have played an equal role in explaining my subsequent behaviour: my guilt upon coming home might have motivated me to work through the night, or it might have discouraged me to the point of postponing work until tomorrow.

This touches on a central aspect of Merleau-Ponty's view on freedom. Freedom, like subjectivity in general, always presupposes a given starting point as it were – a given factual situation in relation to which the subject must decide, either by endorsing or refusing it. Decisions do not happen in a vacuum, as in Sartre's account (at least as Merleau-Ponty reads him). The subject always already finds herself in a certain context and social and cultural factors play a role in determining the field of possible action. Freedom can only exist in the interplay between an agent and her situation.

Given this critique, Merleau-Ponty tries to answer a possible objection from the Sartrean (or indeed idealist) side. The objection is that even though I always already find myself in a given factual situation, this situation has sense only for me, for all value and sense ultimately comes from me and thus I am not constrained by any situation except

those that I myself have 'created' (a version of this imagined objection can be found in PhP 463–4 [502]).

Merleau-Ponty does not deny that sense and value exist only for me, but everything hinges on what 'me' refers to in this sentence. There is a sense in which the way things appear to me depends on what Merleau-Ponty sometimes calls my personal being or self, that is, my free and explicit projects, my 'explicit intentions' (PhP 464 [503]). In relation to the project of writing this paper, my laptop appears conducive and serviceable; it is more to me than a material thing endowed with objective properties (being black in colour, solid, etc.). When we discussed the problem of intersubjectivity, however, we saw that the self can also refer to my pre-personal body-subject. In this sense, things depend on me as well – a mountain appears objectively impassable to me not because of some project or other that I entertain but in relation to my body (and to the body of every normal-sized human being).[21] This can be said to be the main lesson of the *Phenomenology of Perception*, this time applied to the topic of freedom: although it is true that in some sense the world is dependent on the self, this self is not a free-floating transcendental subject or cogito but rather inhabits the world that it constitutes. As a consequence, I, considered as a reflective, transparent, personal self, did not create the world or situation in which I find myself. There are objective limits to my freedom, limits that are not of my choosing and that do not depend on my explicit valuations and projects.

6.

Merleau-Ponty applies his views on freedom (which stem directly, or so it seems, from a critique of Sartre's conception) to his views on history and Marxist theory. Although our main focus here is not

---

21   The example of the mountain appears in Merleau-Ponty's text (PhP 464 [503]) although it comes originally from Sartre (see *Being and Nothingness* 481ff. [page numbers refer to the English edition]). Unsurprisingly, their respective analyses differ in important respects.

Merleau-Ponty's views on these themes, it is in the framework of his discussion of class consciousness that most of his more detailed views on the relation between sociality and freedom appear.

Concerning the problem of how class consciousness comes to pass, Merleau-Ponty's main contention is that it does not arise from the objective economic situation; that is, it is not something bound to happen as soon as certain objective conditions are met, and thus should not be viewed in a deterministic way. This means that class consciousness and its possible outcome, the proletarian revolution, are in a sense the product of human subjectivity and freedom. But this does not mean (again contra Sartre, who embodies the kind of intellectualist position Merleau-Ponty rejects) that the proletarian revolution is a permanent possibility, independent of actual life conditions in a given society. For the intellectualist who endorses an idea of absolute freedom, it is as if I could at any moment choose to be a 'proletarian' or 'bourgeois', while I 'for-myself', to use Sartre's terminology, remain nothing. In sum, as is always the case when it comes to Merleau-Ponty's way of dealing with philosophical problems, both the objectivist-empiricist and the intellectualist-idealist account of class consciousness and proletarian revolution are revealed to be inadequate. As he himself writes, 'idealism and objective thought equally miss the arrival of class consciousness, the first because it deduces actual existence from consciousness, the other because it derives consciousness from actual existence, and both of them because they are unaware of the relation of motivation' (PhP 473 [511]).

While I am not specifically concerned here with Merleau-Ponty's account of class consciousness, it is interesting to see what this account tells us about his views on intersubjectivity and freedom. Contrary to Sartre's account, social position and economic situation matter to the emergence of class consciousness. However, they matter not insofar as they are *represented* in a purely intellectual and detached manner, but insofar as they are *lived* by me. It is my concrete life conditions as they are *lived* and *felt by me* that may give rise to class consciousness. Merleau-Ponty's point, in this respect, is very much the same as the point he makes in the chapter on intersubjectivity (see section 4 above). Before I take up a position regarding, say, class struggle (and as a presupposition of this decision), I am already socially and economically situated. I exist

as a factory worker, a teacher, a business executive, an academic, etc. This does not mean that my concrete existence infallibly determines the position I ultimately take in my life. I may very well decide to abandon corporate life in favour of voluntary work in Africa. But this decision is only possible because it is preceded by corporate life and all that goes with it, and my serving as a volunteer only makes sense if understood as called for, as motivated, by my previous way of life.

For Merleau-Ponty, my life displays sense (or, as the other meaning of the French *sens* has it, *direction*) that can only be lived and that does not follow from taking an explicit position on it. This sense or direction of my life, which includes the way I identify (or fail to identify) with others around me, is not something that I can represent or know explicitly. It is, according to my interpretation, yet another instance of operative intentionality (see section 4). To illustrate the kind of sense or direction life has, comparisons with the lived body are once again inevitable: just as my body displays a special kind of intentionality by being engaged in motor activities (playing piano, or tennis, or even opening a door), such that if someone were to ask me how to do what I am doing, I would find it difficult to explain and would ultimately simply tell her to practise it herself, so, according to Merleau-Ponty, life exhibits an analogous phenomenon of sense: a sense that can only be recognized through the activity, which in this case is life itself.

For Merleau-Ponty, this spontaneous sense or direction of life can be felt as a common destiny among people. In his example of the development of class consciousness, workers come to view their respective existences and destinies as being tied to those of all other workers and opposed to those of exploiters – a sense of things which might be powerful enough to ground a future proletarian revolution. The latter is not an end posited in advance and is not a conscious choice.

The fact that life can assume this collective sense is a presupposition of the very possibility of historical truth and sense, according to Merleau-Ponty. History is revealed as having the same structure as individual life inasmuch as it is an interchange between given social situations and the way individuals embody or transform them. Even though the sense of history sometimes seems to be determined exclusively by the will and actions of powerful individual figures and leaders, such as Napoleon, their actions must somehow be related to the sense already

prefigured in collective history. If history were the mere outcome of the volition of free individuals, it would display no sense or direction. The idea of distinct periods or cycles would be absurd; anything could happen at any time, depending on the groundless will of certain individuals (PhP 474 [513]).

At the same time, whenever Merleau-Ponty talks of the sense or direction of a community of people, and thus the sense or direction of history, it should be kept in mind that this sense or direction must be lived by the individuals who comprise that people, even if in a non-reflective or 'operative' sense, and that the sense of history should not be mistaken for a necessary (Hegelian) progression towards some predetermined end. Moreover, according to Merleau-Ponty, both at the individual and the collective level, the sense of life is by its very nature multivocal. Since events can always be seen from multiple perspectives, and since there is no absolute perspective given that each of us grasps the world from our own point of view, what the sense of history is in each case depends on *whose* perspective we are talking about. To take the previous example, the bourgeois does not identify with the workers' interests and demands, and nor is his sense of his life bound up with the proletarian revolution. Given human freedom, this does not preclude his joining the workers' struggle, as often happens, but it makes it less likely. At the individual level, since the sense or direction of life is essentially lived, and for this reason also ambiguous, it is always possible to be mistaken about it. Moreover, our temporal nature, the fact that life is always in some sense open, that at every passing moment the sense of my life must be reconfigured, means not only that I can be wrong about myself but that the possibility of changing the course of my life is essentially open to me.

7.

To conclude, I will try to bring together some loose ends and take another look at the argument as a whole, further developing some of its points.

Merleau-Ponty, as much as Sartre, links freedom to subjectivity.[22] At bottom, this idea derives directly from Kant's notion of spontaneity. Spontaneity has a theoretical and a practical dimension. Theoretically, it is manifested in our ability to think ourselves as the subject of our thoughts (the transcendental subject); practically, it is manifested in our capacity to be autonomous, free agents. What is essential to this view is the idea that we can rise above the realm of nature in both a cognitive and a practical sense. Sartre's conception of subjectivity – that is, his concept of the 'for-itself' (*pour-soi*) – again takes up Kant's notion of spontaneity inasmuch as the for-itself equates to our intrinsic reflectivity and our non-determinacy as subjects: the idea that, as subjects, we cannot be determined by being (by the in-itself). According to this view, we are 'nothing' and can thus be said to be free.

Up to this point, Merleau-Ponty would not object to this view of ourselves as spontaneous in the relevant sense – that is, as both cognitive and practical beings. The problem for Merleau-Ponty lies, rather, in the way our spontaneity, and thus also the nature of our subjectivity, is to be understood. As mentioned above (see section 3), I read him as endorsing Sartre's notion of pre-reflective consciousness as a major advance in comparison to intellectualist conceptions of subjectivity: the subject's inherent reflectivity does not require an explicit act of reflection; we are usually 'for ourselves' without performing any explicit act of reflection, any explicit 'I think…', and thus what I call our inherent reflectivity as subjects is implicit, or, to use Sartre's term, 'pre-reflective'. In other words, consciousness is by its nature always self-consciousness in a sense that does not require the self to attend explicitly to itself as such (and this is why I deem it 'inherently reflective').

What Merleau-Ponty objects to in Sartre's conception of the cogito lies in the strict opposition Sartre establishes between it and his other major category of being, the 'in-itself'. For Merleau-Ponty, the strict opposition and mutual exclusion between the 'in-itself' and the 'for-itself' lies at the root of Sartre's failure to account adequately for, among other things, intersubjectivity and freedom. There must be some

---

22   See, for example, the following passage, where freedom is intrinsically linked to subjectivity: "My freedom, that fundamental power I have of being the subject of all of my experiences … " (PhP 377 [418]).

sort of mediation between the 'in-itself' and the 'for-itself' to the effect that the former can be seen as in some sense included in the latter.[23] Merleau-Ponty's main thesis in this regard is that the for-itself (the subject) cannot be set apart from his factical situation. In other words, the for-itself is the for-itself of his body, his tradition, his culture, etc. According to Merleau-Ponty, none of these 'factors' is essentially opposed to the 'for-itself'. On the contrary, the 'for-itself' is only 'for-itself' – that is, a perspective on the world, or, as Merleau-Ponty puts it, a certain grip on the world because he himself is an embodied creature. He is socially, culturally and historically situated and this in such a way that only in an abstract sense can the 'for-itself' be set apart from these features, which make up what we have been calling his 'facticity'. Not only is his 'facticity' not essentially opposed to his being a subject, to being-for-itself, but it is, in truth, its condition.

Applied to the problem of intersubjectivity and freedom, Merleau-Ponty's point is that I can only set myself apart from others (and thus be free *of* them) because I always already find myself situated in relation to them. Put in more Sartrean terms, the 'for-itself' can certainly be conceived as 'negation' and 'nothingness', but as a determinate negation. I am not an absolute nothingness, in which case I would be absolutely free. My nothingness always ensues from the negation of being (of the body, nature, others) and for that reason essentially presupposes it:

> The central phenomenon, which simultaneously grounds my subjectivity and my transcendence toward the other, consists in the fact that I am given to myself. *I am given*, which is to say I find myself already situated and engaged in a physical and social world; *I am given to myself*, which is to say that this situation is never concealed from me, it is never around me like some foreign necessity, and I am never actually enclosed in my situation like an object in a box. My freedom, that fundamental power I have of being the subject of all of my experiences, is not distinct from my insertion in the world. I am destined to be free, to be unable to reduce myself to any of my experiences, to maintain with regard to every factual situation a faculty of withdrawal, and this destiny was sealed the moment that my transcendental field was opened, the moment I was born as vision and as knowledge, the moment I was thrown into the world. (PhP 377 [417–8])

---

23   It is this lack of mediation that Merleau-Ponty would later, in his *Adventures of Dialectics*, identify as Sartre's main mistake. See especially p. 142: 'Contrary to appearances, being-for-itself is all Sartre has ever accepted, with its inevitable correlate: pure being-in-itself.'

The fact that I am socially situated does not preclude my freedom; it is a condition of my freedom. My actions are never the outcome of absolute freedom or absolute determinism. For Merleau-Ponty, this is a false dilemma. Instead, my actions can be said to be motivated by my factical situation, but I am no less free for it. In this regard, my situation is only decisive insofar as it is freely *lived* by me.

References

Barbaras, Renaud (2004). *The Being of the Phenomenon. Merleau-Ponty's Ontology*. Trans. Ted Toadvine and Leonard Lawlor. Bloomington: Indiana University Press.

Carman, Taylor (2008). *Merleau-Ponty*. Abingdon, UK: Routledge.

Dillon, M.C. (1988). *Merleau-Ponty's Ontology*. 2nd ed. Bloomington and Indianapolis: Indiana University Press.

Hass, Lawrence (2008). *Merleau-Ponty's Philosophy*. Bloomington: Indiana University Press.

Heidegger, Martin (1962). *Being and Time*. Trans. J. Macquarrie, E. Robinson. New York: Harper & Row, 1962.

Husserl, Edmund (1960). *Cartesian Meditations. An Introduction to Phenomenology*. Trans. D. Cairns. The Hague: Martinus Nijhoff.

Langer, Monika (1989). *Merleau-Ponty's Phenomenology of Perception. A Guide and Commentary*. Hampshire: Macmillan Press.

Merleau-Ponty, Maurice (2002). *Phenomenology of Perception*. Trans. Donald A. Landes. Abingdon, UK: Routledge, 2012; Trans. Colin Smith. Abingdon, UK: Routledge, 2002.

Merleau-Ponty, Maurice (1964). *Sense and Non-Sense*. Trans. H.L. Dreyfus & P. A. Dreyfus. Evanston: Northwestern University Press.

Merleau-Ponty, Maurice (1973). *Adventures of Dialectics*. Trans. Joseph Bien. Evanston: Northwestern University Press.

Merleau-Ponty, Maurice (2010). *Child Psychology and Pedagogy: The Sorbonne Lectures 1949–1952*. Evanston: Northwestern University Press.

Matthews, Eric (2002). *The Philosophy of Merleau-Ponty*. Chesham, UK: Acumen.

Priest, Stephen (1998). *Merleau-Ponty*. London: Routledge.

Romdenh-Romluc, Komarine (2011). *Routledge Philosophy Guidebook to Merleau-Ponty and Phenomenology of Perception*. Abingdon, UK: Routledge.

Sartre, Jean-Paul (1966). *Being and Nothingness*: *A Phenomenological Essay on Ontology*. Trans. H. Barnes. New York: Washington Square Press, 1966.

Zahavi, Dan (2003). *Husserl's Phenomenology*. Stanford: Stanford University Press.

Zahavi, Dan (2005). *Subjectivity and Selfhood. Investigating the First-Person Perspective*. Cambridge (MA): MIT Press.

Zahavi, Dan (2014). *Self & Other: Exploring Subjectivity, Empathy, and Shame*. Oxford: Oxford University Press.

The plural role of emotions in Ethics: the case of ethical dilemmas

Dina Mendonça and João Sàágua

This paper identifies a plurality of roles for emotions in practical reasoning by examining how different ethical approaches enable the recognition of the role of emotions in decision-making. By using ethical dilemmas we show how emotions appear in Intuitionism, Utilitarianism, and Deontological Ethics. The analysis reveals that emotions can appear in different places: sometimes as motivational forces, other times as ends-in-view, sometimes as overarching contextual modes, and finally that they are sometimes reasons for action and at other times they are causes for actions. In addition, we argue that the most recent developments of philosophy of emotions demand a more complex perspective about emotions and that the impact of emotions in ethics should take into consideration layers of emotions (Mendonça 2013). We conclude that the issues raised reinforce Bernard Williams' claim that our notion of rationality is incomplete without emotions and sentiments for 'it would be a kind of insanity never to experience sentiments of this kind towards anyone, and it would be an insane concept of rationality which insisted that a rational person never would' (Williams 1981: 29), and suggest a series of future issues to be explored which would further explain the connection between values and emotion as to do justice to a more complete and rich notion of rationality.

1. General motivation

Almost any reflection about emotion reveals their complex structure and the complex connections with other aspects of the mind and action. Philosophers, psychologists and neuroscience try again and again to

provide a picture of the impact of emotional experience to untangle the rich reality of feelings and emotions and to provide a good taxonomy of the emotional phenomenon. The issue is far from settled and some theorists of the emotion field have established that it is not possible to find a definition that includes all emotions for they 'vary so much in a number of dimensions – transparency, intensity, behavioral expression, object-directedness, and susceptibility to rational assessment – as to cast doubt on the assumption that they have anything in common' (De Sousa 2014: 6). Nevertheless it is possible to identify some common traits. One of these is that emotions have an ambivalent position with regard to their role for certainty and action. As Peter Goldie pointed out, though we 'are inclined to say that emotional experience can sometimes tell us things about the world that reason alone will miss' (Goldie 2004: 249), it is also the case that 'we are inclined to say that our emotions can and do profoundly distort our view of things: in anger or jealousy, for example' (Goldie 2004: 249). That is, sometimes emotions seem to be the source of certainty as when, for instance, a strong gut feeling sensation tells us what course of action to take even though the evidence does not add up in a secure way to give us that certainty; however, at other times, emotions are the source of deception and self deception as when, for instance, fear can turn every detail into an indication of danger or when loving someone blinds the one in love to aspects that everyone clearly recognizes as undesirable.

We think that finding the ways to understand the normative force of emotions within the rich and complex world of emotions may provide some clues to a better understanding of this ambivalent stance of emotional reality. Thus, this paper aims to provide a general picture of the various ways in which emotion can appear in action and decision-making so as to provide a first step to grasp their normative force to better disentangle their ambivalent nature. We begin by outlining three templates for Ethical Theory (Intuitionism, Utilitarianism and Deontological Ethics) and show how contemporary research has made it clear that a good understanding of rationality is incomplete without including the experience of sentiments and emotions. Then we suggest that by looking at an ethical dilemma we can better see how the different ethical templates understand the role of emotions in decision-making. We go on to explore the trolley problem and imagine how

different ethical positions reveal different roles for emotions in order to grasp the overwhelming richness of the impact of emotion on decision-making. Finally, we indicate that understanding the role of emotion in decision-making requires the need to integrate emotional complexity. The notion of emotional reflexivity revealed in emotional layers highlights that the ambivalence of emotions may be better understood once emotional complexity is introduced. We conclude by pointing out several future work directions for the benefit of both Ethics and Philosophy of Emotions.

2. Three templates for Ethical Theory

Ethics, as a philosophical discipline, is populated with theories, some more 'in' than others; in a sense every philosopher of Ethics worthy of her/his name has her/his own theory. Among other things, these theories are supposed to address 'big' moral questions – e.g. What is the right thing to do? What is an adequate theory of good? – and to frame and, if possible, to solve moral dilemmas (e.g. the Trolley Problem, about which more below). It is uncontroversial and well known that this myriad of theories can be grouped, if we ascend to a more abstract level, in families of theories. For the present purposes we are going to considerer three such families and give a conceptual snapshot of them. This will allow us further in this paper to suggest what the different roles are that emotions can play within each family, and this, in turn, will illuminate a bit more the philosophical DNA of what each family is. The three families we are going to considerer can be named as follows: Intuitionism, Utilitarianism and Deontological Ethics. It is clear to us that these three families by no means exhaust the actual families of ethical theories on the philosophical scene. For reasons we are not going to enter into, in order to make the snapshot of each family we will concentrate on the 'Right' instead of on the 'Good'.

Faced with a situation or with a prospect of an action with moral import, Intuitionism states that your most powerful tool to cope with it is your *moral intuition*. This moral intuition is as natural to mankind

as, say, reasoning or perception, and like these two it can also be edu-
cated. So, for instance, if you are preparing yourself to perform action
A with moral import, then (unless you are hampered) you know from
that moment on if this is a right or wrong action to perform. This is a
very powerful starting point for the theory since there is large empirical
evidence and cogent philosophical arguments (into which we are not
going to enter) that support the relevance of moral intuition in practical
decisions with moral import. Two of the most salient problems faced
by this family are: (1) possible conflicting moral intuitions associated
with moral situations (for instance: you may have the intuition that you
should do A, but also the intuition that you should do B, A and B being
incompatible; or you may have the intuition that doing A is both right
and wrong); (2) context or even individual sensitiveness of the 'moral
intuition faculty' (not every culture, much less each individual, has the
same moral intuitions, not even the same individual at different times
of his life).

Faced with a situation or with a prospect of an action with moral
import, Utilitarianism states that you should always maximize pleasure
and minimize pain for all concerned, perhaps allowing for: (a) a 'bonus'
relative to the ones who are closer to you (e.g. if you can only save one
drowning person, save your daughter, not the young girl near her that
you are not acquainted with); (b) an adjustment within different pleas-
ures (arguably, some pleasures/pains are more worthy to have/ more
pressing to avoid than others); and (c) some prudential judgments on
your practical reasoning (while fostering actual pleasure, consider if
it does not jeopardize future ones). One of the most salient problems
faced by this family is: possible justification of impingement of great
pain on a smaller number of people to enhance the pleasure of a larger
number of people (e.g. slavery, and see below the Trolley Problem).

Finally, faced with a situation or with a prospect of an action with
moral import, Deontological Ethics states that you should always freely
act according to your duty, and for the sake of it and nothing else (e.g.
not to be praised by others, or not to be caught doing some illegal ac-
tion). Kant, who is the father of this family, stated this maxim in differ-
ent ways, with the two most interesting statements of it being (roughly):
to act in such a fashion that your acting might be considered a universal
rule; and, in your acting never consider the others as means but always

as ends in themselves. (Note that the latter statement 'kills' immediately the Utilitarian view.) Two of the most salient problems faced by this family are: (1) possible conflicting moral duties applied to the same action or situation (you can easily adapt the examples just given to illustrate the akin problem for Intuitionism); and, (2) possible sensitiveness of moral duties (the same examples given for Intuitionism apply here *mutatis mutandis*).

All the three families, it is clear, try to make sense in three different even incompatible ways of our moral experience and moral actions. Let us now turn to the emotions side and see what is going on there, and then try to connect both as we suggested at the beginning of this section.

3. Rationality is incomplete without emotions and sentiments

It has become more and more visible that our conception of rationality falls short if we do not include how feelings, emotions and sentiments make part of it. Though traditionally the capacities for deliberation and judgment have been taken as more rational than the capacity for emotion, and consequently 'any other mental state (such as the emotions) that conflicts with the outcomes of deliberation and judgment must *ipso facto* be irrational' (Helm 2000: 4), it is by now completely clear that Reason is no longer in opposition to Emotion. That is, the present state of affairs is now such that everyone would totally agree with Williams when he writes that our notion of rationality is incomplete if we do not include such sentiments in rationality for 'it would be a kind of insanity never to experience sentiments […] towards anyone, and it would be an insane concept of rationality which insisted that a rational person never would' (Williams 1981: 29). Thus, the last ten years of development of emotion theory have enabled the settlement that emotions are crucial for rationality even though they may at times come apart and 'present one with inconsistent perspectives on the world' (Helm 2000: 9). Nevertheless, it is still not clear what such a complete picture looks like because we cannot obtain this more complete picture simply by adding

emotions to our notion of rationality. It requires a reconceptualization of rationality and how cognition incorporates and interacts with emotion and a good understanding of emotions themselves. Ultimately, it will also need an explanation of the long historical legacy of thinking of emotions as opposed to reason because the final integrated picture will have to provide an insightful explanation as to why emotions appear to be sometimes in conflict with some thought processes.

One way to contribute to the reconceptualization of rationality in light of the developments in emotion theory is to consider the role of emotions in decision-making. We think dilemmas are promising because we can place the response of a specific theoretical position and then verify what added insight is given by looking into the role of emotion so as to provide a way to see the role emotions play in decision for action. We have chosen the dilemma of the Trolley problem, first introduced by the philosopher Phillippa Foot (1978), because of the way it reveals crucial aspects of the way consequences and principles interact in face of moral dilemmas in which both possible alternatives imply a tragic outcome. We do not aim to offer a detailed defense of a specific answer to the dilemma, nor to fully point out what the different answers to the different trolley problem situations reveal about the guiding principles that guide us in moral situations. Instead, we simply show that the three identified ethical templates give insufficient answers to the dilemma such that Ethical Intuitionism fails and the other two approaches provide answers with irreconcilable aspects for their theoretical structure, and then list the different ways in which emotions appear to play a role in the dilemma within each Ethical template. The point is to argue as 'Taylor suggests that one's emotions are somehow central, but exactly how is left unclear.' (Helm 2000: 3), and that even though we may not exactly be certain as to the impact of emotions in decision-making we are sure they have an impact which is part of rationality and not in opposition to it. Examining the way emotions appear in face of an ethical dilemma increases our theoretical awareness of the plural positions emotions can take in decision-making.

4. Emotions in Dilemmas: Trolley Problem and the phenomenology of hesitation

In his 1985 article entitled 'The Trolley Problem', Thompson describes the dilemma in the following way:

> 'Suppose you are the driver of a trolley. The trolley rounds a bend, and there come into view ahead five track workmen, who have been repairing the track. The track goes through a bit of a valley at that point, and the sides are steep, so you must stop the trolley if you are to avoid running the five men down. You step on the brakes, but alas they don't work. Now you suddenly see a spur of track leading off to the right. You can turn the trolley onto it, and thus save the five men on the straight track ahead. […] Is it morally permissible for you to turn the trolley?' (Thompson 1985:1395) The description of the dilemma comes with a second hypothetical case in which you have to imagine yourself to be a surgeon, a truly great surgeon. Among other things you do, you transplant organs, and you are such a great surgeon that the organs you transplant always take. At the moment you have five patients who need organs. Two need one lung each, two need a kidney each, and the fifth needs a heart. If they do not get those organs today, they will all die; if you find organs for them today, you can transplant the organs and they will all live. But where to find the lungs, the kidneys, and the heart? The time is almost up when a report is brought to you that a young man who has just come into your clinic for his yearly check-up has exactly the right blood-type, and is in excellent health. So, you have a possible donor. All you need do is cut him up and distribute his parts among the five who need them. You ask, but he says, 'Sorry. I deeply sympathize, but no.' (Thompson 1985: 1396)

What the problem and the second hypothetical case raise is a refinement of the issues surrounding the Trolley Problem when it was first suggested by Foot and brings forth the question 'Why is it that the trolley driver may turn his trolley, though the surgeon may not remove the young man's lungs, kidneys, and heart?' (Thompson 1985: 1396).

Of course there is an expanding body of experimental evidence showing that people make choices for (at least in part) unknown reasons, and then make up reasonable justifications while remaining unaware of the gap between their real motivation and their ex-post rationalization (T. Wilson: 2002) and, with regard to the role of emotion in moral dilemmas, experts have mostly identified the role of moral emotions as crucial elements that take part in the final decision outcome.

For example, in 'An Experimental Investigation of Emotions and Reasoning in the Trolley Problem' Lanteri, Chelini & Rizzello found that 'the immediate responses may be traced to moral emotions as opposed to moral reasoning' (Lanteri, Chelini, & Rizzello 2008: 793). They conclude their paper stating that, 'humans probably have a set of hard-wired moral emotions immediately triggered by some features in a choice situation – for instance, among others, personal-moral features' (Lanteri, Chelini, & Rizzello 2008: 801).

We would like to build upon their effort and examine what other emotional input can be identified in face of a dilemma using the three ethical templates described above. Of course we know that all three theoretical positions can take up more complex identities than the ones we described. For instance, Utilitarian can also take the form of rule-utilitarianism. However, for the purpose of this paper it is sufficient to take up the proposed three templates and explore how Intuitionism, Utilitarianism and Deontological Ethics bring forth different ways in which emotions can play a role in decision-making, and establish what are the differences that can make a difference in the normative role of emotional processes.

When faced with the Trolley Problem, intuitionists recognize that intuitions about the dilemma change depending on how the case is presented such that notions of intention, proximity or the general framing modify the given intuitions about the right thing to do (Stratton-Lake 2011: 15–18). When we add to ethical intuitionism the insight of emotions, we capture one way to ground the self-evident nature of basic moral propositions and understand how moral emotions ground what is intuitively grasped and stand as a non-inferential judgment (Roeser 2006: 38). As Sabine Roeser writes, '[w]e can understand emotions as fulfilling the role of non-inferential judgments or intuitions' (Roeser 2006: 42) and even though not only emotions are required to focus on what is morally relevant, emotions are normative judgments such that '[p]aradigmatically, moral intuitions are emotions' (Roeser 2006: 42). In addition, emotions provide the space for empathy and sympathy when emotions give rise to different moral emotions in different people because they provide the arena to establish a dialogue in which each party gives examples and draws analogies as to promote in the other a

similar emotion to share the emotion and intuition regarding the case in hand (Roeser 2005: 83).

Utilitarianism deals with the Trolley dilemma by using their maxim that we should always act so as to produce the most amount of happiness for the greatest number of people. The outcome is impossible to maintain for the one principle, which is supposed to guide all human action, asks that the surgeon remove the organs of one patient to save all the others and consequently offers an answer that goes against the best judgment. Utilitarianism is well known for how its single maxim can imply a justification of a great pain to some in the name of the happiness of many, and how this asks the Utilitarian to find a way to acknowledge the feelings of respect for persons giving rise to something similar to a moral intuition suggested in the first ethical template. So though we would assume that the role of emotions for Utilitarianism is the end result and consequence of the feeling of happiness, the confrontation with the dilemma reveals that the feeling of respect for persons also needs to be taken into account. Thus the greatest number of people also requires incorporating other feelings and the feeling of happiness is not an isolated emotional experience. The insight for philosophy of emotion is that when an emotional order is the end in view, it necessarily incorporates other emotional aspects that must be coordinated in order to avoid the self-effacing change (Stoker 1976).

In contrast with Utilitarianism, when a Deontological ethics takes up the Trolley Problem the action is taken to be more important than its consequences. Within this frame of work, to act morally one must follow the rules and the action must be such as to be according to the maxim that you can also will that it would become a universal law (to use Kant's first formulation of the Categorical Imperative) and it looks like feelings and emotions have no possible role to play. It is common to take this reading of deontological ethics: it is in clear contrast with the utilitarian position where the emotional mark of Happiness (overall Happiness) dominates the decision with regard to the emotional input for decision-making such that we could say that in decision-making it is the foreseen possible emotional outcome that dictates the role of emotion in rationality. However, if you follow Greenspan you can take a deontological position and argue that emotions play more than one role

in ethical deliberation and that, just as intuitionists who integrate emotions argue, there is a normative role of emotion reinforcing and favoring reasons motivating action (Greenspan 2011: 43). The overall picture of this interpretation of deontological ethics implies that principles do not stand on their own since they are motivated by emotional input, leaving the theory also open to the self-effacing charge and arguing for an unsustainable position.

Nevertheless, one interesting aspect of Greenspan's work is her departure from the widespread assumptions that a stress on the role of emotion is essentially Humean (Greenspan 2011: 44) and how she argues that emotions play two roles in ethics: first, emotions supply moral judgments with motivational force and secondly, that emotions stand as sources of reasons. As Greenspan explains, when someone is treated unjustly, feeling upset if the injustice goes unchallenged is not only a reason for the state of emotional discomfort carries normative implications (Greenspan 2011: 45). Greenspan writes, '[t]he fact that one is uncomfortable about something counts in itself as a reason for action – action to prevent the feeling from continuing – apart from any properties attributed to its object. So an appropriate emotion, besides having an evaluative component that reflects a practical reason, can add as a further reason a criticism, from the agent's standpoint, of her own state of feeling. The fact that she is in a state of discomfort is normative insofar as it actually counts against her failure to act to relieve it, whether or not the discomfort or her recognition of it also serves as a motive.' (Greenspan 2011: 45) That is, emotions can appear elements that reinforce reasons and have a causal force while at other times they stand as independent reasons (Greenspan 2011).

The presentation of the three possible ethical postures in face of the Trolley Problem reveals that emotions can play a plurality of roles and the complete description of their impact in ethics is too wide for the scope of this paper. We want to propose that the recent developments of philosophy of emotions indicate that the fact that emotions also come in layers (Pugmire 2005, Jäger & Bartsch 2006; Mitmansgruber et al. 2009; Mendonça 2013; Jäger & Bänninger-Huber 2014; Norman & Furnes 2014; Howard 2015), which is not directly contemplated by any of the ethical positions explored, is decisive for a good understanding of the role of emotions in ethics.

5. Reflexivity of emotions

Previous work has shown that meta-emotions cannot be handled as a special case of emotion because reflexivity modifies the nature of our emotional world. (Mendonça 2013) The reflexivity of emotions is such that one emotion about another changes the meaning and value of the first-order emotion such that being angry about being sad and being proud of being sad ends up being a completely different emotional experience of sadness. Since 'meta-emotions necessarily have an impact on the value of the first-order emotion […] This means that when we feel a meta-emotion, its object (the first-order emotion) changes and with it also changes the emotional experience.'(Mendonça 2013: 394) In addition the added significance brought by the meta-emotion is not a simple addition to the meaning of the first-order emotion but has a transformative effect because the 'information obtained with the description of meta-emotions is not simply a matter of having more information about the experience; the extra knowledge we get from meta-emotions may change the meaning of the experience altogether.' (Mendonça 2013: 394) This means that the inclusion and recognition of the role of meta-emotions in decision-making and deliberation may enable us to make sense of the earlier suggestion that sometimes emotions work as causes of actions and other times as reasons for actions, such that emotional input in decision-making may be the result of refinement of meta-emotional processes.

Though psychologists and philosophers usually illustrate meta-emotions as strategies for healthy emotional regulation, there is no trait in reflexivity that guarantees this and consequently, the relationship between emotions and meta-emotions can be far more complex and negative than the usual positive regulative connections (Howard 2015: 11–15). That is, though reflexivity does not come necessarily with a positive self-corrective direction, just like thinking about thinking does not, it is a privileged ground for instances of regulatory mechanisms and just as thinking about thinking can guide and correct thinking, emotions about emotions can refine and correct feeling. Thus, the meta-emotional mark may be the crucial item to explain why there is a sense in which people are and feel responsible

for their emotions even though it is also the case that people are not always in control of their emotional experiences.

Concluding Remarks and Future Work Directions

The general picture provided by the exploration of the place of emotions in the different ethical templates shows that emotions appear in different places in decision-making. Thus, the way emotions play a role in moral judgment indicates that some kind of version of sentimentalism is needed in ethics to do justice to the way emotions contribute to our ethical decisions (Avramova & Inbar 2013: 170). However, a comprehensive view needs to include insight from different ethical theories so as to grasp the full variety of roles for emotions. We have shown that emotions appear as follows: 1) emotions can be non-inferential judgments such as intuitions and be the normative judgments indicating what is the right action in a moral situation; 2) emotions can be the necessary base for sympathy and dialogue with people who have different judgments about a moral situation; 3) emotions can appear as the goal and consequence to obtain (happiness) and this is different from other emotional outcomes of a specific moral action; 4) emotions can provide the needed motivational force such as feelings about values, a sentiment of respect for the moral law, and when this is taken into account it is crucial to differentiate emotions that appear as causes connected to the descriptive realm of moral action or appear as reasons connected to the normative realm of moral action (for example, anger may be a reason to demand justice (reason) while it can also be a cause for an unjust behavior (cause)); and finally 5) the emotional layer can modify by reinforcing or erasing the force of a first-order emotion.

We think that this plurality of possibilities and added complexity of layers of emotions is at the heart of the ambivalent role of emotion in decision-making and that further research into the plurality of roles will ultimately provide a clearer understanding of the normative force of emotion. Thus, we would like to suggest some follow-up research directions raised by the analysis undertaken that constitute added steps

towards a clearer picture of the normative role of emotions which would show how emotions are one of the ways rationality is guided by values.

First, it would be important to further explore how emotions stand as reasons, building on Davidson's distinction between reasons and causes and following Michael Brady's arguments that emotions stand as a source of reasons and rationality in general (Brady 2013). When an agent runs away out of fear within a moral situation, the agent is at the mercy of an emotion in a causal mode and fear here can only be described within the descriptive realm. However, imagine that the same agent feeling fear does not run away but acts in caution both postponing taking up an action as much as possible while at the same time gathering the most amount of information about the situation. In this case the action of suspension of action seems to be grounded on fear as a reason for caution and one can easily state the general norm that in case of fear one ought to act in a cautious manner, while it is impossible to state a norm that says that in case of fear one ought to run away as fast as possible for in some instances the dangerous situation asks for not running away. The double possibility of the place of emotions is important to seriously take up the neo-sentimentalist proposal for it provides a way to understand that arguing for the importance of emotions for the right action does not mean that all emotions can be subsumed under this description. It is perhaps much harder to establish which ones can be included but, however, the above analysis suggests that only emotions that can be described under the normative realm can be morally integrated and that it is important to identify that the same emotion can be taken within the descriptive realm without annihilating its normative force.

Second, the acceptance of the complexity given by the reflexivity of emotion proves once again that the impact of emotion on ethics goes beyond that of its valence and appeals to a sense of emotional coherence that can be best analyzed in the notion of character. Thus it would be crucial to explore how virtue ethics deals with ethical dilemmas such as the Trolley Problem and if it stands up to the charge of falling into a similar self-effacing position (Pettigrove 2011). Finally, this will ask us to verify in what way this understanding of the role of emotions in ethics has an impact on the education of emotion and the use of education of emotion for moral education.

   Third, the provided analysis clearly asks for an evaluation of the pertinent connection between emotions and values. One possible way to uncover the links of the hierarchy of values implicit in people's moral choices is to explore a phenomenology of hesitation in face of moral dilemmas similarly to the way Cooke explored a phenomenology of error and surprise by comparing the work of Peirce, Davidson and MacDowell (Cooke 2011) in order to better understand the role of surprise within the cognitive experience of error. Likewise, it is possible to draw a phenomenology of hesitation built upon the reflection about a dilemma in order to better understand the various roles of emotion in decision-making and in which way emotions give emphasis to this or that value when people face moral decisions in which they hesitate. A phenomenology of dilemmas and hesitation may further explain how the connection of emotion and character is tied to our understanding of emotional depth and do justice to the way in which emotional depth is connected to excellence of character (Pugmire 2005) and healthy feelings (Dewey 1887). Since dilemmas offer a privileged field to analyze the role of certain specific moral emotions such as regret, compassion and anger, it would be interesting to explore the suggestion that moral emotions represent a desired pattern of emotional structure about other ethical situations. In addition, it may also provide a good interpretation to the way people face dilemmas in daily life that, though they may not have the dramatic edge of the trolley problem, may feel similar to those who experience them.

   In conclusion, when we examine the role of emotions in ethical dilemmas by ethical templates, it becomes clear that though emotions occupy a plurality of roles they have a normative function when they appear as reasons for action. A better understanding of their reflexive nature and how it influences and changes the way in which they can work, sometimes as causes and other times as reasons, will provide a clearer account of their normative force. Ultimately, we think that this normative trait of emotion is partly responsible for the structured format of the plurality of values in which rationality is embedded. Thus, the level of coherence or incoherence and the internal conflicts that can occur in decision-making within a plural conception of values can never be fully understood without an integration of the place and role of emotions. More importantly, we can never attain a complete image

of emotion without investigating the decisive factors that make them sometimes function as reasons for action. In sum, Ethics is incomplete without a good understanding of emotion just as Philosophy of Emotions is deficient if it does not incorporate the way emotions function as reasons in decision for action. That is, at the core of a better understanding of rationality there is an on-going demand for a better understanding of the role of emotions in decision-making.

References

Avramova, Y. R., & Y. Inbar (2013). 'Emotion and Moral Judgment'. *WIREs Cogn Sci*, 4:169–178.

Brady, M. S. (2013). *Emotional Insight: The Epistemic Role of Emotional* Experience. Oxford: Oxford University Press.

Cooke, E. F. (2011). 'Phenomenology of Error and Surprise: Peirce, Davidson, and MacDowell'. *Transactions of the Charles S. Peirce Society*, Vol. 47, No. 1 (Winter), 62–86.

Davidson, D. (1963). 'Actions, Reasons and Causes'. *Journal of Philosophy,* 60 (23): 685–700.

de Sousa, R. (2014). 'Emotion', in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). http://plato.stanford.edu/archives/spr2014/entries/emotion/. Accessed August 29, 2016.

Dewey, John. (1987). 'Psychology', in *The Early Works of John Dewey 1882–1898, vol. 2, 1887: Psychology*. Carbondale: Southern Illinois University Press.

Foot, P. (1978). 'The Problem of Abortion and the Doctrine of the Double Effect', in *Virtues and Vices and Other Essays in Moral*

Philosophy. Berkeley and Los Angeles: University of California Press, 19–32.

Greenspan, P. S. (1995). *Practical Guilt: Moral Dilemmas, Emotions, and Social Norms*. Oxford: Oxford.

___ (2011). 'Craving the right: Emotions and Moral Reasons', in Carla Bagnoli (ed.). *Morality and the Emotions*. Oxford: Oxford University Press: 39–61.

___ (2004). 'Practical Reasoning and Emotion', in A. Mele and P. Rawling (eds.), *The Oxford Handbook of Rationality*. Oxford: Oxford University Press: 206–221.

Goldie, P. (2004). 'Emotion, Reason and Virtue', in D. Evans and P. Cruse (eds.), *Emotion, Evolution, and Rationality*, Oxford: Oxford University Press: 249–67.

Howard, S. A. (2015). 'Metaemotional Intentionality'. *Pacific Philosophical Quarterly*, doi: 10.1111/papq.12093.

Jäger C., & A. Bartsch (2006). 'Meta-emotions'. *Grazer Philosophische Studien* 73: 179–204.

Jäger C., & E. Bänninger-Huber (2015). 'Looking into meta-emotions'. *Synthese* 192:787–811.

Jones, K. (2004). 'Emotional Rationality as Practical Rationality', in C. Calhoun (ed.), *Setting the Moral Compass: Essays by Women Philosophers*: 333–52. New York: Oxford University Press.

Lanteri, A., C, Chelini, & S. Rizzello (2008). 'An Experimental Investigation of Emotions and Reasoning in the Trolley Problem'. *Journal of Business Ethics* 83:789–804.

Mendonça, D. (2013). 'Emotions about Emotions'. *Emotion Review* vol.5, No.4 (October): 390–396.

Mitmansgruber, H., et al. (2009). 'When you don't like what you feel: Experiential avoidance, mindfulness, and metaemotion in emotion regulation', *Personality and Individual Differences* 46: 448–453.

Norman, E., & B. Furnes (2016). 'The concept of "metaemotion": What there is to learn from metacognition?'. *Emotion Review* Vol. 8, No. 2 (April): 187–193.

Petersen, M. B. (2010). 'Distinct Emotions, Distinct Domains: Anger, Anxiety and Perceptions of Intentionality'. *The Journal of Politics*, Vol. 72, 2 (April): 357–365.

Pettigrove, G. J (2011). 'Is Virtue Ethics Self-Effacing?' *Ethics* (2011) 15: 191–207.

Pettit, Philip (1991). 'Consequentialism', in Peter Singer (ed), *A Companion to Ethics.* Oxford: Blackwell: 230–240.

Pugmire, D. (2005). *Sound Sentiments: Integrity in the emotions*. Oxford: Oxford University Press.

Roeser, S. (2006). 'A Particularist Epistemology: "Affectual Intuitionism"'. *Acta Analytica* 21 (1): 33–44.

____ (2006). 'Intuitionism, Moral Truth, and Tolerance'. *The Journal of Value Inquiry* 39: 75–87.

Sherman, N. (1990). 'The Place of Emotions in Kantian Morality,' in O. Flanagan and A. O. Rorty (eds.), *Identity, Character, and Morality*: 149–70. Cambridge, Mass: MIT Press.

Smith, A. M. (2005). 'Responsibility for Attitudes: Activity and Passivity in Mental Life'. *Ethics*, Vol. 115, No. 2 (January): 236–271.

Stocker, M. (1976). 'The Schizophrenia of Modern Ethical Theories'. *The Journal of Philosophy*, Vol. 73, No. 14: 453–466.

Stratton-Lake, Philip (2011). 'Intuitionism in Ethics', in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition). http://plato.stanford.edu/entries/intuitionism-ethics/ Accessed September 13 2016.

Tannenbaum, J. (2007). 'Emotional expressions of moral value'. *Philosophical Studies* 132: 43–57.

Thompson, J. J. (1985). 'The Trolley Problem'. *The Yale Law Journal*, Vol. 94, No. 6 (May): 1395–1415.

Vogelstein, E. (2011). 'Morality, Reasons and Sentiments'. *Philosophical Studies* 155: 421–432.

Williams, B. (1965). 'Ethical Consistency'. *Proceedings of the Aristotelian Society* (Supplement), 39: 103–124.

_____ (1981). 'Moral Luck', in *Moral Luck.* Cambridge: Cambridge University Press, 20–40.

Wilson, T. (2002). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press.

John McDowell on practical rationality – is he (really) talking about us?

Susana Cadilha

1.

In what follows I shall try to give an account of John McDowell's conception of practical rationality, using for the most part his collection of articles *Mind, Value and Reality* (1998). My overall aim is to argue that McDowell's conception of human practical rationality is not in line with what we know about the way we think and act. Not being representative/typical of people like us, real agents, it is not, I think, a realistic conception.

I will give particular attention to two papers McDowell wrote arguing against two other famous philosophers – Philippa Foot and Bernard Williams: 'Are moral requirements hypothetical imperatives?' (1978) and 'Might there be external reasons?' (1995). There he tries to answer the following question: *what does it mean to say that someone has a reason to act in a specified way?* Williams (1981) famously argued that there are only internal reasons, meaning that one only has reason to do whatever practical reasoning,[1] starting from one's existing motivations,

---

1 Williams presents no restricted account of practical reasoning, though – a practical reasoning is more than the mere discovery that some course of action is the means to an end; it is 'a heuristic process, and an imaginative one' (Williams 1981: 110). 'A clear example of practical reasoning is that leading to the conclusion that one has reason to ϕ because ϕ-ing would be the most convenient, economical, pleasant etc. way of satisfying some element in S [the agent's subjective motivational set] …. But there are much wider possibilities for deliberation, such as: thinking how the satisfaction of elements in S can be combined, e.g. by time-ordering; where there is some irresoluble conflict among the elements of S, considering which one attaches most weight to … ; or, again, finding constitutive

may reveal that one has reason to do.[2] It is not that one has only reason to do what in a way satisfies some element in one's subjective motivational set, but that those elements govern the practical reasoning leading up to the conclusion that one has reason to do something.

McDowell, on the other hand, defends that there are external reasons – that there are reasons to act unconnected with our existing motivations. How do people acquire such reasons? How do they start believing that there is a reason to act in a certain way, if there is no connection whatsoever with the subject's motivational set? In order to be an external reason, that reason must have been there all along, so that in coming to see it, the agent must be arriving at a proper consideration of the matter. How come we manage to get things right?[3]

Let us focus on a typical domain of practical rationality – the *ethical* domain. According to Williams, ethical reasons are internal reasons; according to McDowell they are external reasons. This means there are ethical reasons for us to do something even if we are not able to see them and there is no practical reasoning or deliberative path that can take us there. The question is: how can we get things right, as would the virtuous person?[4] How do we come to believe there is a reason for acting in a specified way and how can we acquire a new motivation *by* getting things right?

McDowell is not purely Kantian – he does not say that the agent is able to get things right because he is able to deliberate correctly, i.e. through a pure rational procedure. He clearly states that 'the transition to being so motivated is a transition *to* deliberating correctly, not one effected *by* deliberating correctly' (McDowell 1998: 107). No pure rational procedure would make us consider the matter aright – for instance,

---

solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment' (Williams 1981: 110). 'Imagination can create new possibilities and new desires' (Williams 1981: 105).

2   It is important to notice that according to Williams's interpretation it is not required that the agent is actually motivated to do what he has reason to do.

3   By deliberating correctly, Williams would say. If there were external reasons, there would be a procedure of correct deliberation that gives rise to a motivation, but is not controlled by nor connected to the agent's existing motivations.

4   McDowell follows Aristotle and his virtue ethics – he thinks the most important thing in ethics is to be the right person (the well-educated one). The virtuous person is the measure of the right action, and not the other way around.

seeing that we should give back the wallet some passer-by has dropped. But that does not mean there is no reason to do that, and I would be able to see it if I was the right kind of person. If I had a proper ethical upbringing, I would have my eyes opened to some reasons I otherwise cannot see. As with someone who had not the benefit of an artistic education and hence cannot properly enjoy the experience of a work of art, someone who has not been properly brought up cannot see the reason why he should give back the wallet. But that reason exists (it is an external reason) – and 'it might take something like a conversion to bring the reasons within the person's notice' (McDowell 1998: 107).[5]

Getting things right – figuring out which ethical reasons there are – is then a matter of 'tuning up' our moral perception. What distinguishes a virtuous person (who can clearly see what should be done) from a non-virtuous one is not that the former has different motivations – she simply *sees* things differently.

This is what leads us to McDowell's most controversial theses. If acting correctly is just a matter of seeing/perceiving correctly, any moral fault will be a cognitive fault. This means that if I am not able to do the right thing (for instance, to give back the wallet I just found), it is not because I lack motivation to do it, but only because I lack the knowledge that it is the thing to be done. The difference between an honest and a dishonest person is not that they have different motivations or interests; rather, that difference lies exclusively in their ways of perceiving their circumstances. Thus, it would not be possible for two people to have exactly the same understanding of the circumstances and yet see different reasons to act.

Briefly, then, according to McDowell, it is a person's understanding of how things are that gives her a reason for action. And if I were the right person, I would see the right reasons to act.[6] Moral reasons,

---

5   'In moral upbringing what one learns is not to behave in conformity with rules of conduct, but to see situations in a special light, as constituting reasons for acting; this perceptual capacity, once acquired, can be exercised in complex novel circumstances' (McDowell 1998: 85).

6   McDowell is a moral particularist: there is no rule or criterion to define what the right action is; it will always depend on the particular context. The virtuous person is the one who knows how to act in each occasion, who is sensible enough to distinguish the particular features of each situation. As I said before, the virtuous one is the measure of the right action.

in particular, have no direct link with the agent's existing motivations or interests; there is no need, in addition to her understanding of the relevant facts, for the agent to *care* about the situation, meaning that some desire would function as an independent and extra help in order to motivate her. Her belief does that on its own.[7]

The problem is, of course, a Humean one – is it possible for a purely cognitive state (a view of how things are) to entail some disposition to act, or to make the action attractive to its possessor? Hume would put it like this: does reason motivate?

McDowell would simply say that to assume that cognitive and conative/affective states have distinct existences is just a Humean dogma. Similarly, we do not have to take for granted that the world is, in itself, 'motivationally inert'.

My worries about this view, I must say, have less to do with the worldview it presupposes than with the *picture of mankind* this view leaves us with. *Are we really like that?*

According to this view, there is no possible situation in which someone has the relevant understanding of the situation (say, that the thing to do is to give back the wallet) and is not motivated to act accordingly. Thus, believing that giving back the wallet is the thing to do necessarily entails wanting to do it (if the agent A thinks there is a reason to ϕ in a particular case Y, then he must be willing to ϕ in Y).[8]

---

7   This is how T. Nagel puts it: 'That I have the appropriate desire simply *follows* from the fact that these considerations motivate me; if the likelihood that an act will promote my future happiness motivates me to perform it now, then it is appropriate to ascribe to me a desire for my own future happiness. But nothing follows about the role of the desire as a condition contributing to the motivational efficacy of those considerations' (Nagel 1979: 29–30).

8   This is usually referred as motivational internalism: 'The names 'internalism' and 'externalism' have been used to designate two views of the relation between ethics and motivation. Internalism is the view that the presence of a motivation for acting morally is guaranteed by the truth of ethical propositions themselves. On this view the motivation must be so tied to the truth, or meaning, of ethical statements that when in a particular case someone is (or perhaps merely believes that he is) morally required to do something, it follows that he has a motivation for doing it. Externalism holds, on the other hand, that the necessary motivation is not supplied by ethical principles and judgments themselves, and that an additional psychological sanction is required to motivate our compliance' (Nagel 1979: 7).

My doubts about this intellectualist account are the following: *does that description really match the way people are, and act?* Is it really the case that I don't give back the wallet just because I don't know what the thing to do is (I just have the illusion that I know)? Closely related with this is the description McDowell presents of the virtuous agent's moral psychology: the virtuous person simply does not need to weigh reasons, because once he sees what is the thing to do, every other contrary reason that he might have simply vanishes – 'the dictates of virtue, if properly appreciated, are not weighed with other reasons at all, not even on a scale that always tips on their side. If a situation in which virtue imposes a requirement is genuinely conceived as such, according to this view, then considerations that, in the absence of the requirement, would have constituted reasons for acting otherwise are silenced altogether – not overridden – by the requirement' (McDowell 1998: 90).

It seems that McDowell has in mind some kind of *ideal agent*, not a real one. But it is a kind of ideal with no particular function attached, because there is no way to teach a non-virtuous man to become virtuous and hence no definite way to get closer to that ideal. And what about *akrasia*? According to this view it seems impossible that someone may act contrarily to his best judgment. If the akratic person knows he is not acting as virtue demands, then most likely he conceives the circumstances of his action as the virtuous person would conceive them. But then, if acting correctly is just a matter of perceiving the matter correctly, there is no room left to akrasia – if someone conceives the situation as the virtuous person does, then he would know what to do and any other considerations that might constitute reasons for acting otherwise would simply be silenced.

The only solution available to McDowell is simply to posit that the incontinent person's understanding of a situation does not match that of a virtuous person.[9] But in that case, the mere conceptual possibility of akrasia vanishes. If there cannot be a perfect match with the way a fully virtuous person conceives the circumstances of his action,

---

9   'The way out is to attenuate the degree to which the continent or incontinent person's conception of a situation matches that of a virtuous person' (McDowell 1998: 92).

then an akratic action is conceptually impossible since it is never the case that someone acts contrarily to his best judgment; people behave differently just because they have different understandings of what is to be done.

In a nutshell, what I am arguing is that, while not being a pure Kantian, still McDowell inflates the agent's rationality by stating that if the agent thinks he has a (moral) reason to do X, then he wants/is motivated to do X. What I say is that *we are not like that*: sometimes, we really think that we must do X, or that we have reason to do it, but still we want to do something else.

If Hume has a minimalist conception of practical rationality (it has only an instrumental role, one of finding the right means to attain a given end), McDowell stands to blame for the opposite excess, assuming the intellectualist position that practical knowledge necessarily entails motivation to act; that the agent must want to do what he has a reason to do. Neither of these seems to give an accurate account of how rationality and desire combine in order to originate action. If it is true, on the one hand, that we can rationally deliberate about ends and not only about means (that desires are subject to rational criticism), it is also true, on the other, that there is no guarantee that the agent's motivation will always align with the agent's reasons, or that the agent necessarily wants to do what he thinks is the best to do.

2.

So far I have been defending that McDowell gives an inflated account of practical rationality and thus he is not actually speaking about real agents, people like us.

Connected with that thesis, there is another way in which I think McDowell shows his alignment to that classical philosophical conception of mankind according to which *human beings are the exemplars of rationality and autonomy*. In fact, McDowell thinks that man is a creature who stands apart from animals by virtue of his powers of self-control, reasoning, and reflection – that there is a clear line that separates

the human from the animal way of living.[10] This is because McDowell draws a very sharp distinction between conceptual and non-conceptual, cognitive agents and thinkers.

Continuing to think about the ethical domain, it is easy to see how McDowell is prone to recognize the autonomy of any normative domain such as the ethical one. There is an is-ought gap and no communications allowed. That means that moral matters are purely conceptual and rational matters – thinking what we should do is a rational ability only humans have, and any descriptive or psychological aspect of man is pulled apart from that ability. I mean: the instinctive tendencies we share with other animals do not determine that conceptual and rational ability; and that rational ability cannot be explained in a way that is not itself rational.[11]

My doubts are the following: is it really the case that when it comes to moral matters our 'first nature' traits are simply overridden? That we get rid of all of our natural determinations? That with the 'onset of reason', as McDowell puts it, the practical tendencies that are part of our first nature simply vanish? My opinion is that it is not very plausible to think, with McDowell, that there is an abrupt chasm between biologically determined creatures, on the one hand, and creatures moved only by reasons, on the other. Our rational and conceptual abilities do not override our animal nature.

It seems clear to me that only rational beings are capable of *elaborate* moral systems and *sophisticated* forms of moral thinking. Sophisticated forms of moral thinking imply conceptualization and abstract

---

10   '…we do not fall into rampant Platonism if we say the shape of our lives is no longer determined by immediate biological forces. To acquire the spontaneity of the understanding is to become able, as Gadamer puts it, to "rise above the pressure of what impinges on us from the world" (*Truth and Method*, p. 444) – that succession of problems and opportunities constituted as such by biological imperatives – into a "free, distanced orientation" (p. 445)' (McDowell 1994: 115–116).

11   'Moral education enables one to step back from any motivational impulse one finds oneself subject to and question its rational credentials. Thus it effects a kind of distancing of the agent from the practical tendencies that are part of what we might call his first nature. … If the second nature one has acquired is virtue … [its] dictates acquired an authority that replaces the authority abdicated by first nature with the onset of reason' (McDowell 1998: 188).

reasoning. After all, besides being capable of feelings of outrage in face of asymmetry and unfairness (this is an inequity aversion that we share with non-human primates),[12] we are also able to design sophisticated constructs such as theories of justice. The ability to morally evaluate that characterizes us at this point in our development involves the ability to pose what philosophers usually refer to as the 'normative question': think about what should be the case, question the assumptions and the consequences of action. Now, this is not an automatic behavior or an instinct. This fully developed ability to think morally is what characterizes us as moral beings. What seems questionable is to conceive of no continuity whatsoever between one thing and another and to sustain that our ability to think morally is of a fundamentally different nature, which keeps us irremediably apart from the 'mere' dispositions and feelings of non-linguistic animals. What seems questionable is the idea that being a moral agent has to do with ability for conceptual thinking, but not also with the ability to repudiate certain asymmetries in situations. It seems plausible to say that there is a link between this fully developed capacity we exhibit today and the intuitions and dispositions probably exhibited by our ancestors. My point is this: because we are linguistic beings, capable of conceptual and abstract thinking, we come to a level of sophistication in terms of moral thinking that allows us to think in terms of reasons, and to develop theories that allow the justification of moral positions before the members of the community who also have the ability to discuss them. But the fact that we have reached this level does not mean that the ability to assign value to items in the world, and perhaps the content of some evaluative positions, may not have been influenced and shaped by factors other than rational reflection. It seems to me legitimate to think that there was evaluation and value assignment before there was a rational capacity for justification. This basic capacity to experience items in the world as things requiring certain reactions or counting for certain reactions precedes a linguistically mediated reflective ability to pose the normative question. So, because we are sophisticated creatures we can take a step back with respect to these primitive evaluative dispositions or intuitions and not follow them

12   See De Waal & Berger (2000) and Brosnan & De Waal (2003).

compulsorily; but the fact that we are reflective creatures who can take that step back does not entail that such dispositions cannot yet influence our moral judgments.

Another thing which is difficult to believe in is the thesis that our rational and conceptual capacities are completely untainted by other aspects of our psychology. If we take a careful look, for instance, at some experiments on moral psychology,[13] we may be able to see that it is not the case that our moral judgments always arise out of data manipulation and further rational deliberation. Rather, what we usually define as a moral judgment may after all have its basis in a 'gut reaction' and may not be an expression of propositional knowledge. When faced with certain types of morally innocuous transgressions (like using a national flag to wipe the floor, or drinking a glass of water after having spat in it), people show the same kind of reactions that moral transgressions elicit (they are thought of as being universally wrong, of a non-contingent and mandatory nature, their wrongness independent from authority), even though they cannot find a reason to do so. This appears to bring moral judgments close to a certain kind of affective response in which reflection over propositional contents plays little or no role at all.[14]

These experiments are in line with numerous experimental studies that represent the core of cognitive psychology, and that rest on the hypothesis that most of our judgments result from the triggering of fast and frugal heuristics, and not from deliberative processes. It is not absurd to think that the same happens with moral judgments: they result from heuristics and many of them are automatic.[15] (This does not mean moral reasoning has no place, but it looks like its main function is that of a post-hoc rationalization – it is useful to justify previous intuitions or whenever a conflict between moral intuitions arises). In fact, in many different areas of research it has been found that people make evaluations (as to whether an event/object is good or bad, for instance) immediately, unintentionally and without awareness that they are doing it, so

---

13   Cf. Haidt et al. (1993); Haidt & Bjorklund (2008); Greene et al. (2001).
14   See also Nichols and Folds-Bennett (2003).
15   One of the simplest heuristics studied in this field is that which makes us immediately agree with and positively value what is said by people we like.

it may be the case that 'what we think we are doing while consciously deliberating in actuality has no effect on the outcome of the judgment, as it has already been made through relatively immediate, automatic means' (Bargh & Chartrand 1999: 475). It is not absurd to think that the influences of heuristics and biases uncovered in recent cognitive psychology are widespread in everyday ethical reflection.

So, it might be the case that human beings are not paragons of rationality and autonomy. But if we stick to McDowell's theory of practical rationality, it is clear that the capacity that determines, in a given situation, what matters about that situation and that enables us to evaluate it, is a conceptual conscious ability that only rational animals possess (it is the result of being initiated into a 'conceptual space', as McDowell puts it). In a practical syllogism – that can be used to deliberate or to organize an agent's reasons for action – a judgment determining which feature of the situation matters constitutes one of the premises. And it is also clear that the actions through which we manifest our moral character must be chosen; even if McDowell grants, following Aristotle, that virtuous action is the result of habit, we must not understand that as happening out of instinct or inertia – on the contrary, virtue requires that 'specially human capacity for discursive thought' (McDowell 1998: 39). But if we consider virtue-ethical ideals of practical rationality in light of the model of human cognition now emerging, we realize that moral behavior is not immune to cognitive biases and that it does not always flow from reflectively endorsed moral norms or robust traits of character like virtues. Rather, we see that minor situational influences (such as ambient noise, or the fact that someone is in a hurry) determine moral behavior.[16] In fact, various experiments in social psychology revealed that subjects were much more likely to help someone in need if they had just found a dime, or are not in a hurry, or if the ambient noise was at normal levels. Circumstantial and morally irrelevant factors influence moral behavior in a decisive way, and can also influence the way we perceive the situation as an occasion for ethical decision. And it is extremely relevant that those cognitive biases or response tendencies are beyond the reach of individual practical rationality.

---

16   Cf. Isen & Levin (1972); Darley & Batson (1973); Mathews & Cannon (1975).

It thus seems as though not only McDowell's conception of moral abilities but also his idea that it suffices to believe that X is the thing to do in order to be motivated to do it are not in line with what we know about the way we are and think. It is possible to simply argue that real agents are defective practical reasoners, but in that case we need to admit that there is a distance between the picture of human cognition that applies to virtuous people and the model of human cognition now emerging in the cognitive sciences that applies to everyone else. And how useful and illuminating can that be?

My point in this paper was just to argue that from a philosophical perspective, no less than from an empirical one, McDowell's account of practical rationality is not realistic, since it seems to ignore features that are determinative of us as human agents.

References

Aristotle (2009). *The Nicomachean Ethics*. Oxford: Oxford University Press.

Bargh, J., & T. Chartrand (1999). 'The unbearable automaticity of being'. *American Psychologist* 54 (7): 462–479.

Brosnan, S., & F. De Waal (2003). 'Monkeys reject unequal pay'. *Nature* 425: 297–299.

Darley, J., & D. Batson (1973). 'From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior'. *Journal of Personality and Social Psychology* 27: 100–108.

De Waal, F., & M. Berger (2000). 'Payment for labour in monkeys'. *Nature* 404: 563.

Greene, J., Sommerville, R., Nystrom, L., Darley, J., & J. Cohen (2001). 'An fMRI investigation of emotional engagement in moral judgment'. *Science* 293 (21): 2105–8.

Haidt, J., Koller, S., & M. Dias (1993). 'Affect, culture, and morality, or is it wrong to eat your dog?'. *Journal of Personality and Social Psychology* 65(4):613–28.

Haidt, J. (2001). 'The emotional dog and its rational tail: A social intu-
    itionist approach to moral judgment'. *Psychological Review* 108:
    814–834.

Haidt, J., & F. Bjorklund (2008). 'Social intuitionists answer six ques-
    tions about moral psychology', in W. Sinnott-Armstrong (ed.),
    *Moral Psychology, Vol.2: The cognitive science of morality: intu-
    ition and diversity*. Cambridge (MA): MIT Press: 181–219.

Isen, A., & P. Levin (1972). 'Effect of Feeling Good on Helping: Cook-
    ies and Kindness'. *Journal of Personality and Social Psychology*
    21: 384–388.

Mathews, K., & L. Cannon (1975). 'Environmental noise level as a de-
    terminant of helping behavior'. *Journal of Personality and Social
    Psychology* 32: 571–577.

McDowell, J. (1994). *Mind and World*. Cambridge (MA): Harvard
    University Press.

___ (1998). *Mind, Value and Reality*. Cambridge (MA): Harvard
    University Press.

Nagel, T. (1979). *The Possibility of Altruism*. Princeton: Princeton
    University Press.

Nichols, S., & T. Folds-Bennett (2003). 'Are children moral objectiv-
    ists? Children's judgments about moral and response-dependent
    properties'. *Cognition*, 90(2): B23-B32.

Pettit, P., & M. Smith (2006). 'External Reasons', in Cynthia Mac-
    donald and Graham Macdonald (eds.), *McDowell and his critics*.
    Oxford: Blackwell Publishing, 2006: 142–169.

Williams, B. (1981). 'Internal and External Reasons', in S. Darwall,
    A. Gibbard and P. Railton (eds.), *Moral Discourse and Practice*.
    Oxford: Oxford University Press: 363–372.

Rules and personal changing

REGINA QUEIROZ

Introduction

Even if Cavell (1990) admits the socio-political nature of rules, those who suffer political and social injustices see themselves as *excluded and not represented by the rules of society*. Consequently, since the existing rules neither represent them nor allow them to expose the political injustices, thus satisfying their claim to justice, people are voiceless and unable to show how certain institutions are unfair. A persons' recovery of their voice depends on a personal changing, understood as a miracle. Prior to a subjective *judgment* without rules, formulated in a conversation of justice (Cavell 1990), this personal changing seems to correspond to the private approach to rules depicted by Wittgenstein in *Philosophical Investigations* (1953).

Although we acknowledge the contribution of Cavell's exegesis of Wittgenstein to expose the existential issues underlying Wittgenstein's philosophy (Cavell 1969, 1979, 1990; see Mulhall 1994, 2007) in addition to his original interpretation of Wittgenstein's private language argument (Cavell 1969, 1979, 1990; see Mulhall 1994, 2007), we argue that any personal changing requires the mediation of political rules, following criteria, even though these rules are in the end abnormal rules. We recognize the research on Wittgenstein and Cavell (Bernstein 1981; Conant, 2005; Eldridge 2003; Fleming & Payne 1989; Hammer 2002; McGinn 2004; Mulhall 1994). However, neglecting neither that research on Wittgenstein and Cavell nor Cavell's original interpretation of Wittgenstein's private experience, the relationship between the private language arguments and the miraculous personal changing has not received enough attention.

We are also aware of the widespread disagreement on Wittgenstein's private language arguments (e.g. Baker and Hacker 1985; Canfield 1986; Cavell 1979; Hacker 1990; Kenny 1984; Malcolm 1977; 1989; Mulhall 2007; Nielsen 2008) and on rule-following (Cavell 1990; Cook 1965; Fogelin 1976; Kripke 1982; McDowell 1984). Nevertheless, our main aim is not to provide a personal judgment on those challenging and proficient interpretations of Wittgenstein's philosophy of language. Our main aim is to *use* Wittgenstein's main intuitions and concepts embedded in his *later* philosophy of language for the understanding of some political life issues, namely the understanding of individual and political changing. Thus, in spite of the controversy about the explicit political content of Wittgenstein's philosophy of language (Pitkin 1972), and also acknowledging that some authors have already given that contribution (Bloor 1997; Cavell 1979, 1990; Holt 1997; Pitkin 1972; Taylor 1992; Temelini 2015; Tully 1989), we aim to show that personal changing requires the mediation of political rules, following criteria, even though these rules are in the end abnormal rules.

From these premises, in the first section we will present Cavell's description of the miracle of changing, explaining its emergence from Cavell's existential exegesis of Wittgenstein's private language arguments.

In the second section we present some of Wittgenstein's arguments against the private language argument as well as Cavell's perspective on the private experiences issue, and we argue that no personal changing can dispense with political criteria and rules. We also distinguish normal from abnormal rules. We conclude that *no one changes without* explicit or implicit rules framed by criteria.

Following Cavell's use of Ibsen's play, *A Doll's House* (1879 [1981]), we also illustrate our arguments by using the same play.

1. The Miracle of Changing

Even if Cavell (1990) admits the socio-political nature of rules, those who suffer political and social injustices see themselves as *excluded and not represented by the rules of society*. Consequently, since the existing rules

neither represent them nor allow them to expose the political injustices, thus satisfying their claim to justice, people are voiceless and unable to show how certain institutions are unfair. Recovering their voice on justice depends on a subjective *judgment* without rules, formulated in a conversation of justice (Cavell 1990).

In spite of the linguistic nature of that judgment, Cavell argues that 'unless something is shown' (Cavell 1990: 117) the conversation of justice 'cannot go on – there is nothing to say' (Cavell 1990: 117). Since victims face an unquantifiable personal misery and social and political injustice – 'right is not assertible' (Cavell 1990: 12) – the recovery of any political right depends on an inner changing process, understood as the 'miracle of changing' (Cavell 1990: 111). In the context of reflection on the drama of consent, the analysis of which is beyond the scope of our article, Cavell makes use of Ibsen's character in his 1879 masterpiece *A Doll's House* (1879 [1981]) to illustrate this miracle.

Described as resentful, outraged, dishonoured, and a woman shamed by her husband, Torvald, Nora has truly lost her way and is unable to show how unfair the institution of marriage is that treats 'a grown woman, a wife and mother, as a child' (Cavell 1990: 114). Indeed, in the play, Nora Helmer once secretly borrowed a large sum of money so that her husband could recuperate from a serious illness. She never told him of this loan and has been secretly paying it back in small instalments by saving from her household allowance. The man from whom Nora borrowed the money, Nils Krogstad, threatens to reveal Nora's crime, and when Torvald discovers that Nora has forged her father's name, he is ready to disavow his wife even though she had done it for him. He declares that Nora is immoral, unfit to be a wife and mother, and says he will continue to be married to her in name alone. When Nils says that he no longer wants to blackmail the Helmer family, Torvald rejoices, declaring that they are saved. He then says that he forgives Nora and that he still loves her as his little 'caged song bird.' This is a startling wake-up call for Nora Helmer. In a flash, she realizes that Torvald is not the loving, selfless husband she had once envisaged. She also comes to understand that their marriage has been a lie, and that she herself has been an active part in the deception. She then decides to leave her husband and her children in order to find out who she truly is.

Although Cavell's evaluation of Nora's decision as a claim for po-
litical justice is not consensual – Nora's rejection of sacrifice and op-
pression and her subsequent leaving of her husband and three children
has been seen as the expression of an unprincipled, abnormal, neurot-
ic, irresponsible, egoistic, deceitful, manipulative and immoral woman
(Templeton 1989, 1997) – her internal changing is exhibited by external
and nonverbal signs. For example, contrary to her daily routine, instead
of going to bed and taking off her fancy clothes, Nora reappears wearing
her outdoor clothes, signalling that she is free and released from de-
ception (e.g. being imprisoned by *unfair social and political rules and
seeing herself as an accomplished person*).

In fact, since Nora's nineteenth-century political context required
women to sacrifice personal liberty and accept paternalistic treatment
(Langås 2005; Templeton 1989; Yuehua 2009), the claim for a woman's
personal liberty *could be neither heard nor justified* under political and
social rules. Also, in the face of the main sacrificial and paternalistic
rules of Nora's society, her final decision could not avoid being seen as
anything but unreasonable, if not irrational, for whoever would benefit
from unfair rules. Additionally, seen as *a priori* orders (Cavell 1979,
1990), the consequences of which are 'confined and scored' (Cavell
1990: 115), rules prevent understanding the quest for the personal re-
covery of voice.

Thus, the existence of someone completely excluded from the
(outside) current (social and political) rules (Cavell 1990) along with
an underestimation of the rules does not allow an understanding of the
social contribution to personal changing.

In her turn, as a resentful, outraged, dishonoured and shamed wom-
an, Nora has lost her way and cannot achieve an autonomous change.
Corresponding to Cavell's understanding of Wittgenstein's picture of
thinking, i.e. 'one of moving from being lost to oneself to finding one's
way, a circumstance of personal disorder, a defeat not to be solved, but
to be undone' (Cavell 1990: 21), Nora has lost her way. The existence of
someone who has lost their way also implies that the political depres-
sion could not be overcome by reference to internal principles either,
thereby preventing an autonomous personal changing. Consequently,
both exclusions, from the social and the political self, entail personal

changing to be guided neither by external nor by internal rules (Cavell 1990). It is understood as a miracle.

2. Miracles and private language

Wittgenstein makes explicit references to miracles (Wittgenstein 1965:10–11, 1980: 45; see also Phillips 1993). There is also a philosophical controversy about Wittgenstein's religious belief (Barret 1991; Cook 1988; Diamond 2005; Moore 2005; Nielsen 2001; Malcolm and Winch 1993; Phillips 1993; Winch 1988). In Wittgenstein's philosophy miracles are inherently senseless events and he explicitly says: 'For all I have said by shifting the expression of the miraculous from an expression by means of language to the expression by the existence of language, all I have said is again that we cannot express what we want to express and that all we say about the absolute miraculous remains nonsense' (Wittgenstein 1965:11). And he adds: 'I see now that these nonsensical expressions were not nonsensical because I had not yet found the correct expressions, but that their nonsensicality was their very essence. For all I wanted to do with them was just to go beyond the world and that is to say beyond significant language' (Wittgenstein 1965:11).

However, the fact that miracles are beyond significant language does not mean that they have no place in Wittgenstein's philosophy even if, controversially, some have shown the importance of ineffability in Wittgenstein's philosophy (Conant 2000; Moore and Sullivan 2003; Diamond 2005; Hacker 1986; McGinn 1999; Mulhall 2007).

From the ineffable perspective, the communication of a miraculous inner changing translates the senseless miraculous event into a meaningful one. Nevertheless, since the miracle of changing corresponds to an inner personal event – which is firstly an *inner* personal changing – its *individual communication* can correspond to Wittgenstein's description of private experiences (Wittgenstein 1953: §§ 243–315). In Wittgenstein's philosophy private experiences are also senseless experiences (Wittgenstein 1953: §§ 243–315). However, contrary to miracles, whose essence is to be senseless and ought to remain senseless

(Wittgenstein 1965), private experiences are senseless and ought to be challenged as they result from an absurd private quest for meaning (Wittgenstein 1953: §§ 243–315).

Indeed, in spite of the disagreement over Wittgenstein's arguments about private language, and along with its undoubted complexity and elusiveness (see Baker & Hacker 1985; Canfield 1986; Cavell 1979; Hacker 1990; Malcolm 1977; Mulhall 2007; 1989; Nielsen 2008), by associating the search for meaning as 'a [purely] mental state' (Wittgenstein 1953: § 180), the private approach supposes that agents *immediately* understand the meaning of their thoughts, feelings and acts (Wittgenstein 1953: §§ 246–8, 251, 272, 280, 294, 358). For example, though not neglecting the differences and similarity between Wittgensteinian and Kantian philosophy (see Engel 1970; Glock 1997; Mosser 2008), in the purely theoretical domain there is in Kantian philosophy no immediate access to our inner representations – they are always mediated by the *a priori* forms of sensibility, time and space (Kant 1787 [1968]). Nonetheless, in the practical domain, the exclusion of sensibility entails the *immediate order of human practical reason*. Dispensing with the mediation of space and time, the forms of sensibility, in the practical Kantian domain the inner rational rule commands immediately and unconditionally (Kant 1785/6 [1968], 1788 [1968]). Like any private experience, the immediate and unconditionally commanding rules correspond to private rule-following. This supposes that it suffices to *immediately* grasp the *mental image* of the private human soul (Wittgenstein 1953 § 205; see also § 265) and also to *immediately* behave in accordance with it (or to follow it) (Wittgenstein 1953: § 352; see also McDowell 2002; Kripke 1982).

However, from Wittgenstein's perspective '[it] is *not* a hocus-pocus which can be performed only by the soul' (Wittgenstein 1953: § 454). In reality, when refuting the hypothesis of private feelings and thoughts under the broader criticism of the existence of a private language (Wittgenstein 1953: § 265; see also §§ 56, 271; II, xl, p. 207), Wittgenstein argues that from the private approach to feelings and thoughts agents are not able to *remember the connection between the images and feelings* (e.g. pain). This incapacity is more troubling when viewing the ambiguity of the *aspect* of the image – 'The concept of an aspect is akin to the concept of an image' (Wittgenstein 1953: §§ II, xl,

p. 213). As well as the image of the duck-rabbit (Wittgenstein 1953: §§ II, xl, p. 194) and that of the triangle (Wittgenstein 1953: §§ II, xl, p. 200) – the duck-rabbit can appear *as* duck and rabbit, and the triangle as a triangular hole, a solid or a geometrical drawing – individuals *see* the image of the pointing finger *as if* it were pointing *to two different directions*. The incapacity to *remember the connection between the mental image* and lack of *the* connection *between the immediately grasped mental image* and the forms means that people are not able to distinguish between competing images of thoughts, feelings and acts, nor to identify and understand them. For example, in the private approach to rules, the lack of connection *between the mental image* and the rule's order means that the agents are unable to identify *the regular order* of the rule, i.e. the *regularity of the rule itself* (Wittgenstein (1953: §§ 227; see also §§ 208, 223, 225), or to distinguish the right from the good application, thereby *preventing agents from following the rule*.

Wittgenstein clearly argues that 'The essential thing about private experience is really not that each person possesses his own exemplar, but that nobody knows whether other people also have this or something else' (Wittgenstein 1953: § 272; see also 269). Moreover, besides the intersubjective obscurity of the private experience, the private thing cannot be exteriorizable. Wittgenstein's approach to private experience goes further and sustains that when the intersubjective understanding is lacking, *there is nothing to be exteriorized* – '[The] *private* exhibition is an illusion' (Wittgenstein 1953: § 311; see also §§ 376, 377, 378, 382, 347) or chimera (Wittgenstein 1953: II, xi: 27; see also 81, 97, 177). For example, a private rule prevents others from understanding if and how the agent is following a rule (e.g. is obeying or disobeying the rule) because there is *nothing* to understand. 'It is not possible to obey a rule 'privately' (Wittgenstein 1953: § 202).

Accordingly, accepting that private exhibitions are illusions or chimeras, when instead of going to bed and taking off her fancy clothes, Nora reappears with her clothes signalling her inner changing, *not only could the others understand it as the exteriorization of her inner changing, but also even she would not understand what she was doing, i.e. what was happening inside her.* Nora returning with her outdoor clothes would be an arbitrary and senseless gesture not only for Torvald but also, and mainly, for herself. For instance, the incapacity to *remember*

the connection between the mental image and lack of *the* connection *between the immediately grasped mental image* and its exhibition allows for neither identifying (understanding) nor communicating the personal changing.

It is true that Wittgenstein's reflection on language, mainly on rule-following, explicitly mentions the experience of the *immediate* grasping of the meaning of a word in accord with a use (Wittgenstein 1953: §§ 138, 139, 191, 197) to understand the whole thought in a flash (Wittgenstein 1953: § 319) and to 'grasp the rule' (Wittgenstein 1953: § 155, 179, 184, 321, I, 176). For example, occurring during the application of rules, suddenly grasping the rule corresponds to the awareness of the (right) direction to follow. For instance, someone can be in doubt about the direction to take when suddenly the agent realizes under that 'Now I know how to go on' (Wittgenstein 1953: § 179), i.e. 'I know or understand how to follow the rule'.

Nonetheless, Wittgenstein asserts: "What happens when a man suddenly understands?' – The question is badly framed. If it is a question about the meaning of the expression 'sudden understanding', the answer is not to point to a process that we give this name to. – The question might mean: *what are the tokens* of sudden understanding; what are its characteristic psychical accompaniments?' (Wittgenstein 1953: § 321; see also 155, 179; see Addis 1999; Hunter 1977). Lacking these *tokens*, the non-assertible internal miracle of changing, together with the purely mental and private experience of grasping this change, allows understanding neither the images of thoughts, feelings (Wittgenstein 1953: § 243–315), or rules (Wittgenstein 1953: § 202) nor their significant exteriorization.

As Wittgenstein clearly puts it: 'An 'inner process' stands in need of outward criteria' (Wittgenstein 1953: §§ 243, 272, 280, 294), i.e. without external, intersubjective and public criteria *the inner private image is mute*, and the private image is meaningless. The outward criteria, which address the problem of the communication of our personal (but not private) experiences (Wittgenstein 1953 §§ 380, 378), require a valuable justification for one's own and others' criteria, making individual experiences common (Wittgenstein 1953: § 378). Notwithstanding their oracular formulation in *Philosophical Investigations* (1953) – there are only five allusions to the concept of form of

life (see Wittgenstein 1953: §§ 19, 23, 241, II, i, p. 174 e II, xi, p. 226 e) – Wittgenstein's philosophy of language established the agreement on forms of life as the last criterion allowing common individual experiences. Thus, regardless of the debate on the nature of those forms (e.g. unity (Garver 1990) *vs*. plural (Haller 1988); linguistic (Garver 1990), biological (Hunter 1968) or anthropo-socio-cultural (Cometti 1996)), as criteria, customs, institutions and practices allow us to communicate our internal experiences. Accordingly, personal changing communication depends on criteria.

3. Criteria and the fantasy of the private language

Nevertheless, Cavell also refuses the fantasy of the private language argument – it underlies 'the wish to deny the publicness of language' (Cavell 1979; 351). As a fantasy, a private language 'can be understood as an attempt to account for, and protect, our separateness, our unknowingness, our unwillingness or incapacity either to know or to be known [and its failure] signifies: … that nevertheless there is no end to our separateness. We are endlessly separate, for *no* reason' (Cavell 1979: 369; see also 461). A private language denies the ineradicable *gap* (or *separateness*) between the self and the others, and the self itself (Cavell 1979, 1990; see Mulhall 1994).

From this perspective, which affiliates Wittgenstein's philosophy to the existential philosophical tradition of Pascal (Cavell 1969), Kierkegaard (1969, 1979, 1990) and Camus (1969), and without mentioning the tension relationship between the claims of publicity and expressiveness (Cavell 1979; see Mulhall 1994), Cavell stresses Wittgenstein's appeals to criteria. These assess the public value of people's words towards others and themselves (Cavell 1969, 1979, 1990), allowing them to overcome the denial of separateness. Therefore, although people are irremissibly separated, they can share the public content of their words. The appeal to criteria corresponds then to a claim of community because we have to prove to others that we are not wrong, or that our personal convictions neither *isolate us from the others, nor from ourselves*

(Cavell 1979, 1990). The appeal to criteria is also a claim of rationality because the 'intelligibility of others to himself or herself, and of himself or herself to others' (Cavell 1990: xxxi; see also 1979) is based on a rational speaking. Also, to be rational implies the ability to speak for oneself from a requirement of mutuality. For example, when Nora explicitly decided to abandon her husband she was challenging nineteenth-century social and political institutions and customs or criteria that defined the 'right or the good woman' as a woman who sacrifices herself for her husband (Langås 2005; Templeton 1989; Yuehua 2009). Nora's final decision later referred to a new criterion (e.g. the right and the good woman is a free person), which allows *justifying to herself and to society* why she cannot maintain support for an institution that fostered her irresponsibility and dependence.

Although Cavell is seriously committed to the claim for community since the search for criteria (reasons or explanations) comes to an end – Cavell's (1979, 1990) exegesis of Wittgenstein emphasizes the fact that 'Explanations come to an end somewhere.' (Wittgenstein 1953: §1; see Cavell 1979, 1990) – human communication towards others and oneself always risks disagreement, incomprehension, obscurity, scepticism and absurdity (e.g. 'I fail to manifest our criteria accurately or that the other fails to read them accurately' (Cavell 1979: 477)). Indeed, although a private language is a fantasy, refusing individuals' separateness, and criteria challenge this fact, allowing the publicity of experiences, it does not mean that criteria fulfil absolutely the gap between persons, i.e. that the publicity of 'images' is exempt from disagreements, misunderstandings and conflicts (e.g. the conflict between persons' claim of expressiveness, the claim of community and rationality, the inner and the outer, the self and the others (Cavell 1979).

Moreover, in spite of the complex relationships between the personal inner self and the outer others, and even the complex relationship between the self itself (Cavell 1979, 1990; see Mulhall 1994, 2007) or the double opaqueness – towards others and towards oneself – the underestimation of rules does not allow understanding the social contribution to personal changing, as stated above. Indeed, viewing that the claim of community and reason is seen as a radical subjective experience for publishing inner thoughts or feelings – 'the soul's investigation of itself' (Cavell 1979: 15) – Cavell dissociates rules from personal explanations or reasons in this

investigation. 'The soul's investigation of itself' would be challenged by a guiding rule, mainly when someone lost their way. As also stated above, seen as *a priori* orders (Cavell 1979, 1990), the consequences of which are 'confined and scored' (Cavell 1990: 115), rules prevent understanding the quest for the personal recovery of voice.

Nevertheless, admitting that rules are not forcibly *a priori* (Wittgenstein 1953 §§ 68, 84, 94), one may not dissociate personal changing from rules, even if abnormal (moral) rules. Moreover, besides the fact that there is no criterion for defining criteria (Wolgast 1964), the unavoidable criteria are also not *immediately* and *internally* followed by individuals. *Otherwise we would translate the structure of the private experience into the relationship between agents and forms of life*. For that reason, even if, as Cavell (1979, 1990) has clearly asserted, criteria are in Wittgenstein's philosophy of language the *judges* of rules' application (Wittgenstein 1953: §§ 56, 185, 692), challenging the private experience of rules – application of rules requires criteria that offer reasons for choosing them and for establishing their public validity (Wellman 1962) – the search for criteria is not dissociated from the establishment of rules. More accurately, the common agreement or understanding on forms of life, which distinguishes the wrong from the right rules' application, cannot be verified without that application. Thus, beyond any concrete rule application in a concrete case (e.g. the rule of remaining in a marriage that requires sacrifice for women in the nineteenth century in Europe), political criteria (e.g. equal liberty) can be seen as a pious vote or an empty concept. We argue then that persons cannot communicate (their change) without a (common and agreed) rule (Brown 1988; Johannessen 1988; Tymoczko 1984) following criteria (Canfield 1974).

4. The social scope of personal changing: the role of criteria and rules

Seen as socially transmitted and customary normative injunctions or immanently normative dispositions, rules are prescriptive and are understood by analogy with the obedience of an order (Wittgenstein 1953

§§ 202, 228, 230) in spite of the distinction between right and wrong rule, and since rules are differently applied, one may distinguish between abnormal and normal rules (e.g. the student rule of natural numbers written as integers 1, 0, 3, 2, 5, 4 (Wittgenstein 1953 § 143; see also 141, 142)). Following logical consistency, this abnormal rule is also based on a shared form of life. *Otherwise it would not be a rule*. Thus, neither does the abnormality of rules forcibly entail the refusal of any rule, nor is an abnormal *ambiguous rule* a senseless (private) rule. For example, albeit in the context of the nineteenth century, Nora's rule to leave her marriage can be seen as an abnormal rule. This rule is not forcibly senseless. The fact that in the nineteenth century political legislation could hardly include a woman abandoning her husband did not refute that human history was full of similar decisions, and there was still a shared moral agreement on human liberty and equality (Locke 1679 [1960]; Kant 1787 [1968]).

Furthermore, in spite of the different meaning of criteria in Wittgenstein's philosophy (Albritton1959; Canfield 1974; Cavell 1979, 1990), one may not dismiss the fact that before *A Doll's House* was written by Ibsen in 1879, political philosophy had already stressed the *universal value* of human dignity (Kant 1787 [1968]) and the universality of human rights (Locke 1679 [1960]). So, in spite of explicit discriminatory statements, for example against women (e.g. Kant's (1784 [1968]) judgment on the avoidance of liberty by the 'fair sex'), its starting principles (or criteria) were, and are not, compatible with human discrimination (e.g. gender or discrimination). For that reason, although current political society was still not ordered under there being equal political rights for every human being, it does not mean that there is not *a shared moral understanding of the refusal of women's treatment as foolish children*. The abnormal rule did not lack a criterion, i.e. a reason or explanation for its use.

However, the political (and ethical) value of the rule 'leave the cage of dolls' depends on different criteria. For instance, under the criterion 'the right or the good woman is a woman who sacrifices herself for her husband', 'leave the cage of dolls' can be seen as an unprincipled, irresponsible, egoist and immoral rule. Conversely, under the criterion 'the right or the good woman is a free and autonomous woman', 'leave the cage of dolls' can be seen as a fair rule. Similarly, the rule 'remain

in the cage of dolls' can be seen under 'the right or the good woman is a free and autonomous woman' as an unprincipled, irresponsible, egoistic and immoral rule and a moral one under the criterion 'the right or the good woman is a woman who sacrifices herself for her husband'. For that reason, besides constituting what Cavell names the argument of the ordinary (Cavell 1979, 1990; see Mulhall 1994, 2007), Nora's repudiation and acceptance of criteria always entails the choice of rules. Choice of rules shows that they are not forcibly *a priori* limiting the personal free quest for a personal voice. Consequently, rules do not limit personal liberty nor can personal change following criteria be understandable or exteriorizable beyond concrete, even if abnormal, (moral) rules.

Moreover, since 'The more abnormal the case, the more doubtful it becomes what we are to say' (Wittgenstein 1952 § 142), and even Nora's decision has almost become uncontroversial – at least in certain places in the world – one may not neglect the 'amount' of doubt and misunderstanding surrounding someone who, in the nineteenth century, decided 'to leave the cage of dolls' and to live according to the Kantian principle (or criteria) of an equal and free human being (Kant 1785/6 [1968]). Nonetheless, and in spite of the social frame of rules, the 'amount' of doubt and misunderstanding does not dismiss personal responsibility. Applying rules *is in Wittgenstein's philosophy an intrinsically personal act* – 'Don't always think that you read off what you say from the facts; that you portray these in words according to rules. For even so you would have to apply the rule in the particular case without guidance' (Wittgenstein 1953 § 292).

Applying a rule without guidance can entail applying an abnormal rule under a criterion. For example, the rule 'leave the cage of dolls' could be understood as an abnormal rule under a moral criterion for women i.e. 'the right or the good woman is a free and autonomous woman'. Therefore, from our perspective, there is no conflict between politically depressed and deprived persons lacking any political rule, and cruel and despotic persons following unfair political rules. However, there is the conflict between two rules (e.g. the rule of fairness and the rule of inequality), which one may say from Wittgenstein's point of view is the conflict between a normal and abnormal rule.

It is true that Wittgenstein asserts that 'Not only rules, but also examples are needed for establishing a practice. Our rules leave loop-holes

open, and the practice has to speak for itself.' (Wittgenstein 1969 §139). For example, when children are learning a practice, since they do not know the names yet, teachers ought to make use of examples (Wittgenstein 1953 §§135, 208, 210). But the fact that the learner can dispense with explicit references to rules does not mean that practices are anarchic or anomic (Wittgenstein 1969 §139) nor that rules (and criteria) are always explicitly formulated (Canfield 1974).

In sum, since rules are not prior to their application, the process of the inner changing is encompassed by the *establishment of a rule* under a criterion (e.g. women ought to leave their marriages when their husbands do not treat them as ends in themselves (Kant 1785/6 [1968]). Accordingly, we *cannot change* without explicit or implicit (Wittgenstein 1969 §95) rules framed by criteria, i.e. any personal change results from the *public co-evaluation under a guiding rule*.

Conclusions

Even if we acknowledge the contribution of Cavell's exegesis of Wittgenstein to expose the existential issues underlying Wittgenstein's philosophy, we stress the social and public frame of any search for personal changing. Inspired by Wittgenstein's philosophy of language, we argued that any personal changing requires the mediation of political rules, following criteria, even though these rules are in the end abnormal rules. We also argued that although criteria do certify the public content of those rules, criteria and rules are not *immediately* followed by individuals. Customs are embedded in *societies' rules* and institutions are systems of rules (Rawls 1971).

From these premises we concluded that individual changing results from *a guiding public rule* even if it is an abnormal public rule. *We cannot change* without explicit or implicit (Wittgenstein 1969 § 95) rules.

Future research could explain the changing of political rules under the Wittgensteinian concept of 'seeing-as'. Future research could also reassess the understanding of a personal judgment without rules under Kant's (1790 [1968]) theory of taste judgment. Explicitly articulated by

Cavell (1990), the Kantian approach to rules and judgments can also offer sound arguments to the endless claim for justice.

Finally, future research could also clarify the political consequences of erasing the tragic dimension of politics and further the already existing reflection on the relationship between the philosophical thought of Wittgenstein and Nietzsche (e.g. Bowls 2003; Cavell 1990; Williams 1993). Instead of a political consensus on rules and criteria, the tragic dimension of our political existence warns us about the unavoidable gap between an ideal fair society and the current political injustices (Cavell 1967, 1997) and our political responsibility to demand a decent and fair human society.

References

Addis, Mark (1999). *Wittgenstein: Making Sense of Other Minds*. Burlington: Ashgate.

Albritton, Rogers (1959). 'On Wittgenstein's Use of the Term "Criterion"'. *Journal of Philosophy* 56 (22): 845–857.

Baker, Gordon, & Peter Hacker (1985). *Wittgenstein. Rules, Grammar and Necessity.* In *An Analytical Commentary on Philosophical Investigations*, vol. 2: Oxford: Blackwell.

Barret, Cyril (1991). *Wittgenstein on Ethics and Religious Belief*. Oxford: Blackwell.

Bernstein, Charles (1981). 'Reading Cavell Reading Wittgenstein'. *Boundary* 2, 9(2): 295–306.

Bloor, David (1997). *Wittgenstein, Rules and Institutions*. London and New York: Routledge.

Bowls, M. J. (2003). 'The Practice of Meaning in Nietzsche and Wittgenstein'. *The Journal of Nietzsche Studies* 26: 12–24.

Brown, Harold (1988). *Rationality*. New York: Routledge.

Budd, Malcolm (1984). 'Wittgenstein on Meaning, Interpretation and Rules'. *Synthese* 58: 303–323.

Canfield, John (1974). 'Criteria and Rules of Language'. *The Philosophical Review* 83(1): 70–87.

Canfield, John (ed.) (1986). *The Philosophy of Wittgenstein, Volume 9: The Private Language Argument*. New York: Garland.

Cavell, Stanley (1969). *Must We Mean What We Say*? Cambridge: Cambridge University Press.

Cavell, Stanley (1979). *The Claim of Reason*. Oxford: Oxford University Press.

Cavell, Stanley (1990). *Conditions Handsome and Unhandsome: The Constitution of Emersonian Perfectionism*. Chicago: The University of Chicago Press.

Cometti, Jean Pierre (1996). *Philosopher avec Wittgenstein*. Paris: Presses Universitaires de France.

Conant, James (2000). 'Elucidation and Nonsense in Frege and Early Wittgenstein', in Alice Crary and Rupert Read (eds.), *The New Wittgenstein*. London: Routledge: 174–217

Conant, James (2005). 'Stanley Cavell's Wittgenstein'. *Harvard Review of Philosophy* 13(1): 51–65.

Cook, John (1988). 'Wittgenstein and Religious Belief'. *Philosophy* 63 (246): 427–452.

Cook, John (1965). 'Wittgenstein on Privacy'. *The Philosophical Review* 84(3): 281–314.

Diamond, Cora (2005). 'Wittgenstein on Religious Belief: The Gulfs Between Us', in Dewi Philips and Mario Van der Ruhr (eds.), *Religion and Wittgenstein's Legacy*. Burlington: Ashgate: 99–137.

Eldridge, Richard (2003). 'Wittgenstein and the Conversation of Justice', in Cressida Heyes (ed.), *The Grammar Politics*: *Wittgenstein and Political Philosophy*. Ithaca: Cornell University Press: 117–128.

Engel, Morris (1970). 'Wittgenstein and Kant'. *Philosophy and Phenomenological Research* 30 (4): 483–513.

Fleming, Richard, & Michael Payne (eds.) (1989). *The Senses of Stanley Cavell*. Lewisburg: Bucknell University Press.

Fogelin, Robert (1976). *Wittgenstein*. London: Routledge.

Garver, Newton (1990). 'Form of Life in Wittgenstein's Later Work'. *Dialectica* 44: 175–201.

Glock, Hans-Johann. 1997. Kant and Wittgenstein: Philosophy, necessity and representation. *International Journal of Philosophical Studies*, 5(2): 285–305.

Hacker, Peter (1986). *Insight and Illusion: Themes in the Philosophy of Wittgenstein*. Oxford: Oxford University Press.

_____ (1990). *Wittgenstein: Meaning and Mind, Volume 3 of an Analytical Commentary on the Philosophical Investigations*. Oxford: Blackwell.

Haller, Rudolf (1988). *Questions on Wittgenstein*. London: Routledge.

Hammer, Espen (2002). *Stanley Cavell: Skepticism, Subjectivity, and the Ordinary*. Cambridge: Polity Press.

Holt, Robin (1997). *Wittgenstein, Politics and Human Rights*. London: Routledge.

Hunter, John (1968). 'Forms of Life in Wittgenstein's *Philosophical Investigations*'. *American Philosophical Quarterly* 5: 233–243.

_____ (1977). 'Wittgenstein on inner processes and outward criteria'. *Canadian Journal of Philosophy* 7: 805–817.

Ibsen, Henrik (1879 [1981]). *A Doll's House*. in *Four Major Plays*. Trans. James McFarlane and Jens Arup. Oxford: Oxford University Press.

Johannessen, Renate (1988). 'Rule following and tacit knowledge'. *AI & Society* 2(4): 287–302.

Kant, Immanuel (1784 [1968]). 'Beantwortung der Frage: Was ist Aufklärung?', in *Immanuel Kant Werksausgabe,* edited by Wilhelm Weischedel, Band IX: 61–91. Frankfurt am Main: Suhrkamp.

_____ (1785/6 [1968]). 'Grundlegung zur Metaphysik der Sitten', in *Immanuel Kant Werksausgabe,* edited by Wilhelm Weischedel, Band IV: 11–102. Frankfurt am Main: Suhrkamp.

_____ (1787 [1968]). 'Kritik der Reinen Vernunft', in *Immanuel Kant Werksausgabe*, edited by Wilhelm Weischedel, Band III–IV: Frankfurt am Main: Suhrkamp.

_____ (1790 [1968]). 'Kritik der Urteilskraf', in *Immanuel Kant Werksausgabe,* edited by Wilhelm Weischedel, Band X: Frankfurt am Main: Suhrkamp.

Kenny, Anthony (1984). *The Legacy of Wittgenstein*. Oxford: Blackwell.

Kripke, Saul (1982). *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.

Langås, Unni (2005). 'What Did Nora Do? Thinking Gender with a Doll's House'. *Ibsen Studies* 5(2): 148–171.

Locke, John (1679 [1960]). 'Two Treatises of Government', in *Cambridge Texts in the History of Political Thought*, edited by Peter Laslett. Cambridge: Cambridge University Press.

Malcolm, Norman (1977). *Thought and Knowledge*. Ithaca: Cornell University Press.

____ (1989). 'Wittgenstein on language and rules'. *Philosophy* 64: 5–28.

Malcolm, Norman, & Peter Winch (1993). *Wittgenstein: A Religious Point of View?* edited with a response by Peter Winch. Ithaca, NY: Cornell University Press.

McDowell, John (1984). 'Wittgenstein on Following a Rule'. *Synthese* 58 (1984): 325–63.

McGinn, Marie (1999). 'Between Metaphysics and Nonsense: Elucidation in Wittgensten's *Tractactus*'. *The Philosophical Quarterly* 49: 491–513.

____ (2004). 'The everyday alternative to scepticism: Cavell and Wittgenstein on other minds', in Denis McManus(ed.) *Wittgenstein and Scepticism*, London: Routledge: 240–258.

Moore, Gareth (2005). 'Wittgenstein's English Parson: Some Reflections on the Reception of Wittgenstein in the Philosophy of Religion', Dewi Z. Phillips and Mario Von der Ruhr (eds.), *Religion and Wittgenstein's Legacy*. Burlington: Ashgate: 209–228.

Moore, A.W., & Peter Sullivan (2003). 'Ineffability and Nonsense'. *Aristotelian Society Supplementary Volume* 77 (1): 169–193.

Mosser, Kurt (2008). 'Kant and Wittgenstein: Common Sense, Therapy, and the Critical Philosophy'. *Philosophia* 31(1): 1–20.

Mulhall, Stephen (1994). *Stanley Cavell: Philosophy's Recounting on the Ordinary*. Oxford: Oxford University Press.

Mulhall, Stephen (2007). *Wittgenstein's Private Language: Grammar, Nonsense, and Imagination in Philosophical Investigations, §§ 243–315*. Oxford: Clarendon Press.

Nielsen, Kai (2001). 'Wittgenstein and Wittgensteinians on Religion' in Robert L. Arrington and Mark Addis (eds.), *Wittgenstein and Philosophy of Religion*. London: Routledge: 137–166.

Nielsen, Kai (2008). *The Evolution of the Private Language Argument*. Burlington: Ashgate.

Phillips, Dewi, & Peter Winch (eds) (1989). *Wittgenstein: Attention to Particulars*. Houndmills: Macmillan.

Phillips, Dewi (1993). *Wittgenstein and Religion*. New York: St. Martin's Press.

Pitkin, Hanna (1972). *Wittgenstein and Justice: On the Significance of Ludwig Wittgenstein for Social and Political Thought*. Berkeley: University of California Press.

Rawls, John (1971). *A Theory of Justice*. Oxford: Oxford University Press.

Taylor, Charles (1992). 'To Follow a Rule'*,* in Mette Hjort (ed.), *Rules and Conventions: Literature, Philosophy, Social Theory.* Baltimore: John Hopkins University: 167–185.

Temelini, Michael (2015). *Wittgenstein and Politics*. Toronto: University of Toronto.

Templeton, Joan (1989). 'The Doll House Backlash: Criticism, Feminism, and Ibsen'. *PMLA* 104 (1): 28–40.

_____ (1997). *Ibsen's Women*. Cambridge: Cambridge University Press.

Tully, James (2003). 'Wittgenstein and Political Philosophy: Understanding Practices of Critical Reflection' in Cressida Heyes (ed.), *The Grammar Politics: Wittgenstein and Political Philosophy.* Ithaca: Cornell University Press: 17- 42.

Tymoczko, Thomas (1984). 'Gödel, Wittgenstein and the Nature of Mathematical Knowledge'. *Proceedings of the Biennial Meeting of the Philosophy of Science Association* (2): *Symposia and Invited Papers*: 449–468.

Wellman, Carl (1962). 'Wittgenstein's conception of a criterion'. *Philosophical Review* 71: 433: 447.

Williams, Bernard (1993). 'Nietzsche's Minimalist Moral Psychology'. *European Journal of Philosophy* 1(1): 4–14.

Winch, Peter (1987). *Trying to Make Sense*. Oxford: Basil Blackwell.

Wittgenstein, Ludwig (1953). *Philosophical Investigations*. 2nd Edition. Translated by E. M. Anscombe. Cambridge: Blackwell Publishers.

_____ (1965). 'Wittgenstein Lecture on Ethics'. *Philosophical Review* 74: 311–339.

_____ (1966). *Lecture and Conversations on Aesthetics, Psychology, and Religion*, edited by Cyril Barrett. Berkeley: University of California Press.

_____ (1969). *On Certainty*. Translated by G. E. M. Anscombe and G.H. von Wright. Oxford: Blackwell.

_____ (1980). *Culture and Value*, edited by. G. H. Von Wright and H. Nyman, translated by Peter Winch. Oxford: Blackwell.

Wolgast, Eike (1964). 'Wittgenstein and criteria'. *Inquiry* 7: 348: 366.

Yuehua, Guo (2009). 'Gender Struggle over Ideological Power in Ibsen's *A Doll's House*'. *Canadian Social Science* 5 (1): 79–87.

The Ethico-Political Dimension of Foucault's Thought

Marta Faustino

1. Introduction

Influential scholarship has criticized Foucault for not offering a positive account of ethics. More moderate readings find it difficult to reconcile his earlier views on power, society and subjectivity with the concern with ethics and politics expressed in the final period of his thought. Together with his analyses in the 1970s of the mechanisms of social control and normalization in modern societies – analyses which seem to leave little room for any consistent idea of human freedom, autonomy or sovereignty – his criticisms of humanism, rationality and universalism have frequently been used to build an image of Foucault as a quasi-nihilistic thinker in whose view no satisfying ethics can ever be realized. And yet, if there is something that characterizes Foucault's work as a whole – and not only his later writings, in which ethical questions are most evidently present – it is his constant challenging of settled values and beliefs about ourselves and the social world surrounding us. As Richard Lynch puts it, 'Foucault challenges, questions, criticizes, and 'dereifies' social norms, structures, and institutions; he calls into question our presuppositions about society and individuals, including ourselves. Foucault disconcerts us in much the same way as Socrates disconcerted his fellow Athenians in the agora 2400 years ago' (Lynch 2016: 3). In Foucault's own view, this critical function, which lies at the heart of his understanding and practice of philosophy, indeed derives from the Socratic injunction 'Take care of yourself', which he interprets as a call to '[m]ake freedom your

foundation, through the mastery of yourself' (*ECS*: 301)[1]. Foucault
remains one of the few philosophers of his time to genuinely incorpo-
rate this injunction and we find resonances of it throughout his entire
work. In fact, if we view ethics as concerning not obedience to certain
codes of behaviour or compliance with specific systems of values and
beliefs, but rather, in Foucauldian terms, the relationship between the
self and itself and the way it constitutes itself as the subject of its own
actions, then Foucault's thought is ethical through and through, and
his late focus on the ethics of care of the self is nothing but the apex
of his entire philosophical endeavour.

This paper aims to contribute to a recent trend in Foucault studies
that is currently bringing to light the deep ethical and political dimen-
sions of Foucault's thought.[2] Although I focus on his works from the
1980s – namely the second and third volumes of *The History of Sex-
uality* (*The Use of Pleasure* and *The Care of the Self*) and the lecture
course on *The Hermeneutics of the Subject*, together with interviews
and short essays from the same period – my aim is to highlight top-
ics that, while explicitly thematized in the last period of his thought,
also shed light on the ethical and political significance of his phil-
osophical project as a whole, demonstrating the continuity of these
later writings with his earlier works from the 1960s and 70s. Through
analysis of Foucault's emphasis on the technologies of the self as a
possible means of mastery over oneself (2), his encouragement of the
reestablishment of an ethics of the self as a practice of freedom and
liberation (3), and his ethical ideal of an aesthetics of existence (4),
I hope to make apparent both Foucault's lifelong ethical and political
engagement with the challenges of his time and his use of philosophy
as a weapon of resistance and promotion of individual governance,
freedom and liberation (5).

---

1   Abbreviations are used for references to works of Foucault and Nietzsche. Cf.
     references at the end of the article.
2   See, for example, O'Leary (2002), Harrer (2005), Kelly (2013), Simons (2013),
     Koopman (2013), Smith (2015), and Lynch (2016).

2. The Technologies of the Self

In a seminar delivered at the University of Vermont in 1982, Foucault identifies four types of technologies of practical reason through which individuals acquire self-knowledge and self-understanding while at the same time submitting to a certain form of domination that transforms their conduct and attitudes: technologies of production (which allow individuals to produce things); technologies of sign systems (which allow individuals to use meaningful signs and symbols); technologies of power (which determine the individual's behaviour, submitting and objectifying it to certain ends); and technologies of the self (which allow individuals to produce certain changes in themselves in order to realize individual ends) (cf. *TS*: 225). Even though these technologies 'hardly ever function separately', and Foucault's declared aim is 'to show their specific nature and their constant interaction', his focus has mainly been the last two in an attempt to build 'a history of the organization of knowledge with respect to both domination and the self' (*TS*: 225). In the same seminar, however, Foucault himself recognizes that he might have overstressed the impact of technologies of power and domination in modern societies, resulting in an unbalanced, asymmetrical and reductive view on power (*idem*). In the late period of his writing, he seeks precisely to re-establish this balance by bringing to the fore 'the interaction between oneself and others' and 'the technologies of individual domination, in the mode of action that an individual exercises upon himself by means of the technologies of the self' (*TS*: 225).

These 'technologies of the self', a form of training and modification of individuals which Foucault explicitly opposes to the technologies of power or domination that occupied him for most of his productive life, are described as techniques that 'permit individuals to effect by their own means, or with the help of others, a certain number of operations on their own bodies and souls, thoughts, conduct, and way of being, so as to transform themselves in order to attain a certain state of happiness, purity, wisdom, perfection, or immortality' (*TS*: 225; cf. *UP*: 10–11). Thus, even though it does not correspond to a completely new interest for Foucault, instead representing a shift in focus to a problem that concerned him 'for more than twenty-five years' (*TS*: 224),

this late turn to the technologies of the self does indeed add complexity to his accounts of social order and subjectivity, dispelling possible (mis)readings of his thought. More concretely, Foucault's criticism of disciplinary power and social control gave way to a renewed concern with ethics as 'care of the self' and with the means by which people can, despite those oppressive and disciplining powers, offer resistance by governing, creating and shaping themselves. Considering the whole picture, subjects are thus far from being 'formless, conditionable creatures', to use Honneth's expression (Honneth 1991: 199), dominated by trans-subjective systems of micro-powers that oppress, control and discipline them, but are rather individuals who cope (or at least can cope) with those oppressing external forces in order to fashion their lives and selves.

The Greek and Latin tradition of *epimeleia heautou* (care of the self), to which Foucault devotes a major part of his late work, is a paradigmatic example of the full realization and embodiment of these 'technologies of the self' as an 'art of living' (*tekhnē tou biou*), or a 'government of oneself' (cf. *UP*; *CS*; *HS*). Particularly striking to Foucault in this tradition of thought is how ethics was conceived in total independence of any kind of institutional, juridical, authoritarian or disciplinary structure (*GE*: 260) and was thus unassociated with any attempt to normalize the population (*GE*: 254). Stoic ethics, for example, was neither universally imposed nor applicable to all: 'it was not a question of giving a pattern of behavior for everybody. It was a personal choice for a small elite' (*GE*: 254).[3] The injunction to take care of oneself was, to be sure, a commandment that was fully ingrained in Greco-Roman culture and thus cannot be conceived as a pure individual initiative or creation, free of any kind of external influence or power. In this sense, Foucault stresses the extent to which practices of the self were models that were 'proposed, suggested, imposed upon [the individual] by his culture, his society, and his social group' (*ECS*: 291). What is at stake here is therefore not the opposition between purely autonomous self-constitution and a coercive heteronomous model of subjectivation, but rather the

---

3   Cf. also *GE* (271): 'In antiquity, this work on the self with its attendant austerity is not imposed on the individual by means of civil law or religious obligation, but it is a choice about existence made by the individual. People decide for themselves whether or not to care for themselves'.

significant distinction between a completely *passive subject*, who is simply the product of micro external powers exerted over him, and a relatively *active subject* who, among and despite those external powers, can still emerge as an ethical subject and be actively engaged in his own self-constitution.

The passivity and activity of the subject are not contrasting elements; they are complementary perspectives in Foucault's later conception of subjectivity since both refer to different forms of constitution (or 'objectification') of the subject.[4] In this sense, it seems legitimate to argue that Foucault's so-called 'ethical turn' constitutes a development and complexification of – rather than a break with – his earlier views on subjectivity.[5] Throughout his work, Foucault repeatedly rejects, probably under the influence of Nietzsche,[6] the idea of the subject as a substance (which is completely different from rejecting the concept of the subject altogether): the subject is not a substance but a form, and this form is fluid, flexible and changeable, 'not primarily or always identical to itself' (*ETS*: 290). Thus, subjects are indeed 'formless, conditionable creatures', but in a fundamentally different sense from the idea expressed by Honneth: they are formless and conditionable because they do not have a fixed, eternal, unchangeable, monadic essence and are thus subject to modification or molding by everything with which they are in contact and to which they are related – be it an external coercive structure or a model they voluntarily apply to themselves.[7] In this sense, the ascetic practices that characterize the ancient culture of the self,

---

4   See *SP* (326–327), where Foucault describes the 'three modes of objectification that transform human beings into subjects', the last of which is 'the way a human being turns him- or herself into a subject'.

5   This is Sebastian Harrer's position. He argues that 'there is a 'conceptual continuity', rather than a break, between Foucault's earlier works on normalizing power, and his later works, on ethical self-constitution' since "fabrication' and 'self-constitution' are but two aspects of subjectivation' (Harrer 2005: 75). See also Kelly (2013), Simons (2013), Koopman (2013) and Smith (2015: 145).

6   On Nietzsche's conception of subjectivity, see Constâncio, Branco & Ryan (2015).

7   Cf. *AE* (50–51): '… the subject is constituted through practices of subjection, or, in a more autonomous way, through practices of liberation, of liberty, as in Antiquity, on the basis, of course, of a number of rules, styles, inventions to be found in the cultural environment'.

for example, are also disciplinary practices; the fundamental difference lies in the fact that it is the subject himself who voluntarily decides to submit himself to them in an attempt to govern himself, or to become master of himself.[8]

Foucault's later conception of ethics and subjectivity thus implies the rejection of both the model of a pure autonomous constitution of the subject, independent of any power relations, and the model of a completely heteronomous subject, fully determined by them. As Daniel Smith puts it, 'ethical practices are of course saturated with power relations … But it is not true, either, that the subject is *totally* determined by external influences; not in the sense that there is always a point of absolute freedom hidden deep within us that cannot be completely subjected to power, but rather that the self, in addition to being influenced by outside forces, also affects itself' (Smith 2015: 145). Foucault's late acknowledgment of the possibility of a power relation of the *self over the self* (that is, the possibility that individuals become masters of their own processes of subjectification) rules out 'the reductionist idea of a one-sided rule of force' that Honneth, among others, too simplistically ascribes to his social theory.[9]

3. The Ethics of the Self as a Practice of Freedom

In one of his last interviews, Foucault explicitly claims that freedom is the ontological condition of ethics or, in another formulation, that ethics is the conscious practice of freedom (cf. *ECS*: 284). This statement might seem paradoxical when one considers one of the most persistent criticisms of his thought – namely, the apparent incompatibility between his supposed conception of power and any idea of human freedom or

---

8   On this topic, see Harrer, (2005).
9   For a discussion of Honneth's criticism of Foucault (and Nietzsche), see Constâncio/Faustino (forthcoming). See also Kelly (2013) and Ingram (2006) for a defence of Foucault against Habermas's criticism, which in this respect is similar to Honneth's.

autonomy. This widespread criticism is, however, based on a common misinterpretation of his notion of power, which, in the same interview from 1984, Foucault laments and tries to correct:

> … the claim that 'you see power everywhere, thus there is no room for freedom' seems to me absolutely inadequate. The idea that power is a system of domination that controls everything and leaves no room for freedom cannot be attributed to me. (*ECS*: 293)

Such a view cannot be attributed to Foucault because, once again following the track laid by Nietzsche, his conception of power necessarily involves the ideas of *relation*, *resistance* and thus also *freedom*. Far from being incompatible with the omnipresence of power, freedom is rather the very condition of the possibility of power – that is to say, of *relations of power*.[10] For Foucault, just as for Nietzsche, every human relationship, be it personal, amorous, sexual, political, social, economic, institutional or educational, is a relation of power in the sense that it always involves one person's trying to dominate and control the conduct of the other. But since this power is always *relational* and the same aim can be found in the other party, power relations are necessarily mobile, modifiable, unstable and reversible. The possibility of change and reversion is the very essence of a power relation: 'If one of them were completely at the other's disposal and became his thing, an object on which he could wreak boundless and limitless violence, there wouldn't be any relations of power' (*ECS*: 292). This means, first of all, that power relations require at least a certain degree of freedom on both sides.[11] This applies even to situations where the power equilibrium is clearly unbalanced and asymmetrical: power can only be exerted if the dominated party still has the slightest ability to revert the situation – for example, at the limit, by killing himself or the

---

10   Cf. *ECS* (291): 'I scarcely use the word *power*, and if I use it on occasion it is simply as shorthand for the expression I generally use: *relations of power*.' See also *SP* (340–342).

11   In this sense, Foucault explicitly inverts the position that is often ascribed to him: 'I refuse to reply to the question I am sometimes asked: 'But if power is everywhere, there is no freedom.' I answer that if there are relations of power in every social field, this is because there is freedom everywhere' (*ECS*: 292).

domineering party. Secondly, however, this also implies that any power relation necessarily involves resistance: 'in power relations there is necessarily the possibility of resistance because if there were no possibility of resistance (of violent resistance, flight, deception, strategies capable of reversing the situation), there would be no power relations at all' (*ECS*: 292). Agonism thus lies at the very heart of Foucault's conception of power; the entire net of power relations is characterized by 'permanent provocation', which is at the same time 'mutual incitement and struggle' (*SP*: 342). There are, of course, cases of states of domination which remain 'blocked' or 'frozen' for a long period of time, where the degree of freedom of the oppressed person(s) is extremely limited and narrow, but even in such rare cases, resistance is there, even if only as a possibility: 'In such cases of domination, be they economic, social, institutional, or sexual, the problem is knowing where resistance will develop' (*ECS*: 292).[12]

Foucault distinguishes practices of freedom from practices of liberation, which might be needed to make practices of freedom possible – in cases of strong political, cultural or social states of domination, for example – but are not in themselves sufficient to produce acceptable, satisfying lives. Thus, the ethical question is not 'what do I have to liberate myself from?' but rather 'how can I practice my freedom?', or 'what can I do with my available freedom?' (cf. *GE*: 276; *ECS*: 282– 284). The Greco-Roman tradition of the care of the self is once again Foucault's model in this regard. For the Greeks, freedom was equivalent to non-slavery: to be free meant not being the slave of another person, city, ruler or one's passions and appetites. The care of the self was a particular means of properly practising freedom – that is, of knowing oneself, shaping oneself, overcoming oneself and becoming master of oneself: 'with respect to oneself one establishes a certain relationship of domination, of mastery, which was called *arkhē*, or power, command' (*ECS*: 286–287).[13] In this sense, the particular ēthos of the care of the

---

12   For an excellent account of Foucault's conceptions of power, freedom and resistance, see Simons (2013). See also Allen (2013) for a retrospective view on Foucault's notions of power and subjectivity.

13   Note that concern for oneself does not imply neglecting others, but merely involves acknowledging the ethical priority of caring for oneself over caring for others. Taking care of oneself is required if one is to take good care of others and

self was itself a form of conversion of power, or 'a way of limiting and controlling power' (*ECS*: 288), a form of resistance to the dangers not only of slavery but also of abuse of power. It is in this sense that, inspired by the ethics of concern for the self, Foucault rejects both the possibility of a society freed from all oppressive or repressive structures and the impossibility of active resistance to them:

> The idea that there could exist a state of communication that would allow games of truth to circulate freely, without any constraints or coercive effects, seems utopian to me. […] I do not think that a society can exist without power relations, if by that we mean the strategies by which individuals try to direct and control the conduct of others. The problem, then, is not to try to dissolve them in the utopia of completely transparent communication but to acquire the rules of law, the management techniques, and also the ēthos, the practice of the self, that will allow us to play these games of power with as little domination as possible. (*ECS*: 298)

Thus, contrary to authors like Sartre, Foucault does not view power as evil in itself; as a necessary component of any relationship, it is also a condition of the possibility of things like love and passion, education and transmission of knowledge, parenting, medical practices, and so on (cf. *ECS*: 298–299). The problem is the possible abuse of power and authority, especially by the state and its totalizing institutions, which, according to Foucault, can only be controlled through new forms of subjectivity and the renewal of an ethics based on practices of the self and freedom. This, in his view, is the crucial point of struggles for political rights and against abusive forms of government (cf. *ECS*: 299). Because in Foucault's view our current challenge is not simply to 'discover what we are' but also to '*refuse* what we are', we need, in his words, 'to imagine and to build up what we could be to get rid of this kind of political 'double bind', which is the simultaneous individualization and totalization of modern power structures' (*SP*: 336).

---

to be a good ruler, citizen, husband, father or friend: 'a person who took proper care of himself would, by the same token, be able to conduct himself properly in relation to others and for others. A city in which everybody took proper care of himself would be a city that functioned well and found in this the ethical principle of its permanence' (*ECS*: 287).

4. The Aesthetics of Existence

A last major point of inspiration for Foucault from the ancient tradition of the concern for the self was how the choice to take care of oneself was to a great extent motivated and informed by the wish to live a beautiful life or, in Foucault's preferred formulation, to create oneself as a work of art and establish what he calls an 'aesthetics of existence' (cf. *GE*: 261, 262).[14] In other words, what characterizes this ethical tradition in contrast to later religious or juridical frameworks is that one decides to submit to certain rigid and austere moral codes not out of blind respect or obedience to the (rational, divine or civil) law, but out of individual concern for oneself and a *personal choice* to acquire an ēthos or way of being that is 'good, beautiful, honorable, estimable, memorable and exemplary' (*ECS*: 286; cf. also *GE*: 266–268, 271). Even though Foucault rejects the idea that this Greco-Roman model might offer a plausible alternative to the institutionalization of modern ethics,[15] he believes we can nevertheless learn something from it:

> We don't have to choose between our world and the Greek world. But since we can see very well that some of the main principles of our ethics have been related at a certain moment to an aesthetics of existence, I think that this kind of historical analysis can be useful. For centuries we have been convinced that between our ethics, our personal ethics, our everyday life, and the great political and social and economic structures, there were analytical relations, and that we couldn't change anything, for instance, in our sex life or our family life, without ruining our economy, our democracy, and so on. I think we have to get rid of this idea of an analytical or necessary link between ethics and other social or economic or political structures. (*GE*: 261)

Despite his general academic and descriptive tone and his reluctance to provide 'solutions' (cf. *GE*: 256), Foucault often expresses his

---

14 For criticism of Foucault's overemphasis on the aesthetic (compared to the therapeutic) character of the Hellenistic schools, see Ure (2007).

15 Cf. *GE* (256): 'I am not looking for an alternative; you can't find the solution of a problem in the solution of another problem raised at another moment by other people.' As Paul Veyne (1993: 2) puts it, 'Foucault's affinity with ancient morality is reduced to the modern reappearance of a single card in a completely new hand: the card of the self working on the self, the aestheticization of the subject, in two very different moralities and two very different societies.'

admiration for this Hellenic model of self-creation and his yearning for its rebirth in modern societies:[16]

> The idea of the *bios* as a material for an aesthetic piece of art is something that fascinates me. […] What strikes me is the fact that, in our society, art has become something that is related only to objects and not to individuals or to life. That art is something which is specialized or done by experts who are artists. But couldn't everyone's life become a work of art? Why should the lamp or the house be an art object but not our life? (*GE*: 260–261)

In conceiving the possibility and desirability of an 'aesthetics of existence', Foucault admittedly comes again very close to Nietzsche's own conception of ethics and subjectivity, as expressed in *Daybreak* and *The Gay Science,* for example.[17] In a well-known passage from the latter, Nietzsche claims:

> To 'give style' to one's character – a great and rare art! It is practiced by those who survey all the strengths and weaknesses that their nature has to offer and then fit them into an artistic plan until each appears as art and reason and even weaknesses delight the eye. (*GS*:290)

Nietzsche opposes the 'strong and domineering natures' who devote themselves to the 'long practice and daily work' of aesthetic self-cultivation to 'the weak characters with no power over themselves who *hate* the constraint of style' (*GS*: 290). The ideal, which both Nietzsche and Foucault borrow from the Hellenistic schools, is the exercise of 'a perfect mastery over oneself' (*GE*: 259). A further point of agreement (equally inspired by the ancient tradition of the care of the self) between Nietzsche and Foucault is that the practice of cultivating one's self involves creation and creativity rather than knowledge and the discovery of a supposed essence. That is, cultivating one's self is a matter not of discovering one's true or authentic hidden self but rather of creating it, shaping it, 'giving style' to it. Explicitly acknowledging the affinity between Nietzsche's view and his own in this regard, Foucault claims that 'from the idea that

---

16   On the 'aesthetics of existence' as Foucault's positive alternative to 'ethics in its canonical forms', see Smith (2015). See also Veyne (1993) and Davidson (2006) for Foucault's own appropriation and use of Greek ethics.

17   On the influence of Nietzsche on Foucault's aesthetic conception of ethics, see Ansell-Pearson (1991, 2015).

the self is not given to us, I think that there is only one practical conse-
quence: we have to create ourselves as a work of art' (*GE*: 262). Whereas
Nietzsche illustrates this process with the image of a 'gardener' who cul-
tivates his drives 'as productively and profitably as a beautiful fruit on a
trellis' (*D*: 560), Foucault uses the metaphors of a governor, the head of
an enterprise, a head of household, and 'a sovereign against whom there
would no longer be revolts' (*GE*: 272). According to Foucault, this idea of
perfect governance of oneself remained the central focus of ancient ethics
for centuries, until the advent of Christianity.[18]

If the culture of the self (whose 'golden age' Foucault identifies in
the first and second centuries A.D.) did not disappear with Christianity,
it was nevertheless appropriated, displaced and put to different uses in
the new religious framework. More concretely, 'the problem of ethics as
an aesthetics of existence [was] covered over by the problem of purifi-
cation' (*GE*: 274), and the self, which had been previously the material
for the crafting of a work of art, became something to be deciphered and
renounced, just as concentration on one's self was opposed to God's will,
conceived as a serious obstacle to the soul's salvation (cf. *GE*: 271; *HS*:
250). Foucault sees Christianity as one of the biggest causes of the ob-
scuration and neglect of the tradition of the care of the self in modern
societies. Under the influence of the Christian message of altruism and
self-denial, we became used to the idea that self-love is wrong and that
care for oneself is a despicable form of egoism. As a consequence, we
tend to look somewhat suspiciously on a form of morality that revolves
around notions like 'caring for oneself', 'devoting oneself to oneself',
'withdrawing into oneself', and so on (cf. *HS*: 12–13). Another cause
of the obliteration of this tradition of thought is the 'history of truth',
which is related to what Foucault calls the 'Cartesian moment' (cf. *HS*:
14, 17ss.). Very succinctly, Descartes symbolizes the movement that, in
the modern age, requalified and prioritized the principle 'know yourself'
over 'care for your self' by placing self-evidence at the foundation of
philosophical inquiry and making knowledge independent of spiritual
transformation. The care of the self, which had until then been the very
foundation of philosophy, was thereby disqualified and philosophically

---

18   This goal of perfect governance of the self is equivalent to 'a sort of permanent
      political relationship between self and self' (*GE*: 272). Political metaphors for the
      self are also abundant in Nietzsche; see, for example, *BGE* (12, 19).

discredited (cf. *HS*: 14). For these reasons, Foucault believes, the care of the self was ultimately neglected and 'left in the shadow' by Western thought 'in its reconstruction of its own history' (*HS*: 12).

Despite the episodic reappearance of an aesthetics of existence in the Renaissance and in nineteenth-century dandyism[19] – and also the successive philosophical attempts to reconstitute an ethics of the self by, among others, Montaigne, Stirner, Schopenhauer and Nietzsche (cf. *HS*: 251) – Foucault suggests that we 'have hardly any remnant of the idea in our society that the principal work of art in which one must apply aesthetic values, is oneself, one's life, one's existence' (*GE*: 271). There is also no real care of the self today, at least not in the organized and dominant way characteristic of antiquity, which leads Foucault to deny that 'we have anything to be proud of in our current efforts to re-constitute an ethic of the self' and to suspect that 'we find it impossible today to constitute an ethic of the self' (*HS*: 251–252). Nevertheless, Foucault believes that the constitution of a new ethic of the self 'may be an urgent, fundamental, and politically indispensable task' since 'there is no first or final point of resistance to political power other than in the relationship one has to oneself' (*HS*: 252).[20]

5. Philosophy as a Weapon of Resistance

In one of the late interviews quoted above, Foucault agrees that an im-portant function of philosophy has always been to warn of the dangers of power, adding that 'philosophy is that which calls into question domina-tion at every level and in every form in which it exists, whether political,

---

19   On Baudelaire's *dandysme*, see for example *WE* (312): 'Modern man, for Baude-laire, is not the man who goes off to discover himself, his secrets and his hidden truth; he is the man who tries to invent himself. This modernity does not 'liberate man in his own being'; it compels him to face the task of producing himself.'

20   In an interview from 1984, Foucault tempers this by claiming that the relation-ship of the self to the self is not the *only* possible point of resistance to political power: rather, the relationship of the self to itself is essential to his concept of 'governmentality' which, in turn, 'makes it possible to bring out the freedom of the subject and its relationship to others' (*ECS*: 299–300).

economic, sexual, institutional, or what have you' (*ECS*: 300–301). Fou-
cault consistently pursues this aim throughout his work, even if its strong
ethical and political significance was not brought to the fore until the final
period of his thought. If the most challenging philosophical, ethical and
political task nowadays is to 'refuse what we are' (*SP*: 336) through the
establishment of new practices of subjectivity and a new ethics of the self,
then Foucault's writings, in particular his warnings against modern disci-
plinary power and his presentation of alternative modes of self-constitu-
tion, were to a great extent meant to function as a powerful catalyst in this
regard.[21] In effect, to warn of the dangers of power and call into question
domination at every level is at the same time to promote individual liber-
ation and freedom – a role with which Foucault identified significantly:

> My role – and that is too emphatic a word – is to show people that they are much
> freer than they feel, that people accept as truth, as evidence, some themes which
> have been built up at a certain moment during history, and that this so-called
> evidence can be criticized and destroyed. To change something in the minds of
> people – that's the role of an intellectual. (*TPS*: 10)

In the same interview, Foucault adds that one of his most important
aims has always been to show how our most familiar landscapes are not
necessary or universal entities but rather a product of very complex po-
litical and social processes which have been forgotten (*TPS*: 11). Bring-
ing them to light is both the first step in overcoming them and Foucault's
particular contribution in this regard. Thus, Foucault's analyses in the
1970s on the coercive and oppressive nature of modern disciplinary so-
cieties are far from expressing a spirit of consent or resignation: within
limits, resistance to disciplinary power is possible, and Foucault's gene-
alogical work aims to contribute precisely to this aim. In this context, it
seems appropriate to quote Veyne's description of Foucault's philosoph-
ical activity, following his endorsement of Jean-Claude Passeron's char-
acterization of Foucault as a 'warrior in the trenches' (Veyne 1993: 2):

> Despite what the justificatory or self-protecting philosophers assert, the spectacle of
> the past brings to light no reason in history other than the struggles of men for some-
> thing that is undoubtedly neither true nor false but that imposes itself as truth to be
> told. If this is so, a philosophy has only one possible use, which is making war: not

---

21   See Kelly (2013) and Koopman (2013).

the war of the day before yesterday, but today's war. And for this, it has to begin by proving genealogically that there is no other truth of history but this combat. […] To be a philosopher is to make a diagnosis of present possibilities and to draw up a strategic map – with the secret hope of influencing the choice of combats. (Veyne 1993: 6)

As Veyne stresses in the same text, as a genealogist the philosopher cannot claim that he is right and all others wrong; he can only show that others are wrong in claiming that they are right (Veyne 1993: 6). Accordingly, Foucault's aim is not to substitute certain truths for others, but rather to destroy the validity of accepted truths by showing their contingency and arbitrariness, thereby opening a space of freedom for individual self-criticism, liberation and creativity.[22] As a genealogist, his focus might be the past; as a warrior, it might be 'today's war'. As a philosopher, however, his target is and must remain the future: 'the critique of what we are is at one and the same time the historical analysis of the limits imposed on us and an experiment with the possibility of going beyond them' (*WE*: 319). In other words, his genealogies of the formation and constitution of the modern subject are at the same time a provocation and challenge for future change and self-transformation.[23] By constantly challenging acquired truths and beliefs, by questioning our common values and customs, by calling attention to the dangers of domination by our disciplines and institutions, by constantly disconcerting us with regard to our relation to ourselves and our social and political surroundings, and finally, by giving us, in the last period of his

---

22   Cf. *TPS* (11): 'All my analyses are against the idea of universal necessities in human existence. They show the arbitrariness of institutions and show which space of freedom we can still enjoy and how many changes can still be made'.

23   On the two layers – the formation and self-formation of the subject – of Foucault's ethical project, see Koopman (2013), who convincingly argues that 'Foucault's ethical writings are […] located at the hinge between a history of the formation of the subject and the possibility of the future transformation of the subject. […] For ethics, as a first-order practice of emplaced activity rather than a second-order discourse on such activity, requires both the backward-looking historical gaze in which we discern the inheritance that bears on present ethical action and also the forward-looking, future-oriented hope in virtue of which that inheritance is productively transformed with the resources furnished us by our present. Foucault's ethics explicitly takes up the relation between these tasks as we face them in our present.' (Koopman, 2013: 526) On Foucault's philosophy as a contribution to self-transformation, see also O'Leary (2002: 140 ss.).

thought, a lively example of a different way of relating to ourselves and acquiring governance and sovereignty over ourselves, Foucault does indeed promote a sort of awakening and an inspiration to change: urging us to become masters of our own processes of subjectification and contributing to individual liberation. Seen in this light, it seems undeniable that there is a strong continuity in Foucault's work and that it does fulfil the profoundly ethical Socratic imperative, which, in his view, has characterized philosophy from its very beginnings (cf. *ECS*: 301).

References

Works by Foucault and Nietzsche

Foucault, Michel (1986). *The History of Sexuality*, vol. III: *The Care of the Self*. New York: Pantheon Books. (quoted as CS)

Foucault, Michel & Rux Martin (1988). 'Truth, Power, Self: An Interview with Michel Foucault', in Martin, L., H. Gutman, and P. Hutton, P. (eds.), *Technologies of the Self. A Seminar with Michel Foucault.* Amherst: The University of Massachusetts Press: 9–15. (quoted as TPS)

Foucault, Michel (1988). 'An Aesthetics of Existence' in: Kritzman, L. (ed.), *Michel Foucault: Politics, Philosophy, Culture. Interviews and Other Writings 1977–1984*. New York: Routledge: pp. 47–53. (quoted as AE)

Foucault, Michel (1990). *The History of Sexuality*, vol. II: *The Use of Pleasure*. New York: Vintage Books. (quoted as UP)

Foucault, Michel (1997). 'Self Writing' in Rabinow, P. (ed.), *Essential Works of Michel Foucault 1954–1984, Vol. 1: Ethics*. New York: The New Press: 207–222. (quoted as SW)

Foucault, Michel (1997). 'Technologies of the Self', in Rabinow, P. (ed.), *Essential Works of Michel Foucault 1954–1984, Vol. 1: Ethics*. New York: The New Press: 225–251.

Foucault, Michel (1997). 'On the Genealogy of Ethics: An Overview of Work in Progress', in Rabinow, P. (ed.), *Essential Works of Michel Foucault 1954–1984, Vol. 1: Ethics*. New York: The New Press: 253–280. (quoted as GE)

Foucault, Michel (1997). 'The Ethics of the Concern of the Self as a Practice of Freedom', in Rabinow, P. (ed.), *Essential Works of Michel Foucault 1954–1984, Vol. 1: Ethics*. New York: The New Press: 281–301. (quoted as ECS)

Foucault, Michel (1997). 'What is Enlightenment?', in Rabinow, P. (ed.), *Essential Works of Michel Foucault 1954–1984, Vol. 1: Ethics*. New York: The New Press: 303–319. (quoted as WE)

Foucault, Michel (2000). 'The Subject and Power', in Rabinow, P. (ed.), *Essential Works of Michel Foucault 1954–1984, Vol. 3: Power*. New York: The New Press: 326–348. (quoted as SP)

Foucault, Michel (2005). *The Hermeneutics of the Subject,* Lectures at the Collège de France, 1981–82. New York: Palgrave. (quoted as HS)

Nietzsche, Friedrich (1997). *Daybreak*. Ed. M. Clark and B. Leiter. Trans. R. J. Hollingdale. Cambridge: Cambridge University Press. (quoted as D)

Nietzsche, Friedrich (2001). *The Gay Science*. Ed. B. Williams. Trans. J. Nauckhoff. Cambridge: Cambridge University Press. (quoted as GS)

Nietzsche, Friedrich (2002). *Beyond Good and Evil*. Ed. R-F. Horstmann and J. Norman. Trans. J. Norman. Cambridge: Cambridge University Press. (quoted as BGE)

Works by other authors

Allen, Amy (2013). 'Power and the Subject', in Falzon, C., T. O'Leary and J. Sawicki (eds.), *A Companion to Foucault*. Oxford: Blackwell: 337–353.

Ansell-Pearson, Keith (1991), 'The Significance of Michel Foucault's Reading of Nietzsche: Power, the Subject, and Political Theory'. *Nietzsche-Studien* 20: 267–283.

_____ (2015). 'Questions of the Subject in Nietzsche and Foucault: A Reading of *Dawn*', in Constâncio, João, Maria João Branco, e Bartholomew Ryan (eds.), *Nietzsche and the Problem of Subjectivity*. Berlin/ Boston: Walter de Gruyter: 411–435.

Constâncio, João, Maria João Branco, & Bartholomew Ryan (eds.) (2015). *Nietzsche and the Problem of subjectivity*. Berlin/ Boston: Walter de Gruyter.

Constâncio, João, & Marta Faustino (forthcoming). 'Nietzsche and Foucault on Power. From Honneth's Critique to a New Model of Recognition', in Rosenberg, A., and J. Westfall (eds.), *Foucault and Nietzsche*. New York: Bloomsbury Academic.

Davidson, Arnold I. (2006). 'Ethics as Ascetics: Foucault, the History of Ethics, and Ancient Thought' in: Gutting, G. (ed.), *The Cambridge Companion to Foucault,* 2nd ed. Cambridge: Cambridge University Press: 123–148.

Harrer, Sebastian (2005). 'The Theme of Subjectivity in Foucault's Lecture Series'. *L'Herméneutique du Sujet*. *Foucault Studies* 2, 2005: 75–96.

Ingram, David (2006). 'Foucault and Habermas', in Gutting, G. (ed.), *The Cambridge Companion to Foucault,* 2nd ed. Cambridge: Cambridge University Press: 240–283.

Kelly, Mark G. E. (2013). 'Foucault, Subjectivity, and Technologies of the Self', in Falzon, C., T. O'Leary, and J. Sawicki (eds.), *A Companion to Foucault*. Oxford: Blackwell: 510–525.

Kelly, Michael (2013). 'Foucault on Critical Agency in Painting and the Aesthetics of Existence', in Falzon, C., T. O'Leary, and J. Sawicki (eds.), *A Companion to Foucault*. Oxford: Blackwell: 243–263.

Koopman, Colin (2013). 'The Formation and Self-Formation of the Subject in Foucault's Ethics', in Falzon, C., T. O'Leary, and J. Sawicki (eds.), *A Companion to Foucault*. Oxford: Blackwell: 526–543.

Lynch, Richard A. (2016) *Foucault's Critical Ethics*. New York: Fordham University Press.

O'Leary, Timothy (2002). *Foucault and the Art of Ethics*. London: Continuum.

Simons, Jon (2013). 'Power, Resistance, and Freedom' in Falzon, C., T. O'Leary, and J. Sawicki (eds.), *A Companion to Foucault*. Oxford: Blackwell: 301–319.

Smith, Daniel (2015). 'Foucault on Ethics and Subjectivity: "Care of the Self" and "Aesthetics of Existence"'. *Foucault Studies* 19: 135–150.

Ure, Michael (2007). 'Senecan Moods: Foucault and Nietzsche on the Art of the Self'. *Foucault Studies* 4: 19–52.

Veyne, Paul (1993). 'Final Foucault and His Ethics'. *Critical Inquiry*, Vol. 20, No. 1: 1–9.

Some notes on Rawls' critique of Kant's comprehensive moral philosophy

ANTÓNIO MARQUES

'Justice as fairness is a theory of our moral sentiments as manifested by our considered judgments in reflective equilibrium.' (John Rawls, *A Theory of Justice*, 20, 120)

One of the most striking features of the Rawlsian theory of justice developed in his late work is expressed in the claim that it is a political theory and not a kind of religious, metaphysical or moral doctrine. One may think that in the whole of his work Rawls maintains the same central idea, expressed in the following terms: 'Since there is no reasonable religious, philosophical, or moral doctrine the affirmed by all citizens, the conceptions of justice affirmed in a well ordered society must be a conception limited to what I shall call "the domain of the political" and its values'[1]. Nevertheless this is a statement that seems to contradict much of explicit argument developed in his *A Theory of Justice* (*TJ*), where typically moral values, such as equality and self-respect are defined as most important primary goods. In fact as soon as one looks at the formulation of the first principle, which proposes an equal distribution of liberty, it seems that we are facing a theory based upon moral maxims and an egalitarian stance about moral values. But at the same

---

1   Rawls (1996: 199, 38). In our view the Rawlsian understanding and criticism of Kant´s moral philosophy as a kind of constructivism is critical in *PL*, not so much in *A Theory of Justice.* The rejection of the Kantian moral philosophy in *PL*, labelled as comprehensive doctrine, is in sharp contrast with the place that is reserved to the former in his early work. By this move Rawls intends to give an entire *political* status to theory, since only the agreement on political principles can solve conflicting comprehensive views inside a pluralist society. 'The fact of a plurality of reasonable but incompatible comprehensive doctrines – the fact of reasonable pluralism – shows that. As used in *Theory*, the idea of a well-ordered society as fairness is unrealistic' (Rawls 1996: xix).

time, in his 80s and 90s writings Rawls dismisses such qualification and claims that a correct conception of justice must avoid any confusion with other conceptions of the world of the kind mentioned above. In this context the confrontation with the moral and political philosophy of Kant is of central importance. It is not a overstatement the claim that it is just this confrontation that allows Rawls to design his theory as a political construction and to differentiate it from perhaps the strongest moral philosophy, that is the practical philosophy of Kant. In what follows I´ll insist in the notion that Rawls' political philosophy contains moral grounding elements, despite his defense of a pure procedural method without any comprehensive element *and claim that Kant´s moral philosophy not only doesn't ground his own political philosophy, but also that it is not a constructivism of the sort Rawls claims it is*[2].

In order to understand the Rawls argumentation about the political nature of his theory and the correspondent critique of comprehensive doctrines, one needs to clarify the primacy of politics over morals in the period of *PL*. This is only possible if we first evaluate the moral status of his early theory. So let's first briefly consider the qualification of egalitarianism to Rawl's theory, which seems to be evident or unproblematic if applied to the first principle of his *A Theory of Justice* (*TJ*), precisely the equality principle of justice[3]. It is a fact that though Rawls defends the priority of the principle of equal liberty, he considers also instances of unequal liberty without violation of the first principle. But typically situations of unequal liberty show up for the sake of liberty itself, since precedence of liberty means that liberty can be restricted only for the sake of liberty itself' (Rawls 1996: 39, 244). It is interesting that Rawls justifies eventual situations of unequal liberty with a formulation derived from the principle of difference. He remarks in the same passage of *TJ* that 'basic liberties may either be less extensive though still equal,

---

2   Of course many philosophers have noted the difficulty related to the moral or political status of Rawls' theory. For example Bernard Williams says that 'the supposedly political conception, then, is still a moral conception, one that is applied to a certain matter under certain constraints of content' (Williams 2005: 2).

3   In Rawls (1973: 86, 60-1), the first statement of both principles reads: 'First: each person to have an equal right to the most extensive basic liberty compatible with a similar liberty for others. Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all'.

or they may be unequal. If liberty is less extensive, the representative citizen must find this a gain for his freedom on balance; and if liberty is unequal, the freedom of those with lesser liberty must be better secured . In both instances the justification proceeds by reference to the whole system of the equal liberties'. Anyway the second principle, by contrast with the first one, doesn't seem to pose questions in relation to equality and comprehensive issues, just because in it what is at stake is that the differences among individuals are justified if the distribution of socio-economic goods don't damage the expectations of the worst off. Anyway, as it has been remarked, inequality or even growing inequality is not incompatible with the fulfillment of those expectations[4].

Coming back to the theory of justice as a political theory without moral grounds, the question is if it simply operates as a procedural method by which equal and free citizens agree on certain principles of justice. Allegedly Rawls doesn't introduce moral intuitions in his theory, that is, features, which typically belong to what he calls 'comprehensive' elements. By 'comprehensive' doctrines Rawls understands those that operate with moral, religious or some kind of metaphysical values. From the point of view of a theory of justice a comprehensive conception is to be avoided, since it is *not* a conception of forms of life or of the most important virtues that should determine our whole existence. A theory of justice is, say, an austere *political* conception of the principles and avoids to be overloaded by any sort of metaphysical, religious or moral elements.

In what follows we have to take into consideration, *first* the characterization of the original position (OP) as a device of procedural justice that on Rawl's view should not integrate any moral intuition, *second* the interpretation of Kant's moral philosophy as a moral constructivism. As already mentioned neither the justice as fairness is exempt of moral

---

4   Watson (2015) makes the following point: 'The "Rawlsian" position of favouring the least advantage may sound quite radical. However it is not far removed from the statements of politicians who argue for income tax cuts on the basis that these would stimulate economic activity and hence increase revenue that could be used to raise the incomes of the poorest among us. As this argument illustrates, there is nothing intrinsically egalitarian about the Rawlsian objective?' (Watson 2015: 13).

intuitions, nor Kant's moral is constructivist at least in the way Rawls understands it.

To the first point, as it is well known, Rawls' theory begins by presenting a situation (OP), where reasonable individuals choice principles in order to a just distribution of what he calls a set of 'primary goods'. What is required to the parts is a capacity of rational choice, subjected to certain rules of impartiality (veil of ignorance), and a sense of justice in order to fairly distribute those goods. In that position persons under reasonable or fair conditions will select certain principles of justice. As mentioned above it is largely accepted that in the original position individuals don't operate with any moral intuitions whatsoever. The fact that the parties have a capacity or sense of justice[5] doesn't mean that they are motivated by moral convictions, that is, a real comprehensive understanding of society and its values[6].

Nevertheless Rawls sees the parts in the original position as possessing a sense of justice and the inclusion of such a sense in the structure of rational choice. So a correct description of the OP seems to contradict a picture that represents it completely deprived of comprehensive moral motivations. The question is whether in the real process of designing the institutions of justice, the sense of justice starts to work only *after* the identification of the principles. This is a basic problem that deserves going deeper in it. Referring explicitly to the central place of the sense of justice, Rawls says that it shows itself in two ways. 'First, it leads us to accept the just institutions that apply to us and from which we and our associates have benefited' and secondly,

---

5   Rawls (1973: 86, 567-577; 1996: 19). In the same passage of *PL*, Rawls refers to 'two moral powers': sense of justice, that is the capacity to understand, to apply, and to act from the public conception of justice and a capacity for a conception of the good.

6   Samuel Freeman comments at this point: 'In the original position they (the parties) are not moved by moral considerations (e.g. to do what is just or fair, or make a morally right decision) or by benevolence toward other parties; nor are they directly concerned with other's developing their capacities for justice except in so far as it benefits themselves. They are uninterested' in that they are *indifferent to one another* under these extraordinary circumstances of choice in the original position' (Freeman 2007: 168). Interestingly, Thomas Nagel characterizes both principles in the following terms: 'The first principle is a principle of strict equality, and the second a principle of permissible inequality' (Nagel 2003: 66).

'a sense of justice gives rise to a willingness to work for (or at least not to oppose the setting up of just institutions' (Rawls 1973: 72, 474). So there is no doubt that the sense of justice is a fundamental force in the building and preservation of just institutions. Furthermore it seems that it is when individuals act against the principles of justice, the process of full moral development begins, namely it is when persons violate some rule directly tied to the principles that moral feelings as guilty, shame or repentance show up: 'When we go against our sense of justice we explain our feelings of guilt by reference to the principles of justice … The complete moral development has now taken place and for the first time we experience feelings of guilt in the strict sense; and the same is true of the other moral emotions' (Rawls 1973: 72, 474).

The main idea here is that moral feelings and moral life in general takes form by reference to the principles of justice what enforces the conception of an OP including 'sense of justice', despite the absence of its moral content (Rawls 1996: 97). Rawls makes the point that individuals develop their moral forces just because they coincide with the content of the principles of justice, which is the same to say that these ones don't have a logical and chronological primacy over moral principles. 'Individuals in their role as citizens with a full understanding of the content of the principles of justice may be moved to act *upon* them largely because of their bonds to particular persons and an attachment to their own society' (Rawls 1973: 475; my italics). So on must recognize that in the *TJ* the question about the moral content of the principles is not so much an ambiguous one, just because individuals are defined as moral persons. It is clear that in the initial situation I represent myself and the others as moral persons with a sense of justice. Nevertheless in *PL* he insists that the theory is not comprehensive and repeatedly vindicates and emphasizes the status of its unique political not moral characteristic. It is not without saying that Rawls is totally aware of the difficulty of conceiving persons (the kind of persons in the OP) who at the same are not entitled with any moral or metaphysical powers[7].

---

7   Rawls (1996: 29): 'I remarked earlier that the idea of the original position and the description of the parties may tempt us to think that a metaphysical doctrine of the person is presupposed (...) To rebut claims of this nature requires discussing them in detail and showing that they have no foothold. I cannot do that here'.

In order to accomplish the aim of designing a well-ordered powers should be only political, and this is what is necessary for individuals to operate in the OP. So from a *political* point of view, persons in order to operate in the OP are enough entitled with freedom: citizens are conceived as thinking of themselves as free in three respects: first as having the moral power to have a conception of the good, second as self-authenticated sources of valid claims, that is, claims on their institutions, and third as they view themselves and the other as capable of taking responsibility for their ends (Rawls 1996: 29-33). These are political capacities that are supposed to be exempt of any comprehensive element.

The question now consists in making this assumption compatible with the idea of a *pure procedural* conception of justice, which characterizes the OP and its nature as it is clearly assumed in the later *PL*. A procedural method means first of all that no fundamental principles or set of principles are established as the starting and fixed moment of a theory. If the *TJ* is set upon a procedural stance, then confusion with any kind of comprehensive doctrines is avoided just because in this procedural move there is no fixed metaphysical elements. In both, *TJ* and *PL* Rawls refers to the OP as 'a case of procedural justice' distinguished from a '*perfect* procedural justice'. The last one is commonly illustrated by the division of a cake, in which the concerned individuals are required to divide a cake without knowing the order of choice. Under this procedure an outcome is generated , which is just, whatever the outcome may be. *Pure* procedural justice is different in the sense that the parties don't specify from the beginning any set of principles and they arrive to a deliberation by measuring the all reasons in balance. What guides their deliberation is the consideration of mutual interests in a situation of reciprocal action. Furthermore guidance by mutual interest expresses a contractualist agreement or a justified mutual agreement, to which Rawls refers to as a 'reflective equilibrium'. This state is achieved as far as no one has enough reasonable reasons to reject the other's justified point of view. It is not always clear whether a reflexive equilibrium state is negotiated among parts, but at least there is reciprocal consideration of points of view. Accordingly Rawls states that justice as fairness is not procedurally neutral.

This is a view that can be shared by another contractualist authors, namely Thomas Scanlon who proposes a contractarian perspective of

wrong and right, which is structured upon the principal idea that what is good is what can be justified to others on grounds that they, if appropriately motivated, could not reasonably reject'[8]. What is relevant in this notion is that all points of view are articulated in a whole on the basis of a reasonable justified motivation. The moral life in this scanlonian perspective is a permanent process of reciprocal and interactive reasonable justification. This is very near of the structure of the OP, but in the case of Scanlon his contractualism doesn't exclude a central moral motivation.

Rawls conceives a pure *procedural method* on the basis of a *reflexive equilibrium*, in which each personal point of view reaches an agreement with all concerned individuals based in rational justification. Only then it is possible to represent a real, not utopic consensus, which Rawls refers to as an overlapping one. Only then it is also possible to conceive a true contractualist agreement without the threat of comprehensive elements. Seemingly all this conceptual structure that integrates the OP (perhaps one should clarify the differences between the *TJ* and *PL* in respect the OP, but this is particular hermeneutical task that cannot be carried out here) operates only at a political level and in his late writings Rawls designates such dynamic conceptual network, whose unique aim is to formulates the principles of justice, as a 'constructivism'.

This leads us to the discussion of Kant's moral philosophy. In fact, as already mentioned, the best point of view to get a clear perspective of the place of the moral element in Rawls philosophy would be the discussion of Kant's philosophy as a 'moral constructivism' in the terms of Rawls himself. In his view the contrast between his political constructivism and that of Kant's moral is the right standpoint to evaluate the role, if any, of the moral element in the original position. When we use the word 'constructivism' applied to a theory or doctrine what do we

---

8   Scanlon (2000: 5). Scanlon is aware that he needs go deeper in his notion of reasonable justification that no one could reject. This implies substantive judgments and claims made in specific practical situations. See Scanlon (2000: Ch. 5). Furthermore in the words of Scanlon, 'This gives us a direct reason to be concerned with other people's point of view: not because we might , for all we know, actually *be* them, or because we might occupy their position in some other possible world, but in order to find principles that they, as well as we, have reasons to accept' (Scanlon 2000: 191).

mean by that? What kind of description of that doctrine are we pointing to? In Kant the starting moment of his moral system is the proof of the existence of a moral law, the 'fact of reason' as the first grounding act of practical rationality. And if one proceeds rightly from that initial moment, subjects can build a system of rights, duties and virtues, which is the object of the second part of his 'Doctrine of Virtues' exposed in his last systematic work on the foundations of law and morality, *The Metaphysic of Morals*. It seems that it is in this sense it would be fair to qualify or at least to refer to a sort of constructivism in Kant's morals. Another problem consists in saying that it can endorse a political constructivism, to which we'll refer to later.

In fact, with special emphasis in *PL*, Rawls sees the Kantian moral philosophy as a moral constructivism, from which his own political philosophy clearly separates itself. In this work he states that the first difference is that Kant's practical philosophy is comprehensive since it is based upon an ideal of autonomy 'for all life' (Rawls 1996: 99). As a deeper meaning of 'autonomy' it is meant that the order of moral and political values are made or constituted by the work of practical reason. However this conception of autonomy of practical reason doesn't give the basis of justification for a genuine 'public basis of justification' (Rawls 1996: 99)[9], since it seems to dispense each person to enter in agreement with other. In Rawls words, Kant's constructivism 'is deeper and goes to the very existence and constitution of the order of values. This is part of his transcendental idealism' and adds that 'Political Liberalism must, of course reject Kant's constitutive autonomy; yet his moral constructivism can endorse political constructivism as far as it goes' (Rawls 1996: 100). Anyway the Kantian fact of reason and the correspondent 'constitutive autonomy' of the moral person cannot be considered an adequate starting ground to the genuine political constructivism, without any comprehensive intrusion.

---

9   In his 'Reply to Habermas', Rawls refers to the notion of 'public justification' as the existence of an *overlapping consensus* in a society where reasonable citizens, embedded in different comprehensive views, can 'judge from within their reasonable comprehensive doctrines that political values are very great values to be realized in the framework of their political and social existence, ans a shared public life on terms that all reasonable parties may reasonable be expected to endorse' (Rawls 1996: 393).

At this point of the Rawlsian argumentation, one has to ask if there is such a thing as a moral constructivism in Kant and if there is, if it corresponds to that kind of constructivism claimed by Rawls. Furthermore, if the answer is negative, one must ask if the kind of moral constructivism that Kant offers us with, 'can endorse political constructivism'. As far as Rawls associates the notion of constructivism with other elements of justice as fairness, like 'pure procedural justice', 'reflexive equilibrium' and 'overlapping consensus', a constructivist method is generated by practical rationality and provides a set of grounding criteria or procedures for the set up of a ranked network of values and institutions, what he calls the 'basic structure' of a society, like the constitution and some other fundamental institutions. In fact one must say that there is a kind of a constructivist method in late Kant's moral philosophy, but not the kind of one that claimed by Rawls[10]. Assuming that any constructivism starts from some first material to other concepts and rules, these are not to be inferred from the former. In Kant's view construction is a movement of the reason, either theoretical or practical, that proceeds by synthetic amplification. Under his understanding of practical reason this would be an *a priori* amplification, in the sense that one proceeds from some basic geometrical structure and goes on adding other properties that are nor simple deducible from the initial material. In the case of moral practical reason, the first structure is the existence of a self-coercive moral internal law, which corresponds to the very concept of duty. This is the initial and starting point to the construction of other *a priori* figures of duty, that Kant pretends to amplify in a network of what he calls a 'doctrine of virtue'. So he notes that 'the principle of the doctrine of virtue goes beyond the concept of outer freedom and connects with it, in accordance with universal laws, an *end* that it makes a *duty*. This principle is therefore synthetic' (Kant 1996a: 526, Ak. 6: 396)[11]. It is

---

10   Onora O'Neill remarks that 'it is far from clear that Rawls's fundamental strategy of justification is Kantian. Like Kant, Rawls does not appeal to *individual preferences,* or to a notional *hypothetical social contract,* or to an *independent* order of moral values. But unlike Kant, Rawls basic approach to justification is apparently coherentist' (O'Neill 2004: 23).

11   In this work Kant presents what can be seen as a complete scheme of duties of virtues, in fact the outcome of his synthetic method, which starts from internal law of duty to a net work of 'virtues of duty'.

interesting how Kant understands duties as so many ends that can and must be balanced one against each other. But there we stand on a domain clearly separated from the political one. Kantian moral constructivism, as it can be understood in *The Metaphysics of Morals*, doesn't cross whatsoever or contribute to the first principles of a just society or of a state of law. As it is well known the grounding moment of a political society from Kant's contractualist point of view doesn't require any moral powers of the citizens, but only an intelligent choice of coexistence principles. Kantian contractualism has a mere political basis and is of a more Hobbesian sort than it is generally thought. As he famously states in his essay on perpetual peace 'the problem of establishing a state, no matter how hard it may sound, is *soluble* even for a nation of devils (if only they have understanding) … Such problem must be soluble. For the problem is not the moral improvement of human beings but only the mechanism of nature, and what the task requires one to know is how this can be put to use in human beings in order so to arrange the conflict of their unpeaceable dispositions…' (Kant 1996b: 335, Ak. 8: 366). This Hobbesian statement means also that in Kant´s political philosophy the boundaries between moral and philosophy are pretty clear and that if one wants to design some articulation between both domains, it must be made taking into account that neither morals grounds politics, nor the last one grounds the former. This is an important reminder for anyone (as Rawls), who claims for a political philosophy without comprehensive interferences. Kant made the same claim but at the end his track leaded him to a place quite near of Hobbes.

References

Freeman, Samuel (2007). *Rawls*. London: Routledge.

Kant, Immanuel (1996a). *The Metaphysics of Morals*, in *Practical Philosophy: The Cambridge Edition of the Works of Immanuel Kant*. Cambridge: Cambridge University Press.

____ (1996b). *Toward Perpetual Peace: A Philosophical Project*, in *Practical Philosophy: The Cambridge Edition of the Works of Immanuel Kant*. Cambridge: Cambridge University Press.

Nagel, Thomas (2003). 'Rawls and Liberalism', in Freeman, Samuel (ed.), *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press

O'Neill, Onora (2003). 'Constructivism in Rawls and Kant', in Freeman, Samuel (ed.), *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press.

____ (2004). 'Constructivism vs Contractualism', in Stratton-Lake, Philip (ed.), *On What We Owe to Each Other*. Oxford: Blackwell.

Rawls, John (1973). *A Theory of Justice.* Oxford: Oxford University Press.

____ (1996). *Political Liberalism*: New York: Columbia University Press.

Scanlon, Thomas (2000). *What We Owe to Each Other*. Harvard: Harvard University Press.

Watson, Anthony (2015). *Inequality: What can be done?* Harvard: Harvard University Press.

Williams, Bernard (2005). 'Realism and Moralism in Political Theory', in *In the Beginning Was the Deed*. Princeton: Princeton University Press.

Moral Choice without Moralism

Erich Rast

A moralist is a pedant who insists on conforming to strict moral rules in each and every case. This attitude is often a symptom of 'moral tyranny', which arises '…when the claim to superior knowledge of good and evil lacks justification' (von Wright 1963: 189). Both attitudes can lead to a rejection of the underlying moral codes and undermine morality in general. How can a decision maker act morally without becoming a moralist?

In this article, I will address one aspect of this question from a purely decision-theoretic perspective by giving a tentative answer to a narrower and more specific question: What constitutes a moral decision, provided that in a given decision situation a set of morally relevant attributes can be identified and that these attributes do not automatically and unconditionally override all non-moral attributes? To answer this question in a decision-theoretic setting, I propose a theory with lexicographic thresholds that allows a decision maker to deviate from the prescription made implicitly by the morally relevant attributes whenever the stakes are low, but requires the decision maker to follow the morally relevant attributes when the stakes are high. Conditions are laid out under which a decision is morally acceptable (or, in short, 'moral') in this setting. The approach is based on the assumption that the decision making methodology is principally adequate for making moral decisions and is in principle compatible with much of what has been said in the literature on Practical Reasoning. Albeit controversial, this view will not be challenged here but rather supposed for the sake of this article. However, I believe that another assumption needs to be explained in more detail, namely the one that moral attributes do not unconditionally outrank non-moral attributes. In deontic logic a distinction is often made between actions that need to be done (required/obligatory actions; actions that are one's duty), actions that are permissible and

actions that are forbidden. Using these distinctions as a basis, the thesis boils down to the claim that a course of action that one is permitted to do may sometimes outweigh a course of action that one ought to do if not much is at stake from a moral point of view. Another, in my point of view better, way of putting this is based on Meinong's distinction of value classes. He classifies values into four categories: meritorious (*verdienstlich*), correct (*correct*), merely excusable (*zulässig*) and inexcusable (*verwerflich*), where the first two are good and the last two are bad. According to this taxonomy, some courses of action may be excusable in the sense of being morally permitted despite the fact that they may be judged morally bad because their prudential value outweighs their moral badness. The goal of the following sections is to formulate conditions for such excusable courses of action that allow one to distinguish morally acceptable from morally unacceptable decisions, and for this purpose thresholds will be introduced. Which threshold is the right one in a given situation, however, is a substantive moral question that I will not attempt to answer in general.

There are some known criticisms that I would like to exclude before going into the details. Decision theory has often been criticized for its purported inability to deal with moral dilemmas. The example of a resistance fighter torn between obligations to his country and to his sick mother is often given (Sartre 1946), and another type of such a dilemma may concern long-term choices between different ways of life. If such dilemmas exist, then by definition no account of practical reasoning whose outcome is supposed to be an action, choice of action or an intention to act can resolve them, and so with respect to such cases practical reasoning theories like those of Broome (2002) and Horty (2012) are in the same boat as decision theory. A genuine moral dilemma has by definition no rational 'solution' and since this article is concerned with the morality of decisions in situations in which a rational decision can be made, these types of examples need not be considered further in what follows.

In the remainder of this article, first a brief overview of rational decision making is given in Section 1. Then, in Section 2, conditions are laid out for classifying decisions as moral or not given that morally relevant attributes have been identified. These include unipolar thresholds, then bipolar thresholds are introduced to more adequately represent disvalue,

and later the changes needed to deal with uncertainty are discussed. The discussion is rounded up in Section 3 in which the connections of the proposal to existing work in decision making and to utilitarianism are addressed.

1. Rational Choice

The rational decision making tradition and accounts of practical reasoning in moral philosophy have diverged at some point in the history of ideas, and various factors may have contributed to this unfortunate development: What was formerly called Welfare Economics has dropped much of the welfare aspect from its domain of inquiry, the deontic tradition has always been predominant in moral philosophy, and much of the recent work in axiology has focused on special problems of value incommensurability and parity rather than the question of what makes a choice moral.

In order to make this article more or less self-contained, aspects of the decision making perspective on rational choice will be briefly laid in the following paragraphs insofar as they are relevant for the topic of moral decision making, although limitations of space will only allow me to scratch the surface (and there is a dangerous iceberg below). Readers familiar with standard additive decision theory may skip this section. The proposal itself and how it relates to the standard weighted-sum account will be laid out in Section 2.

Among the many topics that cannot be addressed in such a brief survey is the question of the rationality of decision making principles themselves and Expected Utility Theory in particular. This question has been addressed by a vast variety of authors starting with Ramsey (1931), von Neumann & Morgenstern (1947), Savage (1954), Debreu (1959) and Fishburn (1970). See Eisenführ et al. (2010) for a modern introduction to applied decision theory, Keeney & Raiffa (1976) for a more technical overview and Bouyssou et al. (2010) for details.

1.1 Decision Making Under Certainty

The key notion of decision theory are preferences. Among several alternatives a decision maker has preferences which reflect his values: one alternative might be better than another or they might be just as good. So as to be able to talk about the morality of decisions to act, we must assume that moral rules can give rise to corresponding preferences between action alternatives. For example, killing someone for selfish motives is considered murder unless the circumstances are exceptional such as self-defence or wartime. When a decision maker contemplates two hypothetical alternative courses of action and one of them involves murdering someone while the other does not, then avoiding murder should be preferred from a moral point view. In other words, avoiding murder has a higher value than not avoiding it. So does avoiding loss of life in general, but notice that these two alternatives are also comparable; it is commonly presumed that avoiding murder is preferable to avoiding mere loss of life. None of this should be controversial, yet the claim that all kinds of alternatives and aspects thereof are comparable is, of course, not so innocuous and has been attacked from time to time. See for instance the discussion in Chang (1997). But as mentioned above, moral dilemmas that lead to genuine value incommensurability shall not be considered and we focus on the decision-guiding aspects of values. Under this premise, some standard assumptions about the preference relations can be made. First of all, they are complete, i.e. a decision maker either prefers one alternative over the other or considers them equally good.

   To get a fully-fledged decision theory many more assumptions are needed. Generally, it is presumed that preferences can be described by value functions whose outcome is a numerical value. These numerical values can be compared with each other, reflecting the point of view that the decision maker can decide for any two courses of actions which one has a higher value than the other. In other words, if $a \succcurlyeq b$ says 'alternative a is preferred to b or the decision maker is indifferent about the alternatives' (weak preference), then there is a continuous value function $v(.)$ such that $v(a) \geq v(b)$ iff. $a \succcurlyeq b$. This fairly strong assumption is also made by the subjective expected utility theory of Savage (1954) and must be taken with a grain of salt. It implies that

preferences between alternatives are transitive, that there are no principally incommensurable values and that parity (Chang 2002; Gert 2004; Rabinowicz 2008) can be disregarded as well. Depending on how the values of individual attributes are combined into an overall valuation, there are also a number of further, more technical conditions. If they are combined by adding them, the attributes must be mutually preference independent. Other modes of combination such as multilinear models (Keeney & Raiffa 1993), generalized additive independence (Gonzales & Perny 2004) and multiplicative models (Krantz et. al. 1971) impose less strong conditions.

Not all of the standard assumptions about preferences need to be made for choice-guiding actions. For example Fishburn (1991) uses non-transitive preferences and Hansson (2001: Chapter 2) replaces transitivity by weaker conditions ('top transitivity' and 'weak eligibility'). However, for the purpose of this article I will stick to so-called additive models for simplicity. This means that in addition to some more technical conditions, preferential independence and difference independence need to be presumed. These conditions state that a comparison between two attributes does not depend on other attributes. To put it more precisely, preferential independence states that if there are two alternatives a, b that differ only in attribute i and there are two other alternatives a', b' that also differ only from each other in attribute i, and $a'_i=a_i$ and $b'_i=b_i$, then $a \succcurlyeq b$ if and only if $a' \succcurlyeq b'$.[1] Difference independence is an even stronger condition that says that in the same scenario with two preferences that only differ in one attribute a decision maker will be indifferent between the choice of shifting from a to b and a choice of shifting from a' to b'. If these conditions and a few more technical ones are fulfilled, additive value functions express preferences between alternatives by aggregating the values of their individual attributes. In a model with n attributes, the overall value of an alternative is their weighted sum $v(a) = \sum_{i=1}^{n} w_i v_i(a_i)$, where w_i represents the relative weight of an attribute

---

1   Other, more technical conditions such as restricted solvability and the Archimedean principle differ slightly between the case with two and three or more attributes but I suppress these details in what follows. Details can be found in Keeney & Raiffa (1993), pp. 104–117, Eisenführ et al. (2010), pp. 125–155, and Krantz et al. (1971) for the mathematical underpinnings.

(in comparison to the other attributes), v_i is the value function of the i-th attribute and a_i is the value of the attribute. The alternative with the highest value is the one recommended for rational choice; if there are more than one with a highest value, which might of course happen, then the decision maker is indifferent about them and may choose as she likes or attempt to refine the model to decide between these alternatives.

Someone who is less familiar with these kinds of models might ask why a value function is used instead of just directly using values for the respective attributes. To motivate this feature, it is instructive to take a glimpse at the related field of Consumer Theory in economics. In this theory each alternative is taken as a bundle of goods, and it is generally assumed that under normal circumstances the more of a good one has in a bundle while the amount of other goods is kept fixed, the higher will be the value of the bundle. Two bananas and an apple are worth more than one banana and an apple. In addition, however, it is often claimed that the Principle of Marginal Utility[2] holds, which has been confirmed empirically for some domains.[3] According to this principle, under normal circumstances the overall value of a good decreases in relation to other goods in a bundle the more units of the good one possesses. Consider the innocuous example of someone who has one banana and is willing to swap it for two apples. If the same person had twenty bananas, she might be willing to swap three bananas for one apple. The value of bananas in comparison to apples has dropped from two times an apple to one third of an apple. To model such cases, *cardinal value functions* are needed instead of purely ordinal ones; in addition to the ordinal information given by qualitative preferences, these represent information about preference intensities and allow for a comparison of preference differences. For simplicity, take the apple value function to be linear, defined by the two points $v_a(1)=0.1$ and $v_a(20)=0.3$. Furthermore, $v_b(1)=0.2$. Following the example, suppose

---

2   This principle was first formulated by Gossen (1854) as a characteristic (Merkmal) of all human pleasure (Genuss) in the context of a purely pleasure-utilitarian foundation of economics and later strongly criticized by purporters of the Marxist labour theory of value.

3   See for example Inglehart (1990) for the value of economic growth in societies, Liu (2003) for the quality of marital sex and Horowitz (2007) for general experimental results.

now that $v_b(20) \approx 0.3 \approx \frac{1}{3}v_a(20)$. To ensure that the banana value function does not violate the Principle of Marginal Utility, a third point is needed; suppose $v_b(10) = 0.28$ is this point. Given these premises, the value function may be described by a quadratic function $v_b(x) = 1.2 - x^{-0.0354}$ obtained by curve fitting; it is concave, monotonically increasing, has limit 1.2 and satisfies the Principle of Marginal Utility.

   Note that there may be some applications in *moral* decision making with cardinal value functions for which it makes sense to presume this principle and some for which it does not. For example, one might argue that in the evaluation of a young delinquent's misdemeanours the difference between the first and the second consecutive petty theft ought to count more than the difference between the twelfth and thirteenth such cases. On the other hand, consider the virtue-utilitarian value of polite actions per day such as opening a door for someone else or cheering someone up. Perhaps someone believes that ten such actions a day are optimal but twenty of them are just as good as one. After all, it is possible to be too polite. In that case $v(1)$=¼, $v(10)$=1, $v(20)$=¼ .[4] Similar cases also occur in consumer theory, where economists sometimes try to avoid them for technical reasons by only considering the increasing part of the function. I will not presume the principle in general in what follows and the account works for cardinal and ordinal value functions alike.

1.2 Decisions under Risk

While the focus of this article is on decision making under certainty, some remarks about decisions under risk and uncertainty seem to be appropriate as decision theory is usually applied in contexts with risk and uncertainty and this might hold in particular for moral decision making. Risk is generally dealt with by presuming the axioms of *Expected Utility Theory* of von Neumann & Morgenstern (1947). Omitting most

---

4   Again, the function may be approximated. For example by polynomial curve fitting $v(x) = -0.108x^2 + 2.269x - 1.911$ is obtained. It would of course also be possible to abstract from the subjective optimum and find a general 'politeness value function' with a parameter for the turning point, but since the example is a bit contrived this matter shall not be pursued further.

of the details and the (formal) justification of their approach, an additive decision model under risk can be obtained from the one laid out previously by defining a probability over the alternatives under consideration and then computing the expected utility of the alternatives in a way that is very similar to decision making under certainty. Instead of speaking of a value function, it is customary to speak of a utility function that must satisfy the same requirements as a value function. Taking $\vec{a} = \langle a_1, ..., a_n \rangle$ as a vector of attributes like before, we may define a consequence as a tuple $\langle p_j, a_j \rangle$ ($1 \le \leqslant j \le \leqslant m$) and take each alternative to have m consequences. The subjective expected utility of an alternative a is then $u(a) = \sum_{j=1}^{m} p_j \sum_{i=1}^{n} w_{j,i} u_{j,i} \left(a_{j,i} \right)$ in such a model.[5]

It is important to bear in mind that under risk a utility function fulfils two roles that are not clearly separable from each other. On one hand, it represents a value function, which might for example exhibit the non-linearity exemplified by the Principle of Marginal Utility. On the other hand, it may also represent the risk attitude of a decision maker. Suppose the decision maker's value function v is linear, i.e. has a straight line as a graph. That means that a change by any positive amount of an attribute has the same value no matter how much of it was already present. For example, getting four bananas is worth the same to the decision maker whether he already has four of them or none of them. If that is the case, a concave utility function represents risk aversion. It is easiest to see this by looking at bets with losses. Consider a 50/50 bet a with a win of 8 apples and a loss of 4. The expected value of this bet is $\frac{1}{2} v(8) + \frac{1}{2} v(-4) = 2$. Take another bet b with the same expected value, say a 50/50 bet with a win of 20 apples and loss of 16: $\frac{1}{2} v(20) + \frac{1}{2} v(-16) = 2$. Concavity of the utility function implies that $u(a) > u(b)$. Hence, the decision maker is more willing to take bet a with a possible loss of 4 apples than the more risky bet b with a possible loss of 16 apples. A similar argument shows that a convex utility

---

5   The notation and terminology differs slightly from that of Savage (1954), whose setting is more general than the one presented above. For what follows, the differences do not matter.

function implies that the decision maker is risk prone *unless* he already exemplifies the Principle of Diminishing Marginal Utility to a very high degree.[6] As laid out by Schoemaker (1982), the fact that the two functions of expected utility, subjective valuation versus risk attitude, are often not clearly kept apart in the literature has led to confusion. The following sections will focus on decision making under certainty and therefore avoid these problems.

2. Moral Decision Making

Let us return to the initial question of how to classify decisions as moral in a way that does not simply give moral absolute priority over other considerations. The goal is to find conditions for what makes a decision moral without taking *only* moral aspects of a decision situation into account. It seems obvious that such an approach must involve a threshold at one place or another; the decision maker is allowed to deviate from moral rules to some extent, but not when it matters and not too much.

Before going on, a flawed account needs to be addressed because it might seem intuitively appealing at first glance. What about the claim that a decision maker ought not deviate from the prescriptions of underlying moral rules 'too often'? This is not a good option. A highly immoral decision maker may make moral decisions most of the time. What matters in the end is what is at stake with each decision and not how often similar decisions over the same types of action are made. In an overall consequentialist approach what is at stake is determined by potential wins and losses, and a theory that discriminates between different stakes will at some point involve either thresholds or special functions over the outcomes. The threshold view is simpler and, as I believe, the adequate way to make the moral decision process permeable and 'soft'.

---

6   Arguably, it is harder to find examples of this attitude than in the converse case of diminishing marginal utility.

2.1 The Unipolar Threshold View

If a moral component can be distinguished from a non-moral component in decision making in a given situation, then certain attributes of alternatives must be morally relevant while others are not. What counts as morally relevant and not is a matter of the underlying moral theory. In a broadly-conceived deontic setting, the morally relevant attributes are those governed by a moral rule or norm. For example, the attribute 'number of lives lost' of a consequence of some alternative course of action is morally relevant because, notwithstanding certain exceptional circumstances, minimizing loss of life might be considered a moral obligation or one's duty. In contrast to this, the attribute 'pleasure of taste' concerns what von Wright (1963) calls a hedonic good and from a deontic perspective does likely not have moral significance.[7] It falls into the category of prudential value. On the other hand, from the perspective of a classical utilitarian, hedonic goodness might very well be morally relevant for its contribution to the overall welfare of a group. This difference illustrates the dependence of moral relevance on the underlying moral theory, and the precise nature of this connection is an open problem. Moreover, it seems often reasonable to assume that one and the same attribute can be morally relevant in one instance and irrelevant in another, and this issue is closely related to the previous one. Since it would go far beyond the scope of this article to address these problems here, I will assume in what follows that certain persons with moral expertise, such as perhaps ethics committees, can decide between morally relevant and irrelevant attributes in a given choice situation.[8]

Under this assumption, let us write M to denote the set of morally relevant attributes and N for the set of other attributes. A table that depicts

7 There might be a rule, however, stating that no one is permitted to needlessly deprive someone else of personal pleasure. However, it seems striking that this must be based on some threshold as well since there are many situations in which it is customary and legitimate to deprive people of pleasure – think for instance about work, which is rarely *always* fun. Notice further that methods for finding a balance of power and distinctions like that between positive and negative rights might be needed.

8 Which account of morality is the right one is a decidedly moral question for anyone but an extreme moral relativist.

the relevant parts of a value function over alternatives and attributes will from now on be called a *decision table*. Such a table may be said to *rationally ground* a decision if the decision maker acts according to it. Only rationally grounded decisions are considered from now on.

To tackle the problem we are dealing with, let me first introduce the concept of *dominance* and then lay out a similar concept for moral attributes. An alternative *a* dominates an alternative *b* if and only if $w_i v_i(a_i) \geqslant w_i v_i(b_i)$ for all *i* and for at least one *j*, $w_j v_j(a_j) > w_j v_j(b_j)$. Dominated alternatives can be discarded since at least one alternative is always preferred to them. The idea is now to introduce a similar concept for moral versus other attributes. A decision matrix is *fully moral* if and only if for all alternatives *a, b* the following condition holds:

$$\text{If } \sum_{a_i \in M} w_i v_i(a_i) > \sum_{b_i \in M} w_i v_i(b_i), \text{ then } v(a) > v(b) \tag{1}$$

A decision maker whose decisions always satisfy (1) never makes any moral mistakes, is a perfect moral decision maker and perhaps also a moral tyrant in the sense laid out above. There is something eerily wrong about such a person. Since the antecedent holds for arbitrary alternatives, non-moral values only play a role in her decision if the weighted sums of the moral attributes in the antecedent are exactly equal, i.e. if the alternatives have exactly the same moral consequences. Otherwise they can be discarded. The moral attributes absolutely dominate the non-moral attributes.

It might be tempting to reply that a decision maker acts morally as long as *most* decisions satisfy (1), but as mentioned in the beginning of this section, this kind of reply is not acceptable without further elaborating the role of the stakes. A murder cannot be excused by the fact that the murderer acts morally most of the time, even though this fact might count in his favour in court. So it seems that the condition must be relaxed in another way. One solution is to stipulate, as an additional constraint, that the sum of the moral attributes of the first alternative in the antecedent must be higher than a certain threshold. In other words, we are looking for conditions to further narrow the set of attributes that is morally relevant in general down to a smaller set of attributes that are morally relevant in a particular decision situation based on the values of the attributes in question. A decision matrix is *moral relative*

to global threshold α if and only if the following condition holds for all alternatives *a, b*:

$$\text{If } \sum_{a_i \in M} w_i v_i (a_i) > \alpha \text{ and } \sum_{a_i \in M} w_i v_i (a_i) > \sum_{b_i \in M} w_i v_i (b_i), \text{ then } v(a) > v(b) \quad (2)$$

However, this condition will only work as desired as long as it can be ensured that any value of an attribute or combination of attributes that is considered morally relevant exceeds the threshold. Not only is it hard to conceive how a reasonable moral theory could provide such a threshold, but it would also be problematic to elicit such a joint feature of attributes from a person or expert panel. It seems more reasonable to stipulate thresholds for individual attributes instead. A decision matrix is *moral relative to thresholds α$_i$* if and only if the following condition holds for all alternatives *a, b*:

$$\text{If there is an } \alpha_i \; s.t. w_i v_i (a_i) > \alpha_i \text{ and } \sum_{a_i \in M} w_i v_i (a_i) > \sum_{b_i \in M} w_i v_i (b_i),$$
$$\text{then } v(a) > v(b) \quad (3)$$

where α_i is the threshold of the *i*-th attribute a_i. If just one moral attribute exceeds the threshold, then the whole set of moral attributes becomes relevant.

This principle seems to reflect the way in which a person's actions are often judged retrospectively, for example in court or public opinion, and incorporates a certain virtue-ethical stance. For example, when someone is accused of a crime, judges and plaintiffs sometimes take into account moral aspects of additional motives, for instance whether the crime has been committed out of need or sheer selfishness. The goal is hereby to determine the character of the accused and the final verdict hinges to some extent on this assessment. Despite being common practice, this method seems questionable for the present, more general purpose. If a threshold plays the role of determining moral relevance, it seems that an attribute below the threshold ought not enter moral considerations, or otherwise it is no longer clear what the threshold actually does.

By varying the condition slightly, a more appropriate evaluation method can be obtained. Let *T* be the set of thresholds in the model indexed in the same way as the set of moral attributes, and let M_a^T

denote the set of indices i of attributes of alternative a in M such that $w_i v_i (a_i) > \alpha_i$. The antecedent condition is then relativized to this set. A decision matrix is *selectively moral relative to a set of thresholds T* if and only if the following condition holds for all alternatives a, b:

$$\text{If } M_a^T \neq \varnothing \text{ and } \sum_{i \in M_a^T} w_i v_i (a_i) > \sum_{i \in M_a^T} w_i v_i (b_i), \text{ then } v(a) > v(b) \qquad (4)$$

In this condition, only moral attributes whose values exceed their respective threshold are considered relevant in a specific decision situation. Once these attributes have been identified, they are weighed against other moral attributes in the set.

To get an idea of how this condition works, consider the following trivial example. Suppose Bob is contemplating whether he should steal his cousin's chocolate bar and suppose, furthermore, that he does not fear reprimand because his cousin does not keep track of his huge inventory of chocolate bars. Let feature 1 be an artificial measure that is higher when no chocolate is stolen.[9] On the other side of the equation is Bob's personal pleasure, represented by feature 2. For simplicity, all weights are taken to be 1 in this and the following examples and the threshold of the moral attribute 1 in this example is $\alpha_1 = 0.3$. Suppose the following matrix represents his decision:

	a	b
1 – not steal	0.4	0.2
2 – personal pleasure	0.2	0.6
Moral sum	**0.4**	0.2
Total sum	0.6	**0.8**

As a rational decision maker, Bob decides to steal the chocolate bar. Determining whether his decision is morally permitted is straightforward. First, mark all rows with moral attributes. This is the first row in this example. Compute their 'moral sum' in each column and underline the highest values in the moral sum row. In this case the winner

---

9   A better way to model this situation will be laid out in the next section. For the time being, you may, for example, assume that any value above 0.38 indicates that no good has been stolen.

is *a* with value 0.4. Second, compute the total sum and underline the highest values in that row too. In this case, the winner is *b* with value 0.8. Evaluation: If a value is underlined in a total sum column and not in the corresponding moral sum column, then the decision is non-moral; otherwise it is moral. This is just in ordinary terms what Condition (4) says. Using this procedure, it becomes apparent that Bob's decision in the above example is non-moral. In virtue of being non-moral, the decision is also immoral in this case because a morally preferable course of action was available and could have been taken.

   One might think that there is an easier way to evaluate such an example. Instead of summing up the values of moral attributes, as condition (4) prescribes, one might replace the antecedent by individual comparisons, i.e. all moral attributes must satisfy $w_i v_i (a_i) > w_i v_i (b_i)$ in the antecedent. This modified principle would, however, predict that *any* action based on the following matrix is immoral:

	a	b
1	0.7	0.2
2	0.3	0.8
Moral sum	**1.0**	**1.0**
Total sum	**1.0**	**1.0**

In this example two different moral attributes outweigh each other. Consequently, any of *a* or *b* ought to be just as good from a moral point of view, which is correctly predicted by Condition (4) and the corresponding informal evaluation procedure outlined above. The shortcut version does not make the correct prediction here as it only takes into account the relations between individual moral attribute comparisons and the total sum.

   A final abstract example illustrates a mixed case with three alternatives and motivates our talk about decision tables as opposed to actual decisions. These tables encode additional information that may sometimes turn out to be useful for an assessment.

	a	b	c
1	0.7	0.2	0.4
2	0.3	0.7	0.3
3	0.2	0.5	0.7
Moral sum	0.9	0.7	**1.1**
Total sum	1.2	**1.4**	**1.4**

The rule predicts that *a decision based on this table* is immoral because there are two highest values in the total sum row, but only row *c* has a corresponding highest moral sum. However, if the decision maker were to *actually choose* alternative *c* his choice would be morally impeccable on purely consequentialist grounds. However, the underlying table is, at least, problematic from a moral point of view as he could just as well have chosen alternative *b*. Even if the decision maker actually chose *c*, he might not have done so for the proper motive. For example, he could have made the choice by throwing a coin. This example shows that moral applications of decision theory need not be solely consequentialist in nature even when they involve the weighing of alternatives that represent different possible courses of action.

2.2 The Bipolar Threshold View

The above way of laying out decision problems is clumsy to say the least. In the first example, an artificial attribute 'not steal' is used to express the moral value of not stealing instead of expressing the *disvalue* of stealing something. The reason for this was that the thresholds were formulated for positive values. If a moral attribute's value exceeded a threshold, the value was considered morally relevant. Simple additive models may also include negative values, but then the threshold view must be adjusted. The resulting model is *bipolar* as it distinguishes between negative and positive values. Let the set M_a^T contain index i if and only if (Case 1) α_i is a positive threshold and $w_i\, v_i\, (a_i) > \alpha_i$, or (Case 2) α_i is a negative threshold and $w_i\, v_i\, (a_i) < \alpha_i$. Apart from this, no further changes are needed and Condition (4) remains intact.

With this adjustment, disvalues with a corresponding negative moral threshold can be used. Although the change is minimal, its conceptual relevance is huge as now the theory may be used to express positions like *Negative Utilitarianism*. For example, one might believe that it ought to be allowed to cause small amounts of harm to persons without the harm becoming morally relevant. Everyone does this almost every day, for example when arguing or having a bad day, and people are not always harmed for the sake of a greater good. It would amount to moral tyranny to consider any kind of harm done to a person

morally reprehensible. After all, not many executives of a company can do their job properly without occasionally causing displeasure among their employees. What matters is whether the amount of pain exceeds a certain threshold, which is described by the bipolar threshold view. Negative Utilitarianism has been attacked by Smart (1958) and recently by Ord (2013), who gives an excellent overview of the main arguments against it. While the details of this debate are beyond the scope of this article, it is noteworthy that this particular version of what Ord calls 'Lexical Threshold View' fares better than most other variants of negative utilitarianism. Ord considers the sudden change in evaluation once a threshold is reached implausible for small, perhaps even arbitrary increments of disvalue and constructs a form of sorites paradox against it (his 'Continuity Argument'). I believe Ord's argument does not speak conclusively against the above version of the threshold view but would like to leave this matter for another occasion.[10]

2.3 Dealing with Stakes

Let me finally address the role of stakes in decision making under risk and uncertainty very briefly. The bipolar threshold view does not require many modifications in this setting. Basically, $w_i v_i (a_i)$ must be replaced by $p_j w_{i,j} u_i (a_{i,j})$ in the above conditions. The threshold is applied to the weighted outcome times the probability of the occurrence of a consequence. Although Savage (1954) has shown, within a different yet sufficiently similar setting, that some intuitively plausible principles governing preferences and subjective plausibility imply that the factors $\sum_{j=1}^{m} p_j$ indeed constitute, and so subjective expected utility theory as a whole is derivable from few, intuitively compelling principles, there

---

10   There is a strong argument against any view that rejects lexical threshold utilitarianism while endorsing classical utilitarianism. From a purely formal point of view, it seems possible to translate a threshold model into the classical approach by choosing suitable utility functions of potentially unusual shape. If that line of thinking is correct, as I believe it is, then thresholds are merely a convenient way of making decision boundaries explicit, which is what makes them useful for the purpose of moral decision making.

are legitimate concerns about moral implications of always using expected utility. What is sometimes treated as a risk might in reality be based on some form of epistemic uncertainty. In that case, the Precautionary Principle might dictate that maximal losses ought to be minimized. According to this so-called *Maximin* method, instead of multiplying a weighted utility with a probability, the alternative with the smallest loss is the most preferable one. It seems that this method could be justifiable for moral attributes on moral grounds alone as long as the risk in question really is uncertainty in disguise.[11] There are other decision principles such as *Minimax with Regret* to consider if the stakes are high and some form of uncertainty is at play. It would go beyond the scope of this article to enter this debate, but I cannot see any principal obstacles that would keep us from adjusting the bipolar threshold view to such alternative decision principles, and conditions like (4) seem to be applicable to these as well.

Concluding Remarks

Several conditions for evaluating the morality of a decision on the basis of given sets of morally relevant and non-relevant attributes and corresponding thresholds have been investigated in a (simplified) decision-theoretic setting. Among these the bipolar threshold view of Condition (4) turned out to be the most general and adequate. Once moral attributes and their thresholds have been identified, this condition could be used to give moral advice. For example, in a group decision making process, a team of 'moral experts' could provide the value functions and thresholds for moral attributes, which may then be combined during an argumentative decision making phase with the outcome of a preference

---

11   One might argue that this problem just concerns the possibility of incorrect modelling, yet prescriptive theory ought to be based on the assumption that the modelling is correct. However, if the stakes are high enough, a 'Meta-Precautionary Principle' might no longer be clearly discernible from the normal one.

elicitation process of another group consisting of experts, such as public health professionals, about the specific application domain.

Although this suggestion must be taken with a grain of salt in light of the simplifying assumptions mentioned in Section 2, it is worth noting that most of the known criticism of these assumptions apply to decision theory in general and have already been addressed in great detail in the seminal literature. For instance, worries about preferential independence, completeness and transitivity of the underlying preferences are addressed by Fishburn (1991), Hansson (2001) and in contributions to Bouyssou et al. (2010), and a number of alternatives to the additive decomposition of value functions such as multilinear models, multiplicative models and generalized additive decomposition have been on the table for a long time. Within this spectrum, the present account is a lexicographic outranking method, mixed with additive models for simplicity.

Other kinds of criticism based on the fact that we do not make decisions in the way prescriptive decision theory mandates are also well known. See, for instance, Kahneman & Tversky (1979; 2000), arguments by Broome (1999: Ch. 6) for Bolker-Jeffrey utility theory, and the Maximin and Minimax with Regret approaches mentioned above. As suggested in the last section, it seems possible to adjust the bipolar threshold view to such alternative frameworks, but perhaps this is not needed. Some of the approaches mentioned above, such as work by Kahneman and Tversky and work in the evolving field of 'ecological rationality', draw their motivation from empirical aspects of decision making, and there are doubts whether these successfully undermine the normativity of expected utility theory that is established by intuitively plausible rationality postulates. I am personally wary of any account of rationality whose justification is mainly derived from empirical success criteria. Be that as it may, the matter seems to be undecided even among scholars of decision theory.

There is another worry about the threshold view that has to be mentioned. Thresholds constitute sharp boundaries that in practice might make the approach unfruitful. If thresholds for moral attributes are always so low that the cases when they are not relevant are utterly trivial, then the differences between (1) and (4) might become uninteresting. Perhaps the stakes are always high in situations worthy of

thorough decision analysis. Whether this is a problem or not can only be decided in practice.

Finally, one might wonder whether the theory laid out so far could serve as a basis for a utilitarian calculus. The answer is clearly no. First, there are elements of virtue ethics in the account; not only the actual decision counts for an assessment, but the whole decision table on the basis of which it has been made. Second, a fully-fledged utilitarian calculus requires much more than what multi-criteria decision theory can offer. Time has to be taken into account for there is no doubt that many people prefer short amounts of intense displeasure (pain, inconvenience, dissatisfaction) over a long period of lesser suffering and, vice versa, are sometimes willing to tolerate extended amounts of displeasure – morally acceptable variants of the so-called 'necessary evil' – to later obtain a greater good. If at all, the bipolar threshold view can only become a small part of such an approach, allowing one to consider single decision situations from a moral aspect provided that the relevant alternatives including their consequences over time can be identified and an appropriate connection between attribute values, their thresholds and moral rules can be drawn. What such a connection would look like and how it fares with known methods of preference elicitation is a surprisingly open question of moral philosophy though.

Third, while decision making can account for the influence of possible courses of actions on other people, the simple version laid out above cannot represent equilibria between preferred choices of several decision makers. Such equilibria between morally relevant values would have to be the cornerstone of a fully-fledged utilitarian calculus and have not been addressed at all. As is well known, there are a number of obstacles to such an ambitious project, ranging from the question of how to 'tame' respective variants of Arrow's theorem to more philosophical worries about value aggregation and the interpersonal comparability of utilities. It is also well known that Pareto optimality used in economics allows for social states that are inherently unjust, and so a utilitarian calculus worth being taken seriously would have to take into account additional justice criteria. Variants of decision theory like the one laid out above may help in making morally acceptable decisions, but in a social context only on the basis of an existing theory of justice. By itself, decision theory does not contribute to such a theory and also cannot provide the criteria for deciding at what level attributes become morally relevant.

References

Broome, John (1999). *Ethics out of Economics.* Cambridge: Cambridge University Press.

Broome, J. (2002). 'Practical Reasoning', in *Reason and Nature: Essays in the Theory of Rationality*. Oxford: Oxford University Press.

Chang, R. (1997) (ed.). *Incommensurability, Incomparability, and Practical Reason*. Cambridge (MA): Harvard University Press.

Chang, R. (2002). 'The Possibility of Parity.' *Ethics,* Vol. 112: 669–688.

Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. New Haven, London: Yale University Press.

Eisenführ, F., M. Weber, & T. Langer. (2010): *Rational Decision Making.* Berlin: Springer.

Fishburn, P. C. (1970). *Utility Theory for Decision Making.* New York: John Wiley & Sons.

____ (1991): 'Nontransitive additive conjoint measurement.' *Journal of Mathematical Psychology,* Vol. 35: 1–40.

Gert, J. (2004). 'Value and Parity.' *Ethics,* 114: 492–510.

Gossen, H. H. (1854). *Entwicklung der Gesetze des menschlichen Verkehrs und der daraus fließenden Regeln für menschliches Handeln.* Wiesbaden: Vieweg.

Hansson, S. O. (2001). *The Structure of Values and Norms*. Cambridge: Cambridge University Press.

Hilpinen, R. & P. McNamara (2013). 'Deontic Logic: A Historical Survey and Introduction', in Gabbay, D., J. Horty, X. Parent, R. van der Meyden, & Leendert van der Torre. *Handbook of Deontic Logic and Normative Systems*. London: College Publications.

Horowitz, J., List, J., & K. E. McConnell (2007). 'A Test of Diminishing Marginal Value.' *Economica,* Vol. 74: 650–663.

Horty, J. F. (2012). *Reasons as Defaults*. Cambridge: Cambridge University Press.

Inglehart, R. (1990). *Culture Shift in Advanced Industrial Society.* Princeton: Princeton University Press.

Kahneman, D., & A. Tversky (1979). 'Prospect Theory: An Analysis of Decision under Risk'. *Econometrica*, Vol. 47: 263–291.

_____ (eds.) (2000). *Choices, Values and Frames*. Cambridge: Cambridge University Press.

Keeney, R. L., & H. Raiffa, H. (1976, 1993). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: John Wiley & Sons. Cit. in 1993 edition published by Cambridge University Press.

Krantz, D. H.; R. D. Luce, P. Suppes, & A. Tversky (1971), *Foundations of Measurement*, Vol. I. New York: Academic Press.

Liu, C. (2003). 'Does Quality of Marital Sex Decline With Duration?' *Archives of Sexual Behavior,* Vol. 32, No. 1: 55–60.

Meinong, A. (1894). *Psychologisch-ethische Unterschungen zur Werttheorie.* Leipzig: Leuschner & Lubensky.

Ord, T. (2013). 'Why I'm Not a Negative Utilitarian', University of Oxford. Published online at http://www.amirrorclear.net/academic/ideas/negative-utilitarianism/index.html. Accessed February 20 2014.

Rabinowicz, W. (2010). 'Value Relations – Old Wine in New Barrels', in: Reboul, A. (ed.), *Philosophical Papers Dedicated to Kevin Mulligan.* Université de Genève. Published online at http://www.philosophie.ch/kevin/festschrift/.

Ramsey, F. P. (1931). 'Truth and Probability', in *The Foundations of Mathematics and other Logical Essays*. London: Kegan, Paul, Trench, Trubner & Co.: pp. 156–198.

Sartre, J.-P. (1946). *L'existentialisme est un humanism*. Paris: Éditions Nagel.

Savage (1954). *The Foundations of Statistics.* New York: Wiley.

Schoemaker, P. J. H. (1982). 'The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations.' *Journal of Economic Literature,* Vol. 20 (June): 529–563.

Smart, R. N. (1958). 'Negative Utilitarianism', *Mind* 67: 542–3.

von Neumann, J., & O. Morgenstern (1947). *Theory of Games and Economic Behavior.* Princeton: Princeton University Press.

von Wright, G. H. (1963). *The Varieties of Goodness*. London: Routledge & Kegan Paul.

Moral Relativism and Perspectival Values

Pietro Gori and Paolo Stellino

Although the term 'relativism' entered the philosophical vocabulary as a *terminus technicus* only in the nineteenth century,[1] the philosophical position known as relativism can be traced back to Ancient Greek philosophy. As is known, the fundamental proposition of Protagoras of Abdera was that 'man is the measure of all things: of the things which are, that they are, and of the things which are not, that they are not.' (Plato, *Theaetetus*: 152a) Socrates' refusal of Protagoras' proposition in the *Theaetetus* has led and still leads many philosophers to think that relativism is self-refuting:

> [Protagoras' doctrine] has this most exquisite feature: Protagoras admits, I presume, that the contrary opinion about his own opinion (namely, that it is false) must be true, seeing he agrees that all men judge what is … And in conceding the truth of the opinion of those who think him wrong, he is really admitting the falsity of his own opinion? … But for their part the others do not admit that they are wrong? … But Protagoras again admits *this* judgement to be true, according to his written doctrine? … It will be disputed, then, by everyone, beginning with Protagoras – or rather, it will be admitted by him, when he grants to the person who contradicts him that he judges truly – when he does that, even Protagoras himself will be granting that neither a dog nor the 'man in the street' is the measure of

---

1   Maria Baghramian (2004: 11) points out that the first use of the term 'relativism' can be traced to John Grote's *Exploratio Philosophica* (1865). Mi-Kyoung Lee (2005: 34), for his part, mentions an earlier use of the word in writings of Sir William Hamilton and puts forward the hypothesis that the term entered the English language from the German use of 'Relativismus.' As a matter of fact, as Bernd Irlenborn (2016: 7–8) indicates, the word 'Relativismus' can be already found in the fifth volume of Wilhelm Traugott Krug's *Allgemeines Handwörterbuch der philosophischen Wissenschaften*, dating from 1838. Lee (2005: 34) also points out that 'Relativismus' was the term used by nineteenth-century neo-Kantian German philosophers and scholars to refer to the position that nothing can be known in itself, and that all we can know are appearances.'

anything at all which he has not learned. Isn't that so? … Then since it is disputed by everyone, the *Truth* of Protagoras is not true for anyone at all, not even for himself! (Plato, *Theaetetus*: 171a-c)

Socrates draws attention to the fact that, if man is the measure of all things and, therefore, truth is relative to man, then Protagoras must concede the truth of the opinion contrary to his own doctrine, namely the opinion according to which it is false that man is the measure of all things. By doing so, however, Protagoras would be contradictorily committed to both the truth and falsehood of his own doctrine. In order to avoid falling into this contradiction, Protagoras must assume that there is at least one absolute truth, that is, the truth of the proposition 'man is the measure of all things.' But then, once again, this could be seen as a contradictory move, for Protagoras would be maintaining at the same time that all truth is relative and that there is – at least – one absolute truth, namely, that all truth is relative.[2]

As Neil Levy (2002: 19) has pointed out, unlike epistemic relativism, moral relativism is not vulnerable to the contradiction argument. Indeed, no contradiction is involved in claiming that 'moral claims are true only relative to some standard or framework' since this is not itself a moral claim. Even so, moral relativism faces other difficulties. Above all, opponents of moral relativism claim that if moral relativism is true, then we have no means to condemn morally actions that we find profoundly reprehensible or immoral if these actions are performed by members of a different culture than ours. This claim usually takes the form of a slippery slope argument: if we recognize that *a* (we do not have any absolute moral standards in the name of which we can denounce reprehensible or immoral actions), then the result *b* (the way is open for any kind of crime) inevitably follows. Moral relativism would thus fatally undermine morality: if moral relativism is true, so the criticism goes, then anything goes, that is, everything is permitted. But is it really so? Does this way of framing the problem really capture the subtleties and complexities of moral relativism?

It is interesting to notice how the terms in which the debate between moral relativists and moral absolutists is phrased recall the way

---

2   For a more detailed analysis, see Baghramian (2004: 18–31).

in which philosophers have interpreted and still interpret the relation between Dostoevsky and Nietzsche. Needless to say, the argument 'if moral relativism is true, then everything is permitted' has a clear Dostoevskian flavour. In Dostoevsky's last novel *The Brothers Karamazov*, Ivan, one of the brothers Karamazov, puts forward the following idea: if there is no God and if there is no immortality of the soul, then everything is permitted. The parricide, around which the novel revolves, can be considered as a consequence of this idea, whereas the novel itself can be regarded as a grandiose response to it.

Ivan's idea bears a striking similarity to the maxim 'nothing is true, everything is permitted' that appears in Nietzsche's *Thus Spoke Zarathustra*, in *On the Genealogy of Morality*, as well as in some posthumous fragments from 1884 and 1885. This similarity has not gone unobserved and, beginning from the end of the nineteenth century, Russian and European intellectuals have taken it as the key to read the relation between Dostoevsky and Nietzsche. The maxim 'nothing is true, everything is permitted' has been removed from context and read as summing up the core of Nietzsche's philosophy. This has led to the controversial identification of Nietzsche's moral perspectivism with Ivan's moral indifferentism. As a result, Dostoevsky's novel *The Brothers Karamazov* has been seen as anticipation and critique *ante litteram* of Nietzsche's perspectival philosophy.

Beyond the question of the philological and philosophical adequacy of this kind of interpretation,[3] what should not be overlooked here is the logic underlying this kind of reading. Far from questioning whether the maxim 'nothing is true, everything is permitted' could be taken as summing up the message of Nietzsche's philosophy, intellectuals have taken for granted that the logical and inevitable conclusion following from Nietzsche's moral perspectivism (essentially read as a moral relativism) was that 'everything is permitted.' As one can see, what we have is, once again, the argument according to which, if moral relativism – or, in Nietzsche's case, moral perspectivism – is true, then everything is permitted.

In what follows, we will tackle this argument. More specifically, we will take Nietzsche's case as paradigmatically showing that a

---

3   On this, see the second part of Stellino (2015a).

relativization or *perspectivizing* of morality does not imply *ipso facto* that anything goes or that everything is permitted. In order to do this, we will consider two assumptions which are often made uncritically: (1) Nietzsche's moral perspectivism is essentially a moral relativism, and (2) the practical consequence deriving from Nietzsche's moral perspectivism is that everything is permitted.[4]

1. Moral Perspectivism

One of the aspects of Nietzsche's thought that in recent years has catalysed the attention of many scholars is 'perspectivism.' Although occurrences of this term are limited in number (at least, if we take into consideration only the *oeuvre*[5]) and time (they appear almost exclusively in the late period), this notion has been taken as indicating one of the fundamental theories of his philosophy. The reason for the importance that many scholars have given to this notion lies in the fact that perspectivism is considered as a key term used by Nietzsche to define, in a more synthetic and incisive way, his theory of knowledge. Within this context, scholars often focus on a famous passage from GM III 12 ('There is *only* a perspectival seeing, *only* a perspectival 'knowing''), but scarcely consider GS 354, where Nietzsche relates what he considers 'to be true perspectivism' with the *morally-oriented* view of

---

4   Nietzsche's works are cited by abbreviation, chapter (when applicable) and section number. The abbreviations used are the following: BT (*The Birth of Tragedy*), HH (*Human, All Too Human*), GS (*The Gay Science*), Z (*Thus Spoke Zarathustra*), BGE (*Beyond Good and Evil*), GM (*On the Genealogy of Morality*), TI (*Twilight of the Idols*), EH (*Ecce Homo*). The translations used are from the Cambridge Edition of Nietzsche's works. Posthumous fragments (PF) are identified with reference to the Colli & Montinari standard edition. The fragments which do not appear in the Cambridge Edition of the *Writings from the Late Notebooks* are translated according to the Kaufmann and Hollingdale edition of *The Will to Power* (see *References*).

5   See, particularly, BT, *An Attempt at Self-Criticism* 5; HH I, *Preface* 6; BGE, *Preface* and sections 11 and 34; GM III 12; FW 354 and 374.

the herd community. Similarly, the posthumous note 7[60], 1886–87 is often mentioned, but most of the time it is misleadingly and arbitrarily quoted with no reference to its context.[6] Since in this note Nietzsche's aim is to criticize the attitude of positivism, scholars interpret the claim that 'there are no facts, only interpretations' – a statement often considered as summing up Nietzsche's perspectivism – as being exclusively linked to epistemology. By so doing, the same scholars ignore that Nietzsche had already published the maxim in section 108 of *Beyond Good and Evil* (which chronologically predates the posthumous note 7[60]) and that, in that book, the maxim was specifically referred to *moral* phenomena: 'There are absolutely no moral phenomena, only a moral interpretation of the phenomena …'[7]

This does not imply the rejection of the many interpretations that give preference to the epistemological character of Nietzsche's reflections on perspectivism.[8] Still, it is important to point out that, although Nietzsche's perspectivism is grounded on a specific epistemological view, the former cannot be reduced to the latter. Nietzsche himself suggests this idea, for instance, when he argues that our fundamental 'will to truth' forces us to recognize that 'it is no more than a *moral prejudice* that the truth is worth more than appearance' (BGE 34; our italics). On the contrary, Nietzsche writes, we have to acknowledge that 'life could not exist except on the basis of perspectival valuations and appearances.'

The maxim in BGE 108, which can be taken as the 'motto' of Nietzsche's moral perspectivism, reappears two years later and in a slightly different way in the following passage from *Twilight of the Idols*:

---

6   For a thorough examination of this note, see Gori (2016: chapter 2).

7   See also PF 1885–86, 2[165]: '*My main proposition: there are no moral phenomena, there is only a moral interpretation of those phenomena. This interpretation is of extra-moral origin.*' We can find perspectivism and morality strictly related in other posthumous fragments of Nietzsche's (e.g. PF 1884, 26[178] and 1885–86, 2[206]). In PF 1887, 10[154], Nietzsche writes: '*My intention* to show the absolute homogeneity in all that happens and the application of the moral distinction as only *perspectivally conditioned.*' According to Robert C. Solomon (2003: 46), 'Nietzsche's 'perspectivism' is most in evidence and most at issue in his moral philosophy.'

8   On this, see, among others, Clark (1990) and Leiter (1994).

> You have heard me call for philosophers to place themselves *beyond* good and evil,
> – to rise *above* the illusion of moral judgement. This call is the result of an insight
> that I was the first to formulate: *there are absolutely no moral facts*. What moral and
> religious judgements have in common is the belief in things that are not real. Mo-
> rality is just an interpretation of certain phenomena or (more accurately) a *mis*inter-
> pretation. Moral judgements, like religious ones, presuppose a level of ignorance in
> which even the concept of reality is missing and there is no distinction between the
> real and the imaginary; a level where 'truth' is the name for the very things that we
> now call 'illusions'. That is why moral judgements should never be taken literally:
> on their own, they are just absurdities. (TI, *'Improving' Humanity* 1)[9]

According to Nietzsche, to deny the very existence of moral facts (or
phenomena) means to deny the possibility of claiming that the same
facts (or phenomena) are *intrinsically* moral. In Nietzsche's view, re-
ality is morally *neutral*. To believe that there are moral realities is
the consequence of an illusion: what we do have is the existence of
facts or phenomena, to which a moral interpretation is added by us
depending on the specific moral perspective from which we judge.
According to Nietzsche, the moral character of an action has thus not
been *found* or *discovered*, but rather *introduced* in the action by the
human being.

Here we face the question of the so-called *Sinn hineinlegen*, i.e. the
'introduction of meaning' into the world. As Nietzsche puts it in a well-
known passage from section 301 of *The Gay Science*:

> It is we, the thinking-sensing ones, who really and continually *make* something
> that is not yet there: the whole perpetually growing world of valuations, colours,
> weights, perspectives, scales, affirmations, and negations. … Whatever has *value*
> in the present world has it not in itself, according to its nature – nature is always
> value-less – but has rather been given, granted value, and *we* were the givers
> and granters! Only we have created the world *that concerns human beings*! But
> precisely this knowledge we lack, and when we catch it for a moment we have
> forgotten in the next.

As this passage clearly shows, Nietzsche maintains a *projectivist* stance
on valuations.[10] The world appears to be valuable and meaningful

---

9   This passage is often quoted in order to support a reading of Nietzsche's metaeth-
     ics in the light of J.L. Mackie's 'error theory' (see, for instance, Hussain 2007).
10   On this, see Stellino (2015b: 182–184).

because human beings previously gave value and meaning to a value-less and meaningless world. In other words, they projected moral, aesthetic, religious and other kinds of valuations and estimations onto it. By so doing, they created a perspectival and anthropomorphic world and then forgot about their creation, wrongly believing the world to be intrinsically beautiful and meaningful.[11]

The awareness of the intrinsic meaninglessness of the world strongly characterizes Nietzsche's late philosophical thought. Whereas philosophers so far searched for a meaning *of* or *in* the world, Nietzsche becomes conscious that meaning or value has to be *created*. This creation opens up new, unexplored possibilities for the human being: this is the ultimate meaning of the metaphors of the 'new dawn' and the 'open sea' that Nietzsche uses in order to describe the free spirit's reaction to the news that 'the old God is dead' (GS 343). 'The world has once again become infinite to us,' Nietzsche writes in another section of the fifth book of *The Gay Science*, 'insofar as we cannot reject the possibility *that it includes infinite interpretations*' (GS 374).

A superficial reading of Nietzsche's philosophy could take these passages and metaphors as a confirmation that the reasoning mentioned above – according to which, if moral perspectivism is true, then everything is permitted – is validated by Nietzsche himself. As a matter of fact, if, according to Nietzsche's moral perspectivism, (i) every moral interpretation is relative to a judging perspective, and (ii) God is dead, that is, an absolute viewpoint (God's eye view) is lacking, then (iii) every moral interpretation seems to be as true, valid or justified as the others. In other words, everything would be *permitted*. Following this reasoning, Nietzsche is often interpreted as a supporter of an extreme moral relativism as well as of a radical form of normative ethical egoism according to which, given God's death and the perspectival character of reality, moral agents ought to do what is their own self-interest, even if this means to act in detriment to others' interest. In what follows, attention will be briefly focused on both views.

---

11   See also PF 1884, 25[505].

2. Individualism *vs*. Relationalism

In arguing against the view that takes Nietzsche to be a supporter of a radical form of normative ethical egoism, the following premise is needed: it is undeniable that in Nietzsche's writings and posthumous notes one finds abundant textual evidence in favour of moral individualism. In a passage from *Thus spoke Zarathustra*, for instance, Nietzsche writes as follows: 'He will have discovered himself who speaks: 'This is my good and evil.' With this he has silenced the mole and dwarf who says: 'Good for all, evil for all'.' (Z IV, *On the Spirit of Gravity*) This individualistic attitude – a peculiar feature of Nietzsche's philosophy – acquires its full meaning when contrasted with Kantian universalism. This contrast, in particular, is symbolised by the second metamorphosis of the spirit, who first becomes a camel ('Thou shalt') and then a lion ('I will').[12]

It is because of his strong opposition to Kantian universalism that Nietzsche puts particular emphasis, in *Zarathustra* as well as in other writings, on the point of view of the individual in morality. This emphasis has been, however, interpreted in the sense of a radical and extreme form of individualism, which would directly follow from Nietzsche's perspectivism. Nevertheless, although Nietzsche often makes reference to the human tendency to subjugate and tyrannize – a tendency which is the expression of the fundamental feature of the world, the well-known and widely debated 'will to power' – the perspectival talk of a multiplicity of different and opposed perspectives leads to quite different outcomes. This becomes evident when attention is focused on the key question of the *subject* of perspectivism.[13]

Contrary to what one may be led to believe, most of the time Nietzsche does not identify the human being (the individual) as the proper subject of perspectivism; rather, he refers both to supra-individual subjects (e.g. the species or society) and to infra-individual subjects (e.g. the centres of force). Moreover, no matter which is the subject of perspectivism (the individual, the supra-individual or

---

12   Z, I, *On the Three Metamorphoses*. See also GS 355, A 11, and TI, *Morality as Anti-nature* 6.

13   On this, see Cox (1997).

the infra-individual one), the reality lying behind it is always plural and dynamic. This reality is characterised by the mutual relationship between its component parts, according to the view of nature that Nietzsche defends as from 1881.[14] Nietzsche's perspectivism is, therefore, grounded on a *relational model* with no privileged subject. Within this model, the validity of one specific perspective cannot be thought without any reference to the relation (be it conflicting or not) that this perspective entertains with other perspectives.

A brief scrutiny of the most interesting passages where Nietzsche talks about a 'perspectival seeing' can lend support to what has been argued. The wider subject of perspectivism that Nietzsche considers is the *species*, whose perspective on, or interpretation of, reality is shared by all the single individuals that have the same perceptive and cognitive apparatus. Nietzsche has in mind what we could define as a collective subject on a biological basis. During its evolutionary history, every species has developed a particular psycho-physiological structure which is functional to adaptation to the environment. Although each member of the species has a specific viewpoint of the world, she is still part of a wider interpreting perspective of reality which is the result of similar perceptive mechanisms.[15]

Nietzsche follows a similar line of thought when it comes to another wide subject of perspectivism, namely the *social collectivity*. In the well-known section 354 from the fifth book of *The Gay Science* – the only section of the published texts in which Nietzsche uses the term 'perspectivism' and explains what he considers to be 'true phenomenalism and perspectivism' – attention is focused on communication as a prerequisite for the creation of a society. In particular, Nietzsche points out that human consciousness 'actually belongs not to man's existence as an individual but rather to the community- and herd-aspects of his nature.' The herd is here the subject of a generalized and vulgarized

---

14   On Nietzsche's view of the world as an unresting dynamics of force-quanta in mutual relationship, and on the connection between this ontology and the idea of 'will to power,' see Abel (1998) and Gori (2007: chapter 3).

15   On this, see, e.g. GS 110; PF 1885, 43[1] and 5[36]; PF 1886, 7[2]. George Stack particularly focuses on the species as the main reference of Nietzsche's perspectivism. See, for instance, Stack (1991). See also Cox (1997: 274–275).

perspective, a dimension where any individual feeling and willing loses its value in favour of the usefulness to the herd.[16]

Species and society are two plural subjects of perspectivism in which individuality plays no fundamental role. On the contrary, when it comes to the *human being* – considered as the referent of a singular perspective determined not only by its space-time perception, but also by its specific interests and needs – individuality obviously has a more important position.[17] On this level, we have a multiplicity of singular perspectives pertaining to individual subjects whose fundamental tendency, according to Nietzsche, is to affirm their own worldview (their own 'taste') over those of the other subjects. This picture can particularly lead to the dangerous idea that Nietzsche is a supporter of a radical form of normative ethical egoism for, given this conflictive picture, moral agents could seem to be justified in doing what is their own self-interest, even if this means to act in detriment to others' interest.

Without denying that, in Nietzsche's view, individual perspectives conflict with each other and often tend to overmaster different or opposite perspectives, it should be pointed out that this interpretation suffers one serious flaw: it overlooks the constitutive character that *relationalism* plays in Nietzsche's perspectivism. As already mentioned, Nietzsche considers the individual as always making part of a species or a social collectivity. Within both of them, the individual is not like a monad, but is rather situated in a network of dynamic and interpersonal relations. Moreover, even when emphasis is put on the individual, it should not be forgotten that Nietzsche conceives the individual itself in terms of a plural multiplicity, a *collectivity*. This is evident, for instance, in *Beyond Good and Evil*, where Nietzsche's criticism towards the traditional view of the substantialist concept of 'subject' makes reference to 'social structures' like the *soul*, 'a society constructed out of drives and affects' (BGE 12), or the body, made of many souls from which the action that *we call* 'individual' arises (BGE 19).[18] Thus, behind the

---

16   See on this Ibbeken (2008: 75) and Gori (2016: chapter 3).

17   Among others, Clark and Leiter argue that Nietzsche's perspectivism is limited to human consciousness only. Their view is discussed in Cox (1997: 276 ff). On this, see also (Grimm 1977: 68).

18   See also PF 1880, 6[70]. According to Nietzsche, individuals are plural subjectivities made of drives and instincts acting at an 'unconscious' level (see, e.g. PF

individual, as well as behind the species and the social collectivity, there lies a network of relations between singularities, singularities that we ignore in favour of a more unitary and inclusive perspective.

What emerges from this picture is a plural conception of the human being: on the one side, we find a supraindividual (biological and/or social) perspective that includes the individual one; on the other side, there is the plane of the single entities that constitute the human being and that find in him a (merely illusory) unity.[19] Behind these entities, there is the *last* subject that it is possible to find in Nietzsche's writings, namely the *single centre of force*.[20] Here, Nietzsche leads perspectivism to the extreme, considering that the plane of interpretation coincides with that of being, that is, with the plane of pure and necessary relationship among the different perspectives, which can be defined only from within their mutual relation:

> As if a world would still remain over after one deducted the perspective! By doing that one would deduct relativity! Every center of force adopts a perspective toward the entire remainder, i.e., its own particular valuation, mode of action, and mode of resistance. … The 'world' is only a word for the totality of these actions. Reality consists precisely in this particular action and reaction of every individual part toward the whole. (PF 1888, 14[184])

As the analysis developed shows, when Nietzsche talks of 'perspectivism' or 'perspectival seeing,' he always has in mind a relational dynamics. The different interpretations of the world (be they of theoretical or moral nature) are all expression of this dynamics, on the basis of which the internal articulation of the most complex structures existing in the world is grounded. Everything is based on a non-teleological and necessary, but constitutively unstable, action-reaction process. Value judgments can be defined only by reference to this relationship, where, at the micro-level, a centre of force gains 'power' only insofar as it exchanges energy with

---

1885, 40[42]). Within this picture, the *I* (or the subject) is a non-substance entity, a theoretical notion whose ontological ground is only that of the pure activity that we attribute to it. In other words, the *I* is 'a perspectival *illusion* – the illusory unity in which, as in a horizon, everything converges' (PF 1885–86, 2[91]; on this, see also BGE 16, 17 and 19, and, for an examination of this issue, Gori 2015).

19   On this, see Cox (1997: 290).
20   See, among others, PF 1888, 14[184] and [186].

other centres. The kind of mastery grounded on this relationship is therefore not fixed and immutable. On the contrary, once the power of a centre of force is exhausted, the equilibrium of the total mass of energy changes and another centre becomes 'master' for a limited period of time.

The reference to this dynamic relationship avoids the risk of interpreting Nietzsche's moral perspectivism as leading to a form of autarchic individualism. The idea that different, conflicting interpretations can coexist follows indeed from the view according to which *relationship itself* is the constitutive element of a perspectival reality. In other words, we cannot define the centres of force outside their mutual relations or without making reference to the way they react to the obstacles they find when they discharge their energy. As a result, every perspective can affirm itself only through the relation with the other ones and, furthermore, *in alternation* with them. Thus, it would be wrong to think that, within Nietzsche's worldview, a specific evaluative perspective could be valid *in itself*, that is, in isolation from a context that gives to it its specific meaning, or to claim that one can affirm his own view over the others once and for all. This does not amount to any rejection of the individualistic and affirmative tendency pertaining to each perspective. Still, it is important to emphasize that this tendency must face the same attempt of affirmation from other subjects. In this way, conflicting perspectives give birth to a relational dynamics.[21]

3. Relativism

Nietzsche's moral perspectivism has been interpreted not only as an extreme form of ethical egoism, but also as a radical relativism according to which, as mentioned, since (i) every moral interpretation is relative to a judging perspective, and (ii) an absolute viewpoint is lacking, then (iii)

---

21 L. Hatab (1995: 160) argues that Nietzsche's pluralistic perspectivism is different from any other view that defends the coexistence of multiple 'truths' because it puts emphasis on the *agonal* dimension, that is, on the conflict existing between different perspectives. Nietzsche's perspectival view has been used by Günter Abel in order to develop an 'interpretation ethics' (see e.g. Abel 1995: chapter 24).

every moral interpretation seems to be as true, valid or justified (i.e. *permitted*) as the others. In order to understand why this kind of reading fails to capture the real meaning of the radical change that Nietzsche operates in the realm of morality, it is necessary to focus attention on the main goal of Nietzsche's late philosophy. The death of God announced by the madman of *The Gay Science* (§ 125), together with the collapse of the Christian-moral interpretation of the world, leave an axiological and normative void. Far from accepting this void as an inevitable existential condition, Nietzsche aims to face it 'fearless' and 'cheerful' (GS 343), and to fill it through the well-known *revaluation* of values. It is symptomatic, for instance, that although, on the one hand, Zarathustra (Nietzsche's alter ego) presents himself as 'the annihilator of morals' (Z I, *On the Adder's Bite*), on the other hand he puts strong emphasis on the need of creating new values. In other words, Nietzsche is well aware that a new evaluative interpretation must take the place of the former one, and much of his effort in the late period is focused on elaborating this new interpretation.

The attitude that, in the fifth book of *The Gay Science*, Nietzsche claims to be that of the new philosophers and 'good Europeans' shows us that, according to him, one of the consequences of the death of God is the opening of what Karl Jaspers has defined as a '*positive*, creative freedom' (Jaspers 1997: 157) for the human being. In the posthumous fragment 39[15], 1885, Nietzsche clearly writes that, with the death of God, the Christian-European morality has become no longer necessary (the Christian God and morality held themselves together, he claims). Once traditional morality has been denied validity, Nietzsche exhorts the human being to become a self-legislator, that is, to give himself new values and ideals and to set new goals (GS 335). In other words, man must become *autonomous*. This autonomy, however, is not to be conceived in terms of an unlimited or licentious freedom.[22] The consequence of the death of God is rather an assumption of both individual and collective responsibility.[23] This is a key point which Heidegger (2002 [1943]: 189) did not fail to notice, as the following passage clearly shows:

---

22   See on this e.g. Constâncio (2012).
23   On this, see Pfeuffer (2008).

It is easy but irresponsible to be outraged by the idea and the figure of the over-man, which was designed to be misunderstood; it is easy but irresponsible to pretend that one's outrage is a refutation. It is difficult but for future thinking unavoidable to attain the *high responsibility* [*hohe Verantwortung*; our italics] out of which Nietzsche reflected on the essence of that humanity destined … to undertake mastery over the earth. The essence of the overman is not a warrant for a fit of capricious frenzy. It is the law, grounded in being itself, of a long chain of the highest self-overcomings.

With these words, Heidegger gets at the heart of the problem: the axio-logical and normative void left by the death of God and by the collapse of the Christian-moral interpretation of the world is not conceived by Nietzsche as a 'warrant for a fit of capricious frenzy,' to use Heidegger's words. On the contrary, as already mentioned, Nietzsche calls humanity to an assumption of individual and collective responsibility, that is, to an attainment of the awareness that, since humanity's great values and ideals have proved to be hollow, new values and ideals are now required, i.e. must be created.[24]

This is the chief reason for which Dostoevsky's and Nietzsche's an-swer to the question of the consequence of the death of God for moral-ity could not be more opposite. As *The Brothers Karamazov* exemplary shows, Dostoevsky believes that God's existence and the immortality of the soul are two essential pillars of the moral edifice. Without them, what we have is a dangerous slope that leads from atheism to self-deification, and from self-deification to the breaking of all moral rules. This logic becomes evident in the following passage from the dialogue between the devil and Ivan Karamazov (fourth part of the novel):

Once mankind has renounced God, one and all … then the entire old world view will fall of itself, without anthropophagy, and, above all, the entire former moral-ity, and everything will be new. … Man will be exalted with the spirit of divine, titanic pride, and the man-god will appear. … The question now … is whether or not it is possible for such a period ever to come. If it does come, then everything will be resolved and mankind will finally be settled. But since, in view of man's inveterate stupidity, it may not be settled for another thousand years, anyone who already knows the truth is permitted to settle things for himself, absolutely as he wishes, on the new principles. In this sense, 'everything is permitted' to him. Moreover, since God and immortality do not exist in any case, even if this period

---

24   See PF 1887, 11[411].

should never come, the new man is allowed to become a man-god, though it be he alone in the whole world, and of course, in this new rank, to jump lightheartedly over any former moral obstacle of the former slave-man, if he need be. There is no law for God! (Dostoevsky 1992: 648f.)

Unlike Dostoevsky, for Nietzsche the dichotomy 'either God or amorality' is a false dichotomy. Aware that, to put it with Kant (1998 [1786]: 12), 'without any law, nothing – not even nonsense – can play its game for long,' Nietzsche is far from being a supporter of the thesis 'everything is permitted,' at least when this thesis is understood as an absolute lack of laws and values. If so understood, this thesis leads indeed to the nihilistic attitude that Nietzsche diagnoses in the European culture of his own age (with its degenerative effect on humanity) and to whose opposition a large part of his late writings and *Nachlass* is dedicated. On the contrary, as one can read, e.g. in *On the Genealogy of Morality* (III, 27), Nietzsche shows a clear awareness of the fact that European nihilism can and has to be countered with a revaluation of values. This is the groundbreaking task that Nietzsche decides to face, as he himself confesses in his autobiography: 'I have a hand for switching *perspectives*: the first reason why a 'revaluation of values' is even possible, perhaps for me alone' (EH, *Why I Am So Wise*, 1).

One of the fundamental conditions of the new 'doctrine and counter-evaluation of life' to which Nietzsche makes reference in the new preface to *The Birth of Tragedy* (BT, *An Attempt at Self-Criticism*, 5) is, without doubt, the acknowledgment of the perspectival character of existence. This acknowledgment poses a classical problem to Nietzsche, namely that of the conflict between different moralities or different tables of values. Since there is no one absolute morality, but rather a plurality of (often conflicting) moral perspectives, how can one perspective claim to be better than another? Here, again, relativism seems to cast its shadow and one may be led to believe that there is no plausible alternative to the position according to which every moral interpretation seems to be as true, valid or justified (i.e. *permitted*) as the others. However, this would be wrong. Indeed, Nietzsche defends the idea that it is possible – in fact, according to him, necessary – to establish a rank order among values, valuations, men, individuals, types, affects, drives, forces, goods, types of life, societies and cultures. The *Nachlass* bears abundant testimony that this is one of the most pressing tasks of Nietzsche's late philosophy. In

particular, the *problem of values* and the establishment of the *rank order of values* are considered by Nietzsche as *the future task of the philosopher*, as the following passage from the *Genealogy of Morality* clearly shows:

> The question: what is this or that table of values and 'morals' *worth*? needs to be asked from different angles; in particular, the question 'value for *what*?' cannot be examined too finely. … The good of the majority and the good of the minority are conflicting moral standpoints: we leave it to the naïvety of English biologists to view the first as higher in value as *such* … *All* sciences must, from now on, prepare the way for the future work of the philosopher: this work being understood to mean that the philosopher has to solve the *problem of values* and that he has to decide on the *rank order of values*. –

There is little doubt that Nietzsche's attempt to establish a rank order of perspectival values is problematic. Brian Leiter (2000: 277), for example, in his paper on Nietzsche's metaethics, poses the following question: 'is there any sense in which Nietzsche's evaluative perspective can claim some epistemic *privilege* – being veridical, being better justified – over its target?'[25] In other words, as John Richardson (2004: 68) points out, Nietzsche's attempt to establish a rank order of values generates an *interpretive puzzle*: how can Nietzsche reconcile his 'emphatic 'perspectivizing' of all values, including his own, with his equally vehement 'ranking' of values – a ranking that so clearly purports to some privileged status?'[26] To provide an answer to these questions goes

---

25   Leiter seems not to take into consideration the possibility that the privilege claimed by Nietzsche's evaluative perspective is not epistemic, but rather practical. See, for instance, Gerhardt 1989. On the primary function of every perspective as *sinnorientierend*, that is, as providing a meaning though which the human being can be *practically* orientated in the world, see Kaulbach (1980) and Gerhardt (1989).

26   Another way to put the problem is the following: how do we reconcile the metaethics of the values Nietzsche criticizes and the metaethics of the values he defends? As Robertson (2009: 67) puts it, 'If Nietzsche denies the objectivity of value upon which morality's claim to authority rests, he thereby deprives his own positive values of a legitimate claim to objectivity and authority; in that case, the values constitutive of his own positive evaluative outlook are no more objectively justified than or superior to those he rejects; there may then be no objective justification for the claim that we should alter our evaluative commitments or pursue the revaluation through to completion.' On this, see Stellino (2015b).

beyond the scope of this paper. However, within the present context, we may observe that Nietzsche's insistence on the need of a rank order precisely constitutes the chief objection against those readings that equate Nietzsche's moral perspectivism with a moral relativism according to which all evaluative perspectives would have the same status or the same validity. For if Nietzsche would consider all evaluative perspectives to have the same status or the same validity, why would he feel the urgent need to establish a rank order of values?

Werner Stegmaier (1994: 202) stresses quite clearly Nietzsche's original attitude towards relativism:

> According to Nietzsche, to think in a relativistic way means to search for a hold no longer on any highest point – with which, if proved to be untenable, everything would break down – but rather on a network of relations which maintain their hold on one another. For Nietzsche, nihilism was the groundless-becoming of every higher philosophy of absolute, while the relativism of his perspectivism was the disillusion that had to follow and a relief. Philosophy could now give up the search for ultimate criteria for the foundation of truth and good and, instead, explore the changing plausibilities according to which we generally validate truth as truth, good as good and grounded [*Begründen*] as grounded.

In this passage, Stegmaier particularly focuses on the connection between relativism and what we have defined as 'relationalism,' and stresses the importance of considering values and truths as generated by 'a network of relations which maintain their hold on one another' instead of with reference to a single, absolute principle. If we take this viewpoint, then it is easy to understand how a *relativization* or a *perspectivizing* of morality – which is Nietzsche's case – does not imply *ipso facto* that anything goes or that everything is permitted. As we have seen above, Nietzsche is highly aware that the risk of defending a perspectival view in the moral domain is that all evaluative perspectives can be considered to have the same status or the same validity, but he also defends a relationalistic view according to which each truth, each value can be judged only with reference to the network of which they are part. In short, Nietzsche thinks that there should be (or there *has to* be) a rank order of values and, furthermore, that the *criterion* or *standard*, which has to be defined in order to establish this rank order, must take into account the relationalism of values. Thus, the rank order of values cannot be grounded on some kind of individualistic principle or

normative ethical egoism.[27] As Nietzsche suggests in the passage from the *Genealogy* mentioned above, to define this criterion constitutes the future work of the philosopher.

References

Abel, Günter (1995). *Interpretationswelten: Gegenwartsphilosophie jenseits von Essentialismus und Relativismus.* Frankfurt: Suhrkamp.

——— (1998). *Nietzsche. Die Dynamik der Willen zur Macht und die ewige Wiederkehr*. Berlin/New York: de Gruyter.

Baghramian, Maria (2004). *Relativism*. London/New York: Routledge.

Clark, Maudemarie (1990). *Nietzsche on Truth and Philosophy*. Cambridge: Cambridge University Press

Constâncio, João (2012). 'A Sort of Schema of Ourselves': On Nietzsche's "Ideal" and "Concept" of Freedom'. *Nietzsche-Studien* 41: 127–162.

Cox, Cristoph (1997). 'The "Subject" of Nietzsche's Perspectivism'. *Journal of the History of Philosophy* 35/2: 269–291.

Dostoevsky, Fyodor (1992). *The Brothers Karamazov*. London: Vintage.

Figl, Johann (1982). *Interpretation als philosophisches Prinzip. Friedrich Nietzsches universale Theorie der Auslegung im späten Nachlaß*. Berlin-New York: de Gruyter.

Gerhardt, Volker (1989). 'Die Perspektive des Perspektivismus', *Nietzsche-Studien* 18: 260–281.

Gori, Pietro (2007) *La visione dinamica del mondo. Nietzsche e la filosofia naturale di Boscovich.* Napoli: La Città del sole.

——— (2015). 'Psychology without a Soul, Philosophy without an I. Nietzsche and 19th century Psychophysics (Fechner, Lange, Mach)', in: J. Constâncio et al. (eds.), *Nietzsche and the Problem of Subjectivity*, Berlin/Boston: de Gruyter, pp. 166–195.

---

27   Scholars have usually identified this criterion in the will to power. See, among others, Kaufmann 1950, Schacht 1983 and Richardson 2004.

—— (2016). *Il pragmatismo di Nietzsche. Saggi sul pensiero prospettivistico*. Milano: Mimesis.

Grimm, Rüdiger (1977). *Nietzsche's Theory of Knowledge*. Berlin-New York: de Gruyter.

Hatab, Lawrence (1995). *A Nietzschean Defense of Democracy*. Chicago: Open Court.

Heidegger, Martin (2002 [1943]). 'Nietzsche's Word: "God Is Dead"', in *Off the Beaten Track*. Cambridge: Cambridge University Press.

Hussain, Nadeem (2007). 'Honest Illusion: Valuing for Nietzsche's Free Spirits', in: B. Leiter and N. Sinhababu (eds.), *Nietzsche and Morality*. Oxford: Oxford University Press: 157–191.

Ibbeken, Claudia (2008). *Konkurrenzkampf der Perspektiven*, Würzburg: Königshausen & Neumann.

Irlenborn, Bernd (2016). *Relativismus*. Berlin/Boston: de Gruyter.

Jaspers, Karl (1997 [1936]). *Nietzsche: An Introduction to the Understanding of His Philosophical Activity*. Baltimore: Johns Hopkins University Press.

Kant, Immanuel (1998 [1786]). 'What Does it Mean to Orient Oneself in Thinking?', in *Religion within the Boundaries of Mere Reason*. Cambridge: Cambridge University Press.

Kaufmann, Walter (1950). *Nietzsche: Philosopher, Psychologist, Antichrist*. Princeton: University Press.

Kaulbach, Friedrich (1980). *Nietzsches Idee einer Experimentalphilosophie*. Cologne/Vienna: Bohlau.

Lee, Mi-Kyoung (2005). *Epistemology after Protagoras: Responses to Relativism in Plato, Aristotle, and Democritus*. Oxford: Oxford University Press.

Leiter, Brian (1994). 'Perspectivism in Nietzsche's *Genealogy of Morals*', in R. Schacht (ed.), *Nietzsche, Genealogy, Morality. Essays on Nietzsche's* Genealogy of Morals. Berkeley: University of California Press: 334–357.

Leiter, Brian (2000). 'Nietzsche's Metaethics: Against the Privilege Readings'. *European Journal of Philosophy* 8/3: 277–297.

Levy, Neil (2002). *Moral Relativism: A Short Introduction*. Oxford: Oneworld.

Nietzsche, Friedrich (1986). *Human, All too Human.* Trans. J. Hollingdale, Cambridge: Cambridge University Press.

—— (2001). *The Gay Science*. Trans. J. Nauckhoff. Cambridge: Cambridge University Press.

—— (2002). *Beyond Good and Evil*. Trans. J. Norman. Cambridge: Cambridge University Press.

—— (2003). *Writings from the Late Notebooks*. Ed. R. Bittner, Cambridge: Cambridge University Press.

—— (2005). *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings*. Trans. J. Norman. Cambridge: Cambridge University Press.

—— (2006). *Thus Spoke Zarathustra*. Trans. A. Del Caro. Cambridge: Cambridge University Press

—— (2006). *On the Genealogy of Morality*. Trans. C. Diete. Cambridge: Cambridge University Press.

Pfeuffer, Silvio (2008). *Die Entgrenzung der Verantwortung. Nietzsche–Dostojewskij – Levinas*. Berlin/New York: de Gruyter.

Plato (1997). *Complete Works*. Indianapolis/Cambirdge: Hackett Publishing Co.

Richardson, John (2004). *Nietzsche's New Darwinism*. Oxford: Oxford University Press.

Schacht, Richard (1983). *Nietzsche*. New York: Routledge.

Solomon, Robert (2003). *Living with Nietzsche: What the Great 'Immoralist' Has to Teach Us*. Oxford: Oxford University Press.

Stack, George (1991). 'Kant, Lange and Nietzsche: Critique of Knowledge', in K. A. Pearson (ed.), *Nietzsche and Modern German Thought*. London: Routledge, pp. 30–58.

Stegmaier, Werner (1994). *Nietzsches ‚Genealogie der Moral'*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Stellino, Paolo (2015a). *Nietzsche and Dostoevsky: On the Verge of Nihilism*. Bern: Peter Lang.

—— (2015b). 'Problemi di metaetica nietzscheana'. *Rivista di estetica* 58/1: pp. 175–190.

Political ethics as a functional requirement of democracy: sketching a theory of political values in democratic systems

Gabriele De Angelis

Social theory abandoned the topic of political legitimacy long ago. Max Weber set forth the last truly influential theory that a few scholars still occasionally attempt to follow up on (Beetham 1991). Dominated by political moralism, political philosophy disregards the (thus far admittedly insufficient) contribution of the social sciences, and frames legitimacy in terms of adequacy to pre-political moral principles – an approach to which minority opinions only pose a minor challenge despite their at times noble source (Rorty 1991, Williams, 2005). And yet, a theory of legitimacy with both normative relevance and descriptive capacity is possible through an analysis of political ethics as a "functional requirement" of a political system. What follows is a sketch of such a theory that builds on Niklas Luhmann's intuitions (2002, 2008).

1. Political theory between moralism and realism

In contemporary political theory, democratic legitimacy is most commonly conceptualised according to either one of two models: the "moralism" that characterizes the bulk of contemporary political liberalism, or realism. Moralism sees legitimacy as a fulfilment of pre-political ethical principles to which political institutions are supposed to conform. Such principles tell us what legitimate political institutions look like

and, vice versa, real existent political institutions can be said to be legitimate if and only if they conform to such principles.[1]

Realism is more multifarious than moralism. Williams' realism focuses on the Hobbesian task of securing "order, protection, safety, trust, and the conditions of cooperation" (2005: 3). The accomplishment of these tasks requires taking into account the historical circumstances in which order has to be established and qualified, and includes a number of variables of which moral principles are a component among others (2005: 77). Realism is, however, a less cogent theoretical attitude than moralism. A frequent line of thought reproduces the early modern distinction between ethics and politics as a distinction between ends and means.[2] Another, also frequent, version of realism consists in a prudential assessment of the scope of ethical consensus in contemporary societies, and concludes that democratic politics can be justified as avoidance of uncontrolled political conflict, and as a space for peaceful controversy and the exercise of public reason, mediation and negotiation (Bellamy 2010; see also Galston 2010 and Philp 2010).

Both moralism and realism accept the distinction between descriptive and normative approaches to politics, and take up the classic, already Machiavellian, distinction between morals and politics as two different and autonomous systems of norms, the first belonging to the domain of philosophy as a "normative" discipline, the latter belonging to political science as a "descriptive" discipline of political behaviour and institutional systems. Similarly, Dahl distinguishes between theories that are "essentially ethical" in character and theories that attempt to give us a picture of the world "as it really is" (Dahl 1956: 1).

Theories that attempt to do the latter often construe legitimacy in terms of expressed consent: a power relationship is legitimate as long as "it can be justified in terms of [people's] beliefs" (Beetham 1991: 11) independently of whether those beliefs are worth sharing from our point of view. The social scientist suspends his/her own judgement and attempts to look at power and legitimation from the point of view

---

1   On moralism, see Williams (2005).
2   A contemporary version of such a theory is presented by Bobbio (1984) and Coady (1991).

of the actors involved: legitimacy – as long as it is considered as a social fact – is considered to be relative to the beliefs and expressions of consent of the actors whose political behaviour the social scientist observes and interprets.

As methodologically correct as this is, such an objectifying attitude is somehow unsatisfactory when it is applied to our own political world, in which the reasons why we should or should not agree with political acts is often hotly debated, and factual but dubious consent is often opposed to deserved and qualified consent. Such a normative evaluation is usually counted to the domain of political philosophy, while social scientists content themselves with detecting the presence and the sources of factual consent independently of its moral quality. Such a sharp distinction is, however, not necessarily reasonable as it can indeed be shown that our political world is structured in such as way as to make certain moral assumptions both more frequent and more justified than others. While we participate in our political world, moral assumptions are at work that cast mora than a doubt on whether too strict a distinction between social facts and moral justifiability is actually viable.

2. Beyond the opposition between descriptive and normative theories

In the division of labour between descriptive and normative theories there seems to be no room left for the idea that ethical principles might be present in a different form than as simple "beliefs": ideas rooted in the heads of citizens and political actors with no real relevance for what happens in the political domain. In particular, in current political science or political theory, the idea is no longer pursued that a conception of legitimacy is paramount to the comprehension of political systems, so that too strict a distinction between a normative and a descriptive approach is untenable from the point of view of an effective analysis of political institutions as they are in political reality.

Political institutions rest in fact on normative, ethical assumptions without which political processes would cease to make sense both to actors and observers. In other words, ethical principles can be shown as being the "salt" of political institutions and practices. This is an intuition that was clearly present at the beginning of political sociology and political science (for instance in the works of Hans Kelsen and Carl Schmitt) (De Angelis 2009).

A political system rests on a language of legitimacy as a key component of public communication so that a semantics of political ethics is a functional requirement of any political system. In particular, a democratic political system is characterised by a definite set of identifiable semantical layers by means of which actors and observers distinguish and communicate about the legitimacy of political acts and decisions, with such a communication being part of the political process itself.

The analysis of legitimacy as a component of public political communication is descriptive in as much as it identifies a set of semantical references as a presupposition for the functioning and understanding of a democratic political system. As such, it resists the temptation to affirm what ought to be from the point of view of Reason or truth. However, the distinction between a descriptive and a normative approach becomes less and less useful once we acknowledge that the normative assumptions that are inherent to democratic political institutions and communication are what we as citizens of democratic countries act upon in our political behaviour. Such a set of assumptions lies at the core of our political judgement. Political institutions rest on an ethical basis that invites us to question political behaviour against the background of normative assumptions that are valid inasmuch as they are part of the institutions that frame our political behaviour. These kinds of normative assumptions represent therefore a "tacit" knowledge (Polanyi 1966) that makes our political behaviour meaningful and intelligible. In particular, the here suggested approach analyses political institutions as the embodiment of ethical assumptions in a dialectical relationship between *procedures* and *semantics*, the latter being "functional requirements" of the former. Specific institutional mechanisms become intelligible only in the light of specific ethical assumptions. Without the latter, the former cease to make sense to their actors. Moreover, formal procedures are often

structured in such a way as to "make true" certain ethical assumptions, as will become clear in the case of political representation.

3. The "circle" of democratic power

Any political system is characterised by a set of rules that determine who is authorised to make collectively binding decisions and how, i.e. by means of which procedures (Bobbio 1995: 4). In a representative democracy, such determination takes the form of a chain of authorisations that starts with a more or less complex and mediated selection of decision-makers by means of which citizens choose "representatives" in a regularly repeated electoral process. The rules of the game are such that decision-makers are chosen between alternative proposals of competing political personnel and platforms. They aim to ensure "responsiveness" and "accountability", and therefore result in the indirect control of citizens over decision-makers. This is the kernel of democratic "freedom": citizens are not free because they make the rules themselves, they are free inasmuch as decision-makers are not free to make the rules arbitrarily (Sartori 1994: 172). Political freedoms and rights are such as to ensure the smooth reproduction of such a "circle of power".

A political system thus results in the alternation of inputs and outputs. In a democratic political system, the *input* consists principally in an electoral result that works as an authorisation for a definite set of people to make collectively binding decisions. These decisions (and their impact on social life) are the *output* of the political system.

Inasmuch as it is bound up with a mechanism of authorisation – and therefore with a set of expectations and decisions that are supposed to satisfy these expectations ("responsiveness") – the relation of inputs and outputs already is an ethical relation in which a legitimising input represents a "request" that is to be interpreted and responded to through the output. The correctness of such an "interpretation" is tested by the periodical reiteration of the process of authorisation.

4. Input and output legitimacy, and sovereignty

In modern times, the "circle of power" depicted above is explained with the people's "sovereignty". The specific procedures that give to the political system its input-output structure tell us how such sovereignty is exercised and exactly what it consists of: it generally consists of the capacity to periodically (and more or less directly) determine who is authorized to make *collectively* binding decisions, while the legally valid procedures determine the specific articulation of such a general ethical relationship. Procedures are thus intelligible as embodiments of more specific ethical principles (such as the exact entitlements to the exercise of sovereignty, the interpretation that is given to political equality, the constitution of social, geographical and cultural subgroups as political constituencies, etc.). On the other hand, ethical principles only become effective by means of procedures. Ethical principles make procedures "meaningful" inasmuch as they are the reason why certain procedures are in place: for instance, for the sake of popular sovereignty, collective self-determination, freedom, etc. On the other hand, procedures tell us what kind of a difference ethical principles make in the practice. Generally, "popular sovereignty" makes the world different inasmuch as voters contribute to the periodical selection of decision-makers. More specifically, the exact entitlements will determine who is exactly part of "the people", what "popular" therefore means, which specific cultural or geographic interpretation is given of the set of actors who share this entitlement.

However, the liaison between procedures and principles will also show the limits of the latter: in contemporary democracies, sovereignty can only be exercised indirectly by means of a relation of political representation, and is filtered by a complex set of institutional, communicative and organisational infrastructures. Moreover, procedures are independent of the ethical principles that they are supposed to embody inasmuch as time can change some of the sociological assumptions that are part of those principles, while the procedures continue to take place unperturbed. For instance, the feelings of the actors that compose a constituency may change throughout time as regards the reasons for

their coexistence: separatisms may arise just like separations may lose their reason to be. Conversely, procedures can also become obsolete as they may be perceived as more or less satisfactory embodiments of a changing concept of sovereignty, representation, etc. Although institutional procedures are intelligible as embodiments of ethical principles, this does not mean that they are also their truthful implementation, nor does it mean that such an implementation would not be possible otherwise.

Far from being an endpoint, the double relationship between procedures and ethical principles is the starting point of political communication, for both the exact "embodiment" or realisation of ethical principles by means of institutional procedures and the reasonableness of ethical principles in the light of what we can actually achieve procedurally are open for discussion. Thus, the reform of an electoral system will likely be motivated through reference to a "better" representation of "the people", the reform of an institutional system might prompt discussions about the exercise of sovereignty in a complex and multifarious constituency (as has been the case in Italy when elements of a federal system were introduced in 2001).

Despite historical mutations of procedures and semantical nuances, the interpretation of the electoral mechanism as an authorisation and of its reiteration as a form of accountability is a common feature of contemporary democracies. They make sure that decision-makers are not self-referential, but – on the contrary – interrogate themselves as to the "requests" that come to them from "outside" the political system.

What ought to be, however, the result of such an interrogation? Again, the history of political semantics provides us with "representation" (of the people's interests, needs, etc.) as an apt normative reference.

5. "Representation" as an ethical relationship

What ought to be represented, and how, is subject to interpretation, but the semantical reference to representation is a starting point for political communication: something is expected to be represented and somebody

is expected to represent something and somebody else. What, how and by which means will be a further object of political communication and, ultimately, a matter of choice, preference and evaluation by means of the electoral process. Notwithstanding its openness, a relation of representation fulfils the goal of entitling the represented people with the legitimate expectation to be the addressees and the arbiters of decisions and decision-makers. Thus, a formal procedure makes an ethical assumption true inasmuch as it prompts political actors to enact corresponding patterns of behaviour: as a political representative I will have an interest in being accountable to my constituency and will be motivated to interpret my role as a role of representation with a strong ethical character. And, vice versa, as a citizen and a voter I will also be induced to look at political representatives as more or less satisfactory interpreters of my political preferences, for it is the very structure of the political process that motivates me to do so.

Thus, representation calls for responsiveness and accountability. It requires that decision-makers address citizens as the ultimate sovereigns. "Representation" is one of the core assumptions of democratic politics in that it structures political communication around the need to interpret the citizens' "will", "needs", "demands" and so on, and address them adequately. "Representation" shows how the relationship between formal procedures and ethical principles is structured in such a way as to make sure that not anything goes, but is also "loose" enough to open up a space for controversies, alternatives and civilized struggles.

The space for controversies that opens up in the interplay of ethical principles and institutional procedures allows for – roughly – two levels of communication: the first concerns the "content" of inputs and outputs, the second concerns the adequacy of the principles and procedures themselves. This is the point at which the "dialectical" relationship between procedures and semantics sets in: ethical principles become real only through their embodiment in procedures, while procedures are meaningful only against the background of some ethical principle. Political equality is a case in point.

6. Systemic approach vs. moralism: the case of political equality

In a representative democracy, political equality is realised by means of an institutional procedure that endows citizens with an equal voting right. Again, there might be exceptions, but they have to be justified against the background of such a general assumption of equality.

Indeed, as long as political equality is upheld as a principle and "equal voting right" is its key embodiment, exemptions will have to be publicly justified. This is the main consequence of having a principle and a corresponding procedure as normative paradigms: both work as a criterion in political communication in relation to which discussions are conducted, normative expectations structured, and justifications asked.

The ethical relationship between principle and procedure outlined here is different from what political moralism usually understands as a moral foundation of politics. Thus, political equality can surely be understood as the embodiment of a moral principle that *a priori* (i.e. from a non- or pre-political standpoint) determines what is politically legitimate. This is the case whenever we understand political equality as the embodiment of the equal moral standing or dignity of human beings, their mutual recognition, respect, etc. However, the relationship between ethical principles and formal procedures that is being pointed out here is different from the assumptions of political moralism inasmuch as the selfsame relationship is open to a plurality of moral interpretations, theories and scholarly attitudes. Enlightened self-interest or considerations of opportunity may also be sufficient to accept democratic rules as expressions of a necessary *modus vivendi* (Horton 2010: 440).

Thus, political equality and its embodiment through equal voting rights and weight can as well be justified prudentially inasmuch as I accept that the equal representation of all (alongside the constitutionally entrenched fundamental rights) is a guarantee against oppression that is simply necessary in a complex society in which several conceptions of the good and innumerable individual life plans coexist and compete with each other. In other words, I may accept political equality as a guarantee that the law will be obeyed although anyone would be inclined to exempt him/herself from obedience, as is the case in the Kantian "people of devils":

> The problem of the formation of the state, hard as it may sound, is not insoluble, even for a race of devils, granted that they have intelligence. It may be put thus: Given a multitude of rational beings who, in a body, require general laws for their own preservation, but each of whom, as an individual, is secretly inclined to exempt himself from this restraint: how are we to order their affairs and how establish for them a constitution such that, although their private dispositions may be really antagonistic, they may yet so act as a check upon one another, that, in their public relations, the effect is the same as if they had no such evil sentiments (Kant 1795/1903: 153–4).

Goodin also construes the principle "one person, one vote" not only as the enactment of a moral principle of fairness and reciprocity, but also as a safeguard against extreme decisions: since majorities are likely to shift throughout time, I may consider taking on a prudential attitude for fear that those that I put to a burdensome disadvantage when finding myself in the majority will pay me back in the same coin once I find myself in the minority (Goodin 1992: 85). Publicity, accountability and discursive defensibility are further examples of rules whose acceptability may rest on both prudence and morality.

While moralism takes steps from pre-political moral principles, the systemic approach shows that political procedures and principles of public political ethics are compatible with different moral beliefs. Such a pluralist understanding of morals in politics acknowledges that moral principles rarely are the *ratio essendi* or the historical origin of the political procedures that shape our political systems. A closer look at the moral debates that shaped the historical origin of our political institutions would show that moral reasons surely played a role, but it would also show that political institutions are not in place *because* a given moral principle has been universally acknowledged and applied at a certain given time. The historical origin of political institutions is far more complex than this, and results from a complex interplay of moral attitudes and circumstances as well as material, class and sectoral interests.

Nevertheless, the specific result of such an interplay of social forces is that political equality is a normative and institutional assumption lying at the basis of our political systems. It belongs to its semantics, i.e. to the fundamental assumptions of the political communication that takes place in it, and is solidly anchored in its political practice. "Equality" consists of an equal chance to determine who will be a decision-maker and (indirectly) what his/her agenda will be. For such a

chance to be effectively used, further "functional assumptions" must be in place: equal access to information, and therefore freedom to produce information, freedom of speech, freedom to form political organisations, etc. (Dahl 1989: Ch. 8; 2006).

7. General vs. self-serving interest

Representation as an ethical relationship is undetermined enough to make room for different interpretations of what is to be represented and how. It is a choice of the representative to focus on concepts such as "interests", "needs", "demands", or on actors such as "citizens", "groups", "classes", etc. However, the history of modern political semantics provides us with a dominant dichotomy: the one between "self-serving" and "general" interest. Although it is difficult enough to define the concept of general interest (the indeterminacy of Rousseau's concept of a "general will", as opposed to a "will of all", is indicative enough of such a difficulty), it is far less problematic to detect a self-serving interest. A particular interest as the interest of a few is usually what is not to be represented, unless it also serves a general interest or the interest of all (Shapiro 2003: 200) (as for instance in the justification of "trickle-down" economics). Exceptions are indeed possible, but must be motivated (as is for instance the case of "positive discrimination") against a normative expectation that takes for granted that a general, not a self-serving, interest is what ought to be promoted.

What exactly is a correct interpretation of a general interest, or the interest of all, and what is just a self-serving interest is again open for discussion. (for instance, is trickle-down economics really in the interest of all, or is it just neoliberal hocus-pocus?). The conceptual opposition between the two kinds of interest does nevertheless shape citizens' expectations as to the legitimacy of collectively binding decisions – a legitimacy that will be verified again in the course of the electoral process once decision-makers put their behaviour to the test of the ballots. Between one electoral process and the next, the concepts provide participants in political communication – be these opposition parties, protest-

ers, civil society organisations, etc. – with a normative reference that they can use to shift public opinion in their own favour. In democratic politics – and as far as the citizenry is concerned (constitutional courts' rulings are a somehow different matter) – it is ultimately the electoral process, i.e. a numeric result, that decides upon the quality (including the moral quality) of decisions and decision-makers. The interplay between normative criteria of political judgement (here: "semantics") and formal procedure is such that no *telos* leads necessarily towards the discovery of the "real" general interest or political "truth". The political process is not necessarily a cognitive process. The criteria in question are "empty" inasmuch as their content is necessarily controversial: citizens do not need to agree on what a general will is in order to successfully interact in political communication. It is enough that they share sufficient semantical reference points in order to be able to communicate and understand each other despite their disagreement. The simple fact that a general interest (or the interest of all) is what is, in principle, to be achieved in politics, while a particular will is usually what is to be avoided or subjected to specific justification, is enough to make sure that decision-makers will aim to depict their political choices as legitimate representations of the former, while the opposition will aim to make them out as expressions of a self-serving (or a failed expression of a general) interest.

Both the formal procedures of democratic regimes and the semantical opposition between general and self-serving interest serve the goal of stimulating the generalisation of particular, individual or sectoral interests, as political actors strive for the formation of an electoral majority. Whatever kind of interest an actor represents, he/she will hardly be able to reach a political majority if what he/she represents cannot be shown to match some kind of "general" or collective interest. In order to achieve such a majority, political actors must rely on public communication and its categories of political judgement. The restrictions to which particular or self-serving interests are subject in public communication favour the framing of interests in terms of more general, encompassing, overarching or collective interests. Insofar as democratic procedures include public discussion and justification, the generalisation of interests is part of results in a sort of Kantian "public use of Reason".

8. "Individualism" as a functional requirement of democracy

Besides lying at the core of an ethical relationship between represent-atives and constituency, political equality and its institutional embodi-ments produce a further ethical consequence: the imputation of political choices to individuals. Citizens bear responsibility for their electoral choice inasmuch as they are the starting point of the electoral proce-dure. Therefore, a political system based on citizens' choice – exercised through the casting of a vote – institutes individual responsibility for the consequences of the votes cast – and therefore individual autonomy of judgement – as two fundamental moral assumptions. Participation, abstention, knowledge, ignorance, degree of information, etc. will all be ascribed to the individual as a citizen of a democratic society. I will be the author of my behaviour and the sole person responsible even if I choose not to know and not to participate. The individualisation of the procedure paves the way for a series of imputations and claims: the claim to receive sufficient access to relevant information, to be put in the condition to best exercise my political autonomy, etc. Thus, a demo-cratic political system also institutes an *ideal* of the democratic citizen, who is supposed to get access to information, to debate, to develop cognitively consistent attitudes and make his/her choice accordingly. It is well known how much political reality differs from such an ideal of the democratic citizen (Zolo 1992: 111–20). Nevertheless, the prima-ry effect of such a mechanism of imputation of rational or reasonable choice is to open up a space for controversies about access to and qual-ity of information, freedom of the press, manipulation, rationality, etc. It is the normative background against which citizens of a democratic society conduct their debates about the quality and openness of public communication, and the degree to which it is conducive to the kind of responsible and autonomous choice that the institutional context pre-supposes. Again, such a normative background is purely political inas-much as it is independent of whatever "natural" or pre-political moral assumptions I may want to make with regard to human beings. It is a consequence of the political institutions we live in. Nor are these nor-mative assumptions undermined by scepticism towards political elites:

the disenchanted citizen that blames the untrustworthiness of repre-
sentatives actually acts upon the same assumption, and is just shifting
responsibility to the person or organised group that obtained his/her
consent under false pretences.

9. Political ethics in the social sciences

Although "counterfactual", the link between ethical principles and for-
mal procedures bears important consequences for political life in that
it determines the grid of cognitive criteria through which we look at
and communicate about democratic politics. Such a link binds political
actors to give allegiance (or at least pay lip service) to given ethical
standards. On similar grounds Skinner argues against the idea that po-
litical actors' profession of belief in political ethics is at best an ex-post
rationalisation: although the acknowledgement of such principles can
be merely ideological or straightforwardly manipulatory, it does never-
theless make a difference inasmuch as it subjects political behaviour to
collectively valid standards of ethical judgement (Skinner 2002: 145).

The ethical meaning of legal procedures belongs, in sum, to the
"functional requirements" of politics. It structures public communica-
tion around a definite set of normative expectations, offers criteria for
assessing individual behaviours on the basis of ascribed responsibil-
ities, determines political roles and traces the boundaries of political
entities (such as "the people"). In complex societies, such a liaison be-
tween ethical principles and formal procedures includes a "reduction of
complexity" inasmuch as innumerable individual wills, preferences and
interests are summed up in the (ideally) dual relationship of political
minorities and majorities. Citizens are as well encouraged to link up
to other people's different wills, interests and preferences, as their own
claims have having to pass the test of generalisability while striving
for political majorities. This results in a further ethical relationship that
involves the members of a constituency: political behaviour and deci-
sion-making will tend to be structured in such a way as to satisfy the
need to respond to a collective "will" or interest.

The intimate connection between ethical principles and formal procedures provides citizens with a set of cognitive criteria for political judgement by means of which they can communicate and understand each other in public communication. Nevertheless neither the ethical principles nor the formal procedures are fully determined by each other: there are several interpretations of popular sovereignty as there is more than one procedure to enact it. Democracies are also dynamic political systems inasmuch as the relationship between ethics and procedures shifts throughout time, propelled by the struggle for new political majorities, political opportunities and semantical innovations. The "ambiguities" of the ethical principles, their openness for ever further interpretation and reconsideration as well as the questionability of formal procedures (how apt are they to embody the ethical principles on which they rest? what are their unintended consequences?) are all dynamic elements in the evolution of political systems.

The study of "embedded" political ethics opens a new perspective on legitimacy. It shows that the interplay of ethical semantics and formal procedures represents a key element of everyday normative assumptions in political communication and behaviour. Political ethics emerges as an ingredient of formal procedures beyond both the straightforward distinction between normative and descriptive approaches and the opposition between moralism and realism in political theory. It may represent a new paradigm for the study of political ethics from the point of view of the social sciences, while having a positive impact on political theory inasmuch as it casts light on the normative assumptions that are a functional part of everyday political life.

References

Beetham, David (1991). *The Legitimation of Power.* Basingstoke: Macmillan.

Bellamy, Richard (2010). 'Dirty hands and clean gloves: liberal ideals and real politics'. *European Journal of Political Theory* 9 (4): 412–30.

Bobbio, Norberto (1984). 'Etica e politica', in W. Terga (ed.), *Etica e politica*. Roma: Pratiche Editrice: 7–17.

Bobbio, Norberto (1995). *Il futuro della democrazia*. Torino: Einaudi.

Coady, C.A. (1991). 'Politics and the problem of dirty hands', in P. Singer (ed.), *A Companion to Ethics*. Oxford: Blackwell: 373–83.

Dahl, Robert (1956). *A Preface to Democratic Theory*. Chicago: University of Chicago Press.

_____ (1989). *Democracy and Its Critics*. New Haven: Yale University Press.

_____ (2006). *On Political Equality*. New Haven: Yale University Press.

De Angelis, Gabriele (2009). 'Ideals and institutions: Hans Kelsen's political theory'. *History of Political Thought* 30(3): 524–546.

Galston, W. A. (2010). 'Realism in political theory'. *European Journal of Political Theory* 9(4): 385–411.

Horton, John (2010). 'Realism, liberal moralism, and a political theory of modus vivendi'. *European Journal of Political Theory* 9(4): 431–48.

Kant, Immanuel (1795/1903). *For Perpetual Peace*. Transl. M. Campbell Smith. London: Allen & Unwin.

Luhmann, Niklas (2002). *Die Politik der Gesellschaft*. Frankfurt am Main: Suhrkamp.

_____ (2008). *Die Moral der Gesellschaft*. Ed. by D. Horster. Frankfurt am Main: Suhrkamp.

Philp, M. (2010). 'What is to be done? Political theory and political realism'. *European Journal of Political Theory* 9(4): 466–84.

Polanyi, Michael (1966). *The Tacit Dimension*. New York: Anchor Books.

Rorty, Richard (1991). 'The priority of democracy to philosophy' in *Philosophical Papers, Vol. 1*. Cambridge: Cambridge University Press: 175–196.

Sartori, Giovanni (1994). *Democrazia: cos'è*. Milano. Rizzoli.

Shapiro, I. (2003). *The Moral Foundations of Politics*. New Haven: Yale University Press.

Skinner, Quentin (2002). *Visions of Politics, Vol. 1*. Cambridge: Cambridge University Press.

Williams, Bernard (2005). 'Realism and moralism in political theory' in G. Hawthorn (ed.), *In the Beginning was the Deed*. Princeton and Oxford: Princeton University Press: pp. 1–17.

Zolo, Danilo (1992). *Il principato democratico*. Milano: Feltrinelli.

II – Arguing about values

Practical Rationality at Work – A New Argumentation Model

João Sàágua and Michael D. Baumtrog

Introduction

Conceptual ideals typically provide unobtainable standards from which to measure and evaluate the real. In philosophy, ideals are often conceptualized and articulated not with the intent of literally dictating practice, but with the hope of providing norms from which reflective evaluation might lead to improved practice. Such is the case with any ideal model of reasoning or argumentation. No philosophy working with ideal reasoning or argumentation expects real reasoners or arguers to follow the ideal model perfectly or always. Rather, many of the models of reasoning and argumentation put forward thus far serve as standards against which poor or fallacious practices become apparent. Thus, the 'philosophical punch' of an ideal model is two-fold: the model itself is an articulation of the way things ought to be, and if correct, the way things ought to be informs, motivates, and justifies articulations of failures in practice.

In what follows we provide our first articulation of a new ideal model of integrated practical reasoning and argumentation. It is ideal in two senses. First, its method is prescriptive, although it may never be executed perfectly in practice. Second, as an ideal standard, applying it to an instance of real life reasoning or argumentation will provide evaluative insights (see Baumtrog 2015). However, although an ideal model, we have tried to keep its ideal and prescriptive aspects relevant, pertinent, and realistically representative of intuitions regarding what we actually do when we practically reason and argue.

As alluded to above, we are not the first to put forward an ideal model for practical reasoning or argumentation. Philosophers have long

focused on practical reasoning, with recent models/schemata coming from Broome (2002) and Audi (2006), to name but a few. Argumentation scholars such as Walton (1990, 2007, 2013a, b), van Eemeren, Grootendorst, & Snoeck Henkemans (2002 pp.101–102),[1] and Fairclough & Fairclough (2012) have provided recent models of practical argumentation. As such, if we are to be responsible academics, we must then justify the need for our current contribution; we must answer the question 'What is on offer here that cannot be found elsewhere?'

The answer will become clearer below, but here we would like to offer one brief answer explained through three supporting points. The main answer is that the new model is an *integrated* model of practical reasoning and argumentation. Philosopher's working on practical reasoning and ethics have thus far paid little attention to argumentation theory. Similarly, argumentation scholars have paid little attention to the work of those in practical reasoning and ethical decision-making. In short, there is a gap between the two areas. This could be in part due to the explanation Walton (2007: 212) gives that there is a contrast between his 'commitment' model and the 'Belief, Desire, Intention' (BDI) models. As integrative, then, we consider our model a novel contribution.

The three supporting points indicate the main ways the model is integrated. The first is that it can be used by both commitment and BDI proponents. For those viewing the model with a mind to argumentation, we invite the reader to feel welcome to think in terms of commitments interlocutors have and can be held accountable for. Viewing the model as an instance of reasoning, we invite the reader to view the model in terms of attitudes that connect to reasons in an inferential process. In both cases, however, we side with the view that practical reasoning and argumentation conclude in an intention to act rather than, as many in argumentation have it, a *belief* that one should act. For argumentation, this means maintaining a discursive commitment that one intends to perform some act.

The second way the model is integrative is that we reconceptualise and reposition what have come to be known as 'argumentation schemes' as producing *pro tanto* reasons. In this way, we use one concept from argumentation – argumentation schemes with critical questions – while

---

1   They refer to their scheme as 'pragmatic argumentation' rather than 'practical'.

using them as providing support for reasons used on a BDI conception of the role of reasons. The importance of this move will be made clearer below.

The third point of integration is that we attempt to integrate explicit moral considerations into the production and evaluation of practical reasoning and argumentation. While moral components have been a focal point for philosophers of practical reasoning, many in argumentation have stayed away from including them on the basis that procedural accounts of argumentation ought to remain silent on content. As will be shown below, our account takes moral considerations into account, without, however, dictating or advocating for an authoritative moral theory or threshold.

Without further ado then, the next section will introduce our starting points: what we take as background and assumptions from which to proceed. Section 3 provides an overview of the model, looking at its functioning on a macro scale. Section 4 zooms in on each of the model's component parts, detailing and justifying the selection of the schemes so as to provide a picture of the scaffolding of practical reasoning. In section 5 we provide a summary and some concluding remarks.

1. Background and Assumptions

1.1 Background

We take up Broome's characterization of reasoning as 'a process whereby some of your attitudes cause you to have a new attitude' (Broome 2013: 221) and agree that for practical reasoning the new attitude is an intention. Accordingly, in designing this model we have conceived of practical reasoning as an activity of the human mind aiming at forming an intention to complete the actions required for some alteration in the state of the world.

Theoretically, we separate reasoning and argumentation – though recognize the two are intertwined in practice. We consider argumentation to turn on the notion of conflict, and thus conceive of it as a dialectical

situation, which can be individual, dialogical, or polylogical. It is the practice, through which human agents support or criticise a given line of reasoning, or a step of that reasoning (Baumtrog 2017). Reasoning and argumentation are differentiated by the nature of the activity that each one carries out. Reasoning is a mental and usually individual process leading to a conclusion. It is an activity of the mind through which an individual, starting from certain mental states and following a rational process according to rules, leads his mind into a new mental state that concludes the process (Broome: 2004). Argumentation begins when one or more parts of the process of reasoning come into conflict – it is the contestation of reasoning or its conclusion.

In terms of practical argumentation, if the argumentation from an opponent is successful, the proponent can interiorize that recommendation and make it his own intention. Only when someone reasons or argues by himself, does the argumentation immediately result in an intention to perform the action (or not).[2] Practical reasoning and argumentation have the following purpose in common: to produce and serve as rational support for an intention to realize an action and/or a sequence of actions consisting of the means chosen to achieve that action.

1.2 Assumptions

In order to philosophically frame the model, it is first necessary to explain the main assumptions from which we start.

A) Practical reasoning and argumentation:

(A1) *Objectives are intentions*. Objectives are nothing other than intentions linguistically expressed and sufficiently stable to serve as the base for practical reasoning and practical argumentation. Since objectives can be more general or more specific, so too can intentions. In some cases, it is helpful to distinguish more precisely between an objective and an aim. Whereas an objective can be achieved through a traceable

---

2   Thanks to Dima Mohammed for this suggestion.

causal sequence, an aim is a broader objective such as 'being a good person', which may not be as easily traceable. Since both are intentions we consider them both objectives. However, for clarity, we try to use the word 'aim' where it applies and is helpful to clarify.[3] In what follows, and for formal simplicity, we shall use 'G' (or variants) to refer to the common content of the intention or the objective.

(A2) *Complex objectives give rise to plans*. When a certain objective assumed by x is sufficiently complex and for that reason *involves a progressive execution over time*, it gives rise to a *plan*. Plans, among other things, influence our actions beyond the present (Bratman, 1987). In what follows, we shall use 'M' (or variants) to refer to the content of any means or sequence of means, whether they belong to a plan over time or to simpler practical reasoning.

(A3) *The relation between Objective and Mean is contextual*. The first objective of a plan can be a means for another, more inclusive, plan. It should also be noted that, for example, x can have as an objective 'to be in a place of power' and use the sub-plan 'to be Prime-Minister' as a means (and that other means/sub-plans would eventually also be needed). Hence, it becomes apparent that being a means and an objective (end) often depends on context and can be conceived of differently depending on the level of zoom with which the reasoning or argumentation is viewed.

B) Human Agents

(B1) *The relation 'is a reason for' is considered primitive and pro tanto*. To justify their objectives and the means they choose for realizing them, human agents reason and argue in *terms of reasons*. At this point we will not go beyond the intuitive notion of 'a reason' that Thomas Scanlon articulates: 'a consideration that counts in favour' (Scanlon, 1998, p. 17). For example, that 'x is thirsty' is a reason (a consideration that counts in favour) for x to (intend to) drink water. Along with Dancy (2004) we recognize that reasons may count in favour of and/or against

---

3   Many thanks to David Hitchcock for bringing the difference between goals and
     aims to our attention. We acknowledge that further work is needed regarding how
     this might impact our view of practical reasoning and argumentation overall.

the assumption of an objective and we are therefore talking about *contributing* reasons or reasons *pro tanto*.[4] In what follows, if we wish to distinguish between reasons, we will number them as R1, R2, and so on. To qualify reasons, we will write R+ or R-, depending on whether these contribute in a positive or negative way, respectively, for the assumption of an objective, G (or for the adoption of a mean, M). Taking this notation a little further, we will accept that $(R\pm1, \ldots, R\pm n)G^*$ represents the set of reasons, positive or negative, associated to the assumption of an objective G^* where the asterisk identifies that the goal has yet to be assumed and that $(R\pm1, \ldots, R\pm n)M^*$ represents the set of reasons, positive or negative, associated to the adoption of a mean or set of means which have yet to be assumed. Accordingly, and in short, G^* is used to stand for 'proposed goal' and M^* is used to stand for 'proposed means.'

(B2) *Situation, Circumstance, and Context.* We can describe practical reasoning and argumentation in relation to the baseline situation, S', and to a situation of arrival, S*, also called a future state of affairs (Hitchcock, 2011; Fairclough and Fairclough, 2012): x is in situation S', S' has some aspect that leaves x unsatisfied, let us call that aspect 'the absence of G^*', and x assumes G^* as an objective, whose realization will turn S' into S*. x thinks that to realize G^*, he should mobilize the means M^*. The beliefs and evaluations the agent(s) hold about a given situation determine what is relevant for a given occurrence of practical reasoning or argumentation. Accordingly, we call the context of practical reasoning and argumentation the set of relevant circumstances.

(B3) *Plausible Justifications and Defeasible Rules*. Given that incompatibilities exist between objectives, means for objectives, and the means for one objective impacting a different objective, etc., combined with the ever evolving (perceived) knowledge of the agent, it follows that the rules the agents can use to infer a certain conclusion from certain premises are rules of plausible inference and, therefore, remain defeasible and sensitive to context.[5] Although they are never deductive or inductive (or statistical) inferences (even if these enter as components

---

4   The literature on reasons is vast and very complex. Some of our favourite texts are (in alphabetical order) Broome (2001, 2004), Dancy (2004), and Scanlon (1998).
5   As convincingly argued by Walton, Reed, & Macagno (2008). Along the same lines, though more moral, see Dancy (2004: 111–117, 184–187).

of plausible inferences), plausible justifications and defeasible rules should not be seen as a defect or limitation, but rather the condition of the exercise of practical reasoning and argumentation.

With this background and these assumptions in mind, we now move to our view of practical reasoning and argumentation illustrated through a flowchart and accompanying explanation.

Overview of the New Model

Figure 1: Integrated Model for Practical Reasoning and Argumentation.[6]\
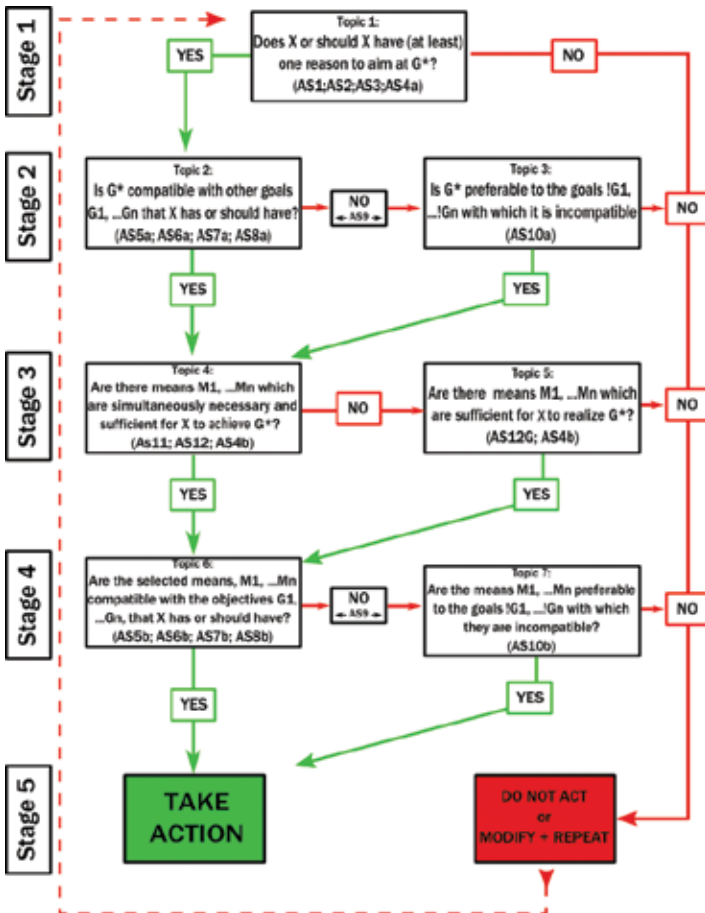
Our model is an *integrative, realistic,*[7] *and normative model*. In a single representation, our model integrates the structure of both practical reasoning and practical argumentation, including the variants usually differentiated in both – i.e., instrumental, normative, and value based. It is realistic in the sense that following the model generally corresponds to the real practice of reasoners and arguers. It is normative in the sense that it prescribes a chain of inferences (for reasoning) or a chain of primitive argumentative schemes (for argumentation) that should occur, and in a certain order, for both to provide maximally plausible formulations, conclusions, and decisions.

3.1 Stages and Topics

The model has *5 Stages*. Stage one addresses the agent's motivation for action; Stage two is concerned with the proposed goal and other goals; Stage three concerns the available means for achieving the proposed goal; Stage four deals with the relation between the means and between the means and other goals; and Stage five is the decision to act, not act, or make a modification to the reasoning or argumentation and start the process over. Given that our model is integrative and that, simultaneously, we think that the assumption of the objectives themselves should be an object of reasoning and argumentation – and not only the choice of means – our model includes two initial stages about objectives, two about means, and one for the decision.

In order to license moving from one Stage to the next, the reasoner must answer one or two 'Yes' or 'No' *Topic* questions. In any case, an affirmative answer results in a 'green light' to move to the next Stage. In some cases, a negative answer or 'red light' will lead to another Topic question and thus a second chance to move to the next Stage. In other cases, a negative answer leads straight to a conclusion not to act.

Each Topic questions an aspect of the general theme of the Stage and conditions the specific practical reasoning and argumentation associated with it. Ideally, the answer should be properly justified through

---

7   Thanks to Eugen Poppa for this term.

an instantiation of one or more primitive argumentation schemes (AS), together with responses to their respective critical questions (CQ). These together determine the basic argumentation structure of the Topic. Discussing (arguing about) those primitive argumentative schemes may require (several) other argumentation schemes.[8] It is not possible to anticipate which schemes those might be, given that they can vary from case to case. We can thus only provide a complete string of what we believe to be primitive schemes.

3.2 Tracks

Given the possibility of providing differing answers to the Topic questions, there are different paths or tracks one can take through the model. The 'fast track' (shortest path)[9] most readily resembles routine reasoning and involves only 'Yes' answers to the Topic questions. In such a case the arguer only addresses Topics 1, 2, 4, and 6. In the most involved cases, the arguer has to address all of the Topics – weighing the reasons for and against the acceptance of both the goal and the means.

3.3 Schemes and Critical Questions

Many of the schemes we include in our primitive list are based on schemes already articulated by others, especially Walton, Reed, and Macagno (2008). We have, however, made efforts to systematize the schemes by including only one term with inferential power per premise. For example, the first argumentation scheme for the argument from teleology includes as the first premise: x has G+ as its finality. In this case 'finality' is the sole term with inferential power.

---

8   The distinction between primitive and derived schemes is contextual and was a suggestion made by Fabrizio Macagno.

9   Nothing guarantees a reasoner or arguer will be able to address the 'short' track Topics quickly or quicker than perhaps all of the long track Topics. We are making a quantitative observation here only that fewer Topics and schemes need be addressed when taking this route.

We have also adjusted the schemes and critical questions to strengthen them for individual use. If you have a separate, critical Other asking the critical questions, then the wording of the questions can be less stringent in light of the opportunity for the Other to 'press harder' – so to speak – if the answerer does not provide a satisfactory answer. In the case where you are the only one responsible for asking and answering the critical questions, a more carefully worded question will make it harder to provide an unsatisfactory answer. As such, we find the wording of the questions is of great importance and have avoided using critical questions with only 'Yes' or 'No' answers.

Accordingly, we asked ourselves 'what makes a critical question, critical?' If you envisage the question being asked by a critical Other, then part of the answer would be 'The disposition with which the question is being asked'. When you are alone and conducting individual dialectical or quasi-dialogical argumentation, however, that critical attitude may be nearly absent. Thus, another part of the motivation for the way we have formulated the critical questions was in an effort to make the questions themselves as critical as possible, while relying less on the person asking them.

Taking the above two considerations into account – avoiding Yes/No questions and making the questions critical – we have designed the questions as a pair of questions, the first of which asks for an explanation of the inferential term's use and the second of which challenges it. Both parts are necessary, in our view, for the critical questioning of the inferential term to be adequate. Thus, in our model, the second (part of the) question ensures the question contains a critical component in every case and regardless of the questioner or their disposition.

Finally, it should also be noted that schemes may be used more than once to answer any given Topic question. For example, when answering the question in Topic one, the reasoner or arguer may put forward two differing reasons from positive values. In addition, they may avoid one of our identified schemes altogether. For example, they might not use any reasons from teleology. Our only contention is that at least one of the argumentation schemes should be used in answering any given Topic. As contributory, how the reasons resulting from the use of the schemes interact is a separate question from which and how many schemes are used.

3.4 Closure

Following a track and using the schemes ensures that the use of our model always ends in a traceably justified decision to act or not act. Whereas other models have left the decision to an unarticulated procedure of weighting, our model provides a way to rationally justify the selection of one of the alternatives. This is especially important for the evaluation or challenging of the process. On our account, the choice of which schemes to employ (or not), as well as the quality with which they were employed, can both be identified and evaluated as precise areas of challenge or critique against which an alternate possibility can be clearly projected.

2. The Structure of Practical Reasoning and Practical Argumentation

Let us now identify the problem each Topic addresses, how this problem can motivate practical reasoning, and the primitive argumentative schemes (and respective critical questions) that should be used to justify a response.

Stage 1. Topic 1
Stage 1, containing only Topic 1, consists in answering the problem: 'Does X or should X have (at least) one reason to aim at goal G*?' It should be noted that a reason here does not have to be a pro-attitude. I can suppose that I should assume G* for another type of reason: maybe G* involves some sacrifice that I have to make (hence, my not having a pro-attitude towards G*), but, if I assume G*, perhaps I feel that I am contributing to realize a certain value ('social equity') that I cherish. We can also include here reasons deriving from 'institutional facts' (Searle, 1995; 2001, p. 56–7).

The rational justification of the answer to the question of Topic 1 seems to depend on three main considerations, articulated through four argumentation schemes.[10]

---

10   For reasoning, consider only the pattern of reasoning without the critical questions.

Teleological Considerations. Practical reasoning and argumentation are teleological in that goals are instantiations or manifestations of a general purpose or aim. If, for example, x is an institution created with the mission G+, we consider that G* can be a manifestation of G+. By arguing that G* results from that objective, one attributes to x a reason for assuming G*. To illustrate:

Major Premise: NATO's mission is to actively contribute to world peace and security (G+)

Minor Premise: Helping Ukraine increase its defensive power (G*) will actively contribute to world peace and security

Therefore, it is plausible to suppose that,

Conclusion: NATO has a reason to help Ukraine increase its defensive power.

More formalized we arrive at:

Argumentation Scheme 1. Assumption of Objectives by Teleology[11]
(AS1)
Premise 1:   x has G+ as its finality
Premise 2:   G* belongs to G+
              Therefore, plausibly
Conclusion: There is a reason for x to assume G*

By definition, answering CQs in plausible argumentation is essentially contextual: it depends on the circumstances (in the sense of 'circumstance' explained above).

Satisfactorily answering the following critical questions provides a plausible justification:

Critical Questions for Argumentation Scheme 1
CQ1:   How does G+ really correspond to the finality of x?
        How can G+ not correspond to the finality of x?[12]

---

11   We agree with Fabrizio Macagno, who suggested that AS1 can be considered a variant of 'Argument from Commitment' (Cf. Walton, Reed, & Macagno 2008: 335).

12   For this and all critical questions we assume there can be more than one response/reason. We use the singular wording only for the sake of simplicity of presentation.

CQ2:   How is G* really a particular case of G+?
        How could G* not be a particular case of G+?

Value Considerations. These are considerations involving moral or social values, *sensu lato*, regarding both individual and collective behaviour (e.g. 'Individual Well-Being', 'Collective Well-Being', 'Keeping a Promise', 'Honesty', etc.)

    For this consideration, we have two types of cases in mind. The first regards the assumption of your value as positive (V+). For example, if you are a political leader who values fairness (V+), you may consider it to be positively promoted by taxing the rich to help fund a free national public health system (G*). In the second type, G* may not directly promote any obvious value. It may, however have consequences positively valued by x, for instance, to quench thirst, thus giving x another kind of reason to assume G*.

    In the first case, the argumentative scheme from positive values (Walton, Reed, & Macagno, 2008, p. 321) generally applies, but with two caveats. First, there are no critical questions associated with the scheme in the literature so, using the method described above, we have taken the liberty of formulating them ourselves. Second, since we formulated these questions to focus on the correct application, in a given context, of the essential term with inferential power, we will propose a slightly modified, simplified version of the scheme that clearly isolates the (only) expressions we consider essential.

> **Argumentation Scheme 2. Argument from Positive Values**
> AS2)
> Premise 1:   value V is positive (= V+)
> Premise 2:   V+ positively values G*
>   Therefore, plausibly
> Conclusion: There is a reason for x to assume G*
>
> **Critical Questions for Argumentation Scheme 2**

CQ1:   What reasons are there for attributing a positive value to V?
        What reasons could count against attributing a positive value to V?

CQ2:   What reasons are there for the positive evaluation of G* by V+?
        What reasons are there for G* not to be positively evaluated by V+?

The answer to CQ1 will likely involve the participants in *a substantial* discussion regarding *values*. The answer to CQ2 most likely consists in the demonstration of the *relevant relation between V+ and G**, which may require sub-argumentation. For instance, if V+ is 'to promote peace' and G* is 'to reinforce military power in Ukraine' there is definitely room for sub-argumentation.

Argumentation Scheme 3. Argument from Positive Consequences
(AS3)

Premise 1:   If G* is realized by x, then the consequences K1, …, Kn will occur.
Premise 2:   K1, …, Kn are to be valued positively.
Therefore, plausibly
Conclusion: There is a reason for x to assume G*

We have 'unfolded' the single premise put forward by Walton, Reed & Macagno (2008, pp. 332–3; cf. Walton, 2013a, p. 102) into two premises to permit a critical question to specifically focus on two issues in two premises. This is because in the actual argumentative process it is possible to accept one of the premises and deny the other, deny both, or accept both. The use of the infinitive in the second premise is deliberate, for it allows a discussion (CQ2) on the positive evaluation: x, the proponent, can positively value K1, …, Kn, but in argumentation, y, the opponent, can value them negatively or be neutral. If we indexed the evaluation to x, the second premise would become undisputable (it would consist in the truism, stated by x, that x values K1, …, Kn positively) and we think that it should be able to be discussed. We have also suppressed the original version's CQ3, because we think that it should be carried out in Topic 3 (where pro and con reasons are pondered), as we shall show below.

Critical Questions for Argumentation Scheme 3

CQ1:   What makes it plausible that G* has K1, …, Kn as consequences? How could G* not have K1, …, Kn as consequences?

CQ2:   Why should K1, …, Kn be positively valued? How could K1, …, Kn not be positively valued?

Operational Consideration. One last basic aspect that could enable x with a reason to assume G* is to know if x has the ability to (contribute to) realize G* and, also, if his assumption of G* is or is not idle regarding the realization of that objective. At this point, it is important to note that we are not here addressing the ability of x to carry out the means. This argument scheme will appear again later in the model where it can be appropriately used for that purpose. For its use in providing a reason to aim at the goal, 'ability' here is to be understood more broadly as 'in a position'. For example, consider a husband who needs to pick his wife up from the train station. His having a driver's licence that his children do not have could be an ability reason that enables him to realize the goal of picking her up. This ability, however, says nothing about the car being functional or otherwise available for him to perform the means of driving to get there. We recognize that if hard pressed, the ability reason does indeed boil down to an ability to perform the means, but think an important part of early practical reasoning rests on a preliminary consideration of an agent's being in a position – having the ability – to achieve the goal. Since this scheme is used again later while addressing the means, it is not crucial to follow the ability chain all the way to the end of the performance of the means here. Its second instantiation functions as an appropriate check on the means at that point.[13] The argumentation at this stage should instantiate the following scheme:

Argumentation Scheme 4. Argument from Ability[14]
(AS4a)

Premise 1:   G* should be positively valued

Premise 2:   x has the ability to realize G*

Premise 3:   x's ability to realize G* is a necessary/enabling condition for the realization of G*

Therefore, plausibly,

Conclusion: There is a reason for x to assume G*

---

13   Changing AS4 to appear in two places is new in this articulation as compared to the forthcoming publication, but has been added in consultation with João Sàágua who maintains reservations about such a decision.

14   Given that we have not found a similar scheme in the literature, we hope this constitutes a modest contribution to the field.

In this scheme, the agent goes from the existence of a reason to carry out G* (Premise 1) to the existence of a reason for *x, and not any other agent*, to carry out G* (Premises 2 and 3). If x were not in a position to realize G*, or if the assumption of G* by x was unnecessary, in the sense that G* would occur anyway even if x would not assume it, then there would not be this reason for x assuming G*. The two reasons are not the same. Going back to the NATO example, the reason to carry out G* (NATO helping Ukraine increase its defensive power) can be, for example, because it 'Promotes Peace', which is considered to be a positive value (V+), while the reason for x assuming G* (and not any other agent) can be, for example, because NATO is in a better position to negotiate with the quarrelling parties, an *operational* reason.

Critical Questions for Argumentation Scheme 4a

CQ1:   How does x have the ability to realize G*?
   What could prevent x from realizing G*?

CQ2:   To what extent is the assumption of G* by x a necessary/enabling condition for the realization of G*?
   Which y exists (such that y≠x) whose ability to realize G* is a necessary/enabling condition for the realization of G*

Since we think that, normatively, it only makes sense to argue through the instantiation of AS4 if its Premise 1 has already been proven by *another type* of argumentation (AS1-AS3), we consider Premise 1 as assumed. For that reason, it does not need the association of a CQ. Further, this illustrates the importance of following the argumentation schemes in order since if AS4 were used first, it would be unsupported.

  Let us imagine that all four schemes were employed on real argumentation and that all of the CQs were answered successfully. While it may mean there are reasons to assume G*, it does not yet mean that x should assume G*. This is because the reasons x has for assuming G* are *pro tanto* and not *pro toto*. We thus have now to consider 'the other side of the scale'.

Stage 2. Topic 2

Topic 2 involves argumentation aimed at founding an answer (positive or negative) to the problem: 'Is G* compatible with other goals, G1, ….Gn, that x has or should have?' As stated, the problem seems to lead to the idea that x has to consider the compatibility of G* with virtually *every* objective (including aims) that x *has*, as well as with *all* those that x *should have*. To complicate the situation further, we assume that there is no safe and sound method for the calculation of (in)compatibilities! Although seemingly extremely complex, this is not an intractable situation. It will be sufficient *to use the Principle of Charity* and, in a sense, *to reverse the burden of proof*. Given that we are speaking of human reasoning, using the principle of charity we shall assume from the outset that x is 1) usually not (knowingly) self-contradictory and 2) is not an inherently evil person. Obviously, there is place for a margin of error: x can overlook conflicting goals, or accidentally contradict himself and x can have instances of evil. Generally, however, we take x to be consistent and morally neutral or good by default, thus reversing the burden of proof and leaving it to the opponent to build an argument to challenge a positive answer to Topic 2.

With these qualifications in mind, we consider the argumentation supporting an answer to the Topic to rest on the four following argumentation schemes: the first argues against the assumption of G* because this objective promotes a negative value; the second argues against the assumption of G* because this objective contradicts or inhibits a positive value; the third argues against the assumption of G* because the enactment of this objective has negative consequences; the fourth is neutral regarding values and evaluations and simply argues that there is an operational incompatibility between G*, if assumed by x, and other objectives x has *already assumed*. The first three schemes thus concern objectives that x *should have*, while the fourth concerns the objectives x *has*. Since these schemes occur in the overall model twice – here as applied to the goal and later as applied to the means – on the flowchart they have been labelled 'a' and 'b' respectively, as was done with AS4 above. This double applicability is represented in each scheme with G*/M*. Assumptions A2 and A3 above address our view on the flux between means and goals.

Argumentation Scheme 5. Argument from Negative Values[15]
(AS5)
Premise 1:   the value V is negative (= V-)
Premise 2:   V- negatively values G*/M*
Therefore, plausibly,
Conclusion: There is a reason for x not to assume G*/M

Critical Questions for Argument Scheme 5
CQ1:   What reasons are there for attributing a negative value to V?
        How could V not have a negative value?

CQ2:   What reasons are there for the negative evaluation of G*/M*
        by V-?
        How could G*/M* not be negatively valued by V-?

Argumentation Scheme 6. Argument Contradicting Positive Values[16]
(AS6)
Premise 1:   Value V is positive (V+)
Premise 2:   G*/M* contradicts (or inhibits) V+
Therefore, plausibly
Conclusion: There is a reason for x not to assume G*/M*

Critical Questions for Argument Scheme 6
CQ1:   What reasons are there for attributing a positive value to V?
        How could V not be valued positively?

CQ2:   What reasons are there to indicate G*/M* contradicting (inhib-
        iting) V+?
        How could G*/M* be congruent with V+?

---

15   See, Argument from Negative Value (Walton, Reed, & Macagno 2008: 321; Walton
      2013a: 103). The two remarks made above regarding the argumentative scheme on
      positive values apply, *mutatis mutandis*, also here, hence, we will not repeat them.
16   Although this scheme cannot be found as such in Walton, Reed, & Macagno
      (2008), or in Walton (2013a), it is considered a variant of the 'Argument from
      Values', easily manageable out of the two schemes that are 'traditionally' includ-
      ed in it.

Argumentation Scheme 7. Argument from Negative Consequences

(AS7)

Premise 1:   If G*/M* is realized by x, the consequences K1, …, Kn will occur

Premise 2:   K1,…, Kn are to be negatively valued

Therefore, plausibly

Conclusion: There is a reason for x not to assume G*/M*

Similar remarks to those made for AS3 are applicable, *mutatis mutandis*, here – with the exception of the ones regarding the existence of a second premise, which, in this case, already appear in the original formulation of the scheme (see Walton, Reed, & Macagno 2008: 332–333; Walton, 2013a: 102).

Critical Questions for Argument Scheme 7

CQ1:   What makes it plausible that G*/M* has K1, …, Kn as consequences?
       How could G*/M* not have K1, …, Kn as consequences?

CQ2:   Why should K1, …, Kn be negatively valued?
       How could K1, …, Kn not be negatively valued?

Other schemes related to AS7 are rightly described by Walton, Reed & Macagno (2008, pp. 318–344) as in the realm of practical reasoning, but are not primitive. A discussion of how they relate to the primitive scheme would be an excellent topic for a further paper.

Argumentation Scheme 8. Argument from Operational Incompatibility[17]

(AS8)

Premise 1:   G is an objective already assumed by x

Premise 2:   G and G*/M* are operationally incompatible

Therefore, plausibly

Conclusion: There is a reason for x not to assume G*/M*

---

17   Given that we have not found a similar scheme in the literature, we hope this constitutes a modest contribution to the field.

It is noteworthy that in premise 1, the objectives of x are restricted to *those already assumed by x* and *do not* include those that the opponent considers x should assume. If the latter were included, the reference to values and evaluations would be unavoidable (and, for that, we already have AS5, AS6, and AS7). This scheme aims at situations in which the existence of a contradiction between the realization of certain objectives already assumed by x and the new objective x is considering to assume, G*, is 'pointed out'. In this way it remains focused on operational incompatibility rather than ideological incompatibility.

Critical Questions for Argumentation Scheme 8

CQ1:   What reasons are there for taking G as an objective already assumed by x?
What reasons are there for doubting x already assumed G?

CQ2:   What makes G and G*/M* operationally incompatible?
How could G and G*/M* not be operationally incompatible?

Let us imagine that AS5, AS6, AS7, and AS8 were actually instantiated in a concrete argumentation and that they passed their respective CQs; or that at least one of them did. In that case, the practical argumentation that took place guaranteed that x has up to four, but at least one reason, for *not* assuming G*. Does this mean that x should not assume G*? Not yet. The reasons x has for not assuming G* are *pro tanto*, and not *pro toto,* so we have to decide between the two sides – we have to weigh the pros and cons. That is the purpose of Topic 3.

Excursus. Negotiation of Objectives
Before analysing the argumentation belonging to Topic 3, it is appropriate to consider a situation in which a contradiction has arisen between an objective to be assumed, G*, and another objective !G. Instead of arguing about which objective is preferable (Topic 3), one can argue for a modification to one of those objectives, or both. This is arguing through negotiation. It is important to emphasize precisely this *argumentative aspect of the negotiation,* because negotiation writ large does not have to be rational as in the case of pure threat, blackmail, or bribery.

A simple[18] example. Let us imagine that x already had as an objective, G, 'To act in an environmentally-friendly way'. Now, x wins the lottery and can buy the car he always dreamt of. x is considering a new objective, G*, 'To buy a Ferrari Testarossa'. Knowing the Testarossa's high fuel consumption, it is obvious that the second objective is incompatible with G – his aim to be environmentally friendly. To mark the incompatibility of G with G*, we will represent G as !G – in which '!' is used to point out that contradiction with G*. Now, in a certain sense, x can choose between determining which of the two objectives, G* or !G, is preferable, thus going to Topic 3. Or x can try to modify one of the two objectives, or both, in order to make them compatible. Let us imagine that x enters into a process of *argumentative negotiation* in which he will have to determine *how far he can go in the modification of his objectives*, G* and !G, *in order to make them compatible, but also to think that he is still assuming that part of those objectives that x considers essential*. Let us imagine, for instance, that at the end of the negotiation (either with y or with himself) x modifies G*, 'To buy a Ferrari Testarossa' into, 'To buy a Citröen DS5'. There is a clear sense in which the objective, G*, was preserved and modified: x now has the objective of buying a more environmentally-friendly car that, although not a Ferrari, is still a fancy car. This is now, so to speak, the 'car of his dreams' insofar as it achieves the assumption of both goals rather than requiring the sacrifice of one.

We consider that the argumentative process just illustrated consists in an instantiation of the following Argument Scheme.

Argumentation Scheme 9. Argument Based on Reasonable Negotiation[19]
(AS9)
AS9.1. Variation on !G
Premise 1:   !G and G*/M* are contradictory
Premise 2:   Modifying !G into Gi preserves the essential in !G

---

18   But it is obvious that this kind of situation can be enormously complex. For example, consider the negotiation between social stakeholders: employers, unions, and government.

19   Given that we have not found a similar scheme in the literature, we hope this constitutes a modest contribution to the field.

Premise 3:   Gi is compatible with G*/M*
Therefore, plausibly
Conclusion: x should assume Gi (instead of !G)

AS9.2. Variation on G*
Premise 1:   !G and G*/M* are contradictory
Premise 2:   Modifying G*/M* into G' preserves the essential
               in G*/M*
Premise 3:   G' is compatible with !G
Therefore, plausibly
Conclusion: x should assume G' (instead of G*/M*)

AS9.3. Variation on G* and !G
Premise 1:   !G and G*/M* are contradictory
Premise 2:   Modifying !G into Gi preserves the essential in !G
Premise 3:   Modifying G*/M* into G' preserves the essential
               in G*/M*
Premise 4:   Gi and G' are compatible
Therefore, plausibly
Conclusion: x should assume Gi and G' (instead of !G and
               G*/M*, respectively)

Since premise 1 works as an assumption imported from the previous
Topic we do no need to question the incompatibility. Accordingly, these
are the remaining Critical Questions associated to this scheme (in any
of its variations).

Critical Questions for Argumentation Scheme 9
CQ1:   How do the modifications of !Gi into Gi or of G*/M* into G',
       respectively, preserve the essential aspect(s) of each of the ini-
       tial objectives?
       How might the modifications of !Gi into Gi or of G*/M* into
       G', respectively, diminish/jeopardize the essential aspect(s) of
       each of the initial objectives?

CQ2:   What makes the schemes resulting from the proposed modifi-
       cations (Gi/!Gi and G'/G*/M*) compatible?
       How might the schemes resulting from the proposed modifica-
       tions (Gi/!Gi and G'/G*/M*) be incompatible?

CQ3:   What incompatibilities with other objectives x has or should have result from the proposed modifications (Gi/!Gi or G'/G*/M*)?

How could one resolve these resulting incompatibilities?

We consider the burden of proof of CQ1 to be on the side of the proponent and the burden of proof in the case of CQ2 and CQ3 to be on the side of the opponent (in line with what was stated about that matter in Topic 2).

Stage 2. Topic 3.
Topic 3 receives a situation of incompatibility (insurmountable, or overlooked, by negotiation) between G* and one or more objectives/aims that x has or should have as input and has to provide a founded answer to the question: 'Is G* preferable to the goals, !G1, …!Gn, with which it is incompatible?'. Intuitively, and simplifying slightly, if G* is preferable to another objective, !G, with which it is incompatible, then that other objective should be abandoned and the reasoning should progress to Topic 4. If !G is preferable, then !G should (continue to) be assumed by x and the practical reasoning on G* ends here. To found the answer to the question, *an argumentative process in favour of the preference* for G* or for !G should be carried out. In addition, that argumentative process should take into account the specific results obtained in Topics 1 and 2. Let us see this in greater detail.

     Topic 1 allowed for four types of reasons in favour of the assumption of G*, of which at least one would have been positively associated to G*. Obviously, we are talking about several types of reasons. As mentioned above, this means that there can be several particular reasons in favour of the assumption of G* by x that are specimens of each one of those types. Topic 2 allowed for four types of reasons against the assumption of G*, of which at least one would have been negatively associated to G*. Here, we are again talking about *types of* reasons and so there can be *several particular reasons* against the assumption of G* by x that are specimens of each one of those types. This time, the *particular* reasons positively associated to G* in Topic 1 are the ones that must be weighed *against* the *particular* reasons negatively associated to G* in Topic 2. Resolving Topic 3 rationally articulates this process of 'weighing' the reasons in favour/against the assumption G* by x.

Argumentation Scheme 10. Argument Based on Rational Preference[20]

(AS10)

10.1 Variation in favour of G*/M

Premise 1:   !G and G*/M* are contradictory

Premise 2:   (R*±1, …, R*±n)G*/M*

Premise 3:   (!R±1, …, !R±n)!G

Premise 4:   (R*±1, …, R*±n)G*/M* are preferable to (!R±1, …, !R±n)!G

Therefore, plausibly

Conclusion: x should assume G*/M* (and abandon !G)

10.2 Variation in favour of !G

Premise 1:   !G and G*/M are contradictory

Premise 2:   (R*±1, …, R*±n)G*/M*

Premise 3:   (!R±1, …, !R±n)!G

Premise 4:   (!R±1, …, !R±n)!G are preferable to (R*±1, …, R*±n)G*/M*

Therefore, plausibly

Conclusion: x should assume !G (and abandon G*/M*)

In theory, the R*± of premise 2 were all identified in Topics 1 and 2. In concrete argumentative practice, if the matter is very serious, one can submit G* to a 'second round' of those very same Topics. It is almost certain that the !R± of premise 3 were not all identified when having G* and not !G in sight. Hence, one should now go through those two Topics *having !G in sight*. For that we do not need additional Topics or schemes. Thus, the individual reasons (R*±1, …, R*±n) and in (!R±1, …, !R±n) are just those reasons identified positively in Topic 1 or negatively in Topic 2 for G* or !G.

What is being weighed ({R*±} vs. {!R±}) in premise 4, when the *relation of preference* is applied?

---

20   Given that we have not found a similar scheme in the literature, we hope this constitutes a modest contribution to the field.

1) All evaluations of reasons considered positive vs. all evaluations of reasons considered negative. For example, 'In this situation S1, it is preferable to slightly sacrifice the value V1+, in order to greatly implement the value V2+'; or another example, 'In the situation S2, it is preferable to slightly sacrifice K1+, to be able (in the future) to enjoy the positive consequence, K2+, that will increase the well-being of x in a more sustained way'.

2) Ideally, the subjective probabilities (possibly conditional) that x and y believe to be associated to both: (a) the success in realizing G* or, alternatively, !G; and (b) the 'coming to existence' of the reasons {R*±} and {!R±} as a result of the realization of that G*, or !G, respectively. For example: G* has a 0.9 probability of being realized, its R+I has a 0.7 probability to be implemented if G* is realized (repeated for each R+i) and its R-I has a 0.2 probability of occurring if G* is realized (repeated for each R-i); and a similar reasoning for !G and its associated reasons.

3) Most of the time and in alternative to 2), the *subjective 'plausibilities'* which are equal to 2, but replacing the probabilistic quantification, between 0 and 1, by qualifiers such as 'very', 'few', and so on. We are not often capable of specifying a probability, even a subjective one, for the success of G* or of reasons that we believe to be associated to G*.[21]

4) The *beliefs* regarding the *circumstances* of the situation.

Importing the critical questions from above for the input premises (1–3), let us now see the CQs for premise 4. As noted above, the critical questions for this scheme depart slightly from the usual 1:1 ratio of critical question per term with inferential power. This is because, we believe, the term 'preference' entails both aspects of goodness and probability. Thus, the questions here, while focused only on the single term 'preference', address both of its component parts.

---

21 How people pick and assign probability and weight to reasons is an interesting and important question, but one which is ultimately a matter for psychologists. Further work could, however, address how one ought to assign probability and weight to reasons (See Lord & Maguire 2016).

Critical Questions for Argumentation Scheme 10

CQ1:   What makes the standard(s) used for the evaluation of the rea-
       sons associated with the goals/means the best for this situation?
       Why might the standard(s) used for the evaluation of the reason
       associated with the goals/means not be the best for this situation?

CQ2:   What makes the standard(s) used to assess the probability or
       plausibility of the reasons used to justify the assumption of the
       goal/means and of the goal/means being assumed the best for
       this situation?
       Why might the standard(s) used to assess the probability or
       plausibility of the reasons used to justify the assumption of the
       goal/means and of the goal/means being assumed not be the
       best for this situation?

In short, these questions are challenging the goodness in the reasons and
the accuracy of the probability of success, respectively. These questions
are notoriously difficult to formulate because it is at this point where
argumentation theory meets choice theory, and both meet moral theory.

Stage 3. Topic 4
This Stage begins when the objective G* has been rationally founded. We
then need to associate one or more means to it. Here is where what has
been called instrumental practical reasoning (Wallace 2014; Hitchcock
2011), or 'means-ends reasoning' begins. The first question each agent
will ask about the means can be vague, of the kind: 'Is there any way to
realize G*?' As an answer to this question, the agent expects that rep-
resentations of actions he can carry out and whose implementation will
bring him closer to the realization of G* until G* is realized will 'pop into
mind' by a process that he usually does not control well. The agent might
use his experience from similar cases along with other tools to marshal
every means offered to him in any more or less fortuitous, more or less
contextual, way. There are also studies pointing out the importance of
automated or innate heuristics to 'the finding of means'.[22] Though the
creation of reasons is a matter generally investigated within psychology,
it also has philosophical implications (Smith 2010; Mizrahi 2014).

---

22   See, e.g., Tversky & Kahneman (1974).

From a philosophical point of view, one part of the important work consists in classifying the means into necessary or possible options. Accordingly, the problem of interest to us at Stage 3, Topic 4 is, 'Are there means, M1, …, Mn (M*), to realize G* which are simultaneously necessary and sufficient for x to achieve G*?'. As a matter of fact, this problem includes two questions: 1) 'are there means that *have* to be used if one intends to realize G*?' and 2) 'are those *all* the means needed to realize G*?'

A 'Yes' to the first question means that, without the use of those means by x, x is not able to realize G*. In that case, those means *have to* be used. Imagine a situation where the only way to beat a competitor is to kill him. While perhaps necessary (and say, sufficient) it is not usually something that *should* be done. Here we deal only with what has to be done, with the foresight of knowing that the 'should' is addressed shortly (Topics 6 and 7).

However, a 'No' to the first question *does not* necessarily imply that there are no means available to realize G*. It can also imply that there are several alternative means that x *can choose between*. In that case, there is the problem of knowing if those means are *sufficient*. That is the problem of Topic 5. If they are, and given that x can choose, then the discussion about the 'best means' will be opened. That problem will be dealt with in Topic 6 and eventually 7. At present, a 'No' to the first question is simply tantamount to going to Topic 5, where we will deal with the problem of the existence, or not, of sufficient means to realize G*.

Let us now imagine that we answer 'Yes' to the question regarding the necessary means (NM). Now we need to know if the set {NM1, …, NMn} is sufficient to realize G*; or if, *some other means besides* {NM1, …, NMn} will still be needed. This is the *raison d'être* for our second question in Topic 5: 'are the means necessary to realize G* sufficient in conjunction?' If the answer is 'Yes', we go to Topic 6. If we answer 'No', it means that the set {NM1, …, NMn} has to be *supplemented with more means from which a choice will have to be made*. 'Supplemented', because {NM1, …, NMn} is not sufficient to realize G* and 'a choice will have to be made,' because if there were no choice, the added means would actually be necessary and would belong to {NM1, …, NMn}. Schematically: to realize G*, x has to use {NM1, …, NMn} and then still use

M1, or M2, or Mn (which do not belong to {NM1, …, NMn}), because without at least one of these means, x cannot realize G*.

Schematized, we arrive at:

> **Argumentation Scheme 11. Necessary Condition Argument**[23]
> (AS11)
> Premise 1:   x has the objective of realizing G*
> Premise 2: ....{NM1, …, NMn} are necessary means for x real-
>   izing G*
> Therefore, plausibly
> Conclusion: x has a reason to carry out {NM1, …, NMn}

There is no CQ for premise 1 because it works as an assumption, in the sense already explained.

> **Critical Questions for Argumentation Scheme 11**
> CQ1:   What makes it plausible that {NM1, …, NMn} are necessary
>   means for x realizing G*?
>   How could any of these means be suppressed while still allow-
>   ing for the realization of G*?

Obviously, the argumentation aimed at showing the (defeasible) necessity of any of the means has an extremely high sensitivity to context. Think of the necessary means for becoming President – a detailed discussion of what those means are will involve differing argumentative schemes derived from AS11, in the sense already explained. For that reason, the 'course' the argumentation will take in each case is difficult to predict. It also seems reasonable to accept that it is the proponent of the argument instantiating AS11 who has the initial burden of proof regarding the necessity of {NM1, …, NMn}, given that it requires the

---

23   We have distanced ourselves from the 'Necessary Condition Schema' (Walton, Reed, & Macagno 2008: 323–324) for two reasons. The formulation of the 'Alternatives Premise' removes the necessity of each of the means by using the expression 'at least one of', making them optional amongst themselves. Also, the formulation of the 'Selection Premise' and of the conclusion clearly shows that the scheme's objective is to select 'the best mean' (referred as 'Bi'), which will only be dealt with by us in Topics 6 and 7. These are not meant as criticism of the scheme, but as justification for not considering it primitive and thus not using it here.

use of those means or the waiving of the realization of G* on reasonable grounds.

Despite this highly contextual character, it is known that an argument in favour of a *necessary* condition (in the present case, a mean) ends with a conclusion in the form. 'If not {NM1, …, NMn}, then not G*', in which the conditional is material, and which is, thus, logically equivalent (by contraposition) to 'If G*, then {NM1, …, MNn}'. Any of those forms of the conclusion can be used to build a *plausible* argument in favour of the necessity of each one of the NMi, an argument whose premises will be, as already stated, strongly dependent on context. To determine if the means are sufficient, we can use the following scheme:

> **Argumentation Scheme 12. Sufficient Condition Argument**[24]
> (AS12)
> Premise 1:   x has the objective of realizing G*
> Premise 2:   If x carries out {NM1, …, NMn}, then x realizes G*
> Therefore, plausibly
> Conclusion: x has a reason to carry out {NM1, …, NMn}

There is no CQ for Premise 1 because it works as an assumption in the sense already explained.

> **Critical Questions for Argumentation Scheme 12**
> CQ1:   How does carrying out all of the necessary means guarantee the realization of G*?
> How might G* remain unrealized despite carrying out the necessary means?

The same observations we made regarding the sensitivity to context of AS11 and its CQ are applicable to AS12 and its CQ, therefore we will not repeat them. The same can be said regarding the matter of the burden of proof.

Likewise, and despite this highly contextual character, it is known that an argument in favour of a *sufficient* condition (in the present case,

---

24   See the previous footnote. Similar considerations can be applied here in regard to '22.3. Sufficient Condition Schema' (Walton, Reed, & Macagno 2008: 323–324).

a mean) ends with a conclusion of the form: 'If {NM1, …, NMn}, then G*', in which the conditional is material, and which is, thus, logically equivalent (by contraposition) to 'If not G*, then not {NM1, …, NMn}'. Any of these forms of conclusion can be used to build a *plausible* argument in favour of the sufficiency of each mean, taken in conjunction – an argument whose premises will be, as already stated, strongly dependent on context.

Finally, it is one thing to recognize necessary and sufficient means, and another to ensure that x has the ability to carry them out. Further, in our view, being the *only* one able to carry out the means can provide an *additional* reason for x to pursue M*. Accordingly, we re-use the scheme for ability from Topic 1 and include it in Topic 4 (and Topic 5 if it should be necessary), in a similar but slightly modified way. The similarity is that, as before, it still only becomes necessary after the other schemes in the Topic have been addressed. In other words, for this Topic, if there are no means then there is no need to consider the agent's ability to carry out the non-existent means. The modification is that, in this instantiation, ability does not refer to 'being in a position' but rather, more directly, to 'being able to perform'.

Argumentation Scheme 4b. Argument from Ability
(AS4b)
Premise 1:   M* are necessary and sufficient (or at least sufficient) for G*
Premise 2:   x has the ability to realize M*
Premise 3:   x's ability to realize M* is a necessary/enabling condition for the realization of G*
Therefore, plausibly,
Conclusion: There is a reason for x assuming M*

In this scheme, we go from the existence of a reason to carry out G* (Premise 1) to the existence of a reason for *x* to carry out M* (Premises 2 and 3). For the same reasons as presented in Topic 1, if x did not have the ability to realize M*, or if the assumption of G* by x was unnecessary, in the sense that G* would occur anyway even if x would not assume it, then there would not be a reason for x assuming M*.

Also for the same reasons as mentioned with this scheme in Topic 1, there are only two Critical Questions needed for this scheme.

Critical Questions for Argumentation Scheme 4b

CQ1: How does x have the ability to realize M*?
What could prevent x from realizing M*?

CQ2: To what extent is the assumption of M* by x a necessary/enabling condition for the realization of M*?
Is there any y (such that y≠x) whose ability to realize M* is a necessary/enabling condition for the realization of M*

Stage 3. Topic 5

An agent only arrives at this Topic if the prior argumentation leads to a negative answer to the question 'Are there means M1, …Mn which are simultaneously necessary and sufficient for x to achieve G*?' (Topic 4). If the Topic reveals that there are necessary but not sufficient conditions, then we are then directed to Topic 5 while bringing with us a set of necessary means (if they were also sufficient we would have gone to Topic 6, without going through Topic 5). However, this difference between having or not having means does not substantially affect the formulation of the scheme, which, in reality, is nothing more than our well-known AS12, now formulated in a more general way: AS12G (here, 'G' means 'General').

Argumentation Scheme 12. Sufficient Condition Argument[25]
(AS12G)
Premise 1: x has the objective of realizing G*
Premise 2: If x carries out {NM1, …, NMn} and {SM1 or, …, or SMn}, then x realizes G*
Therefore, plausibly
Conclusion: x has a reason to carry out {SM1 or, …, or SMn} (in addition to the reasons x may have to carry out {NM1, …, NMn})

Since the question 'Are there necessary means to realize G*?' would have been positively answered in Topic 4 and since a negative answer would be inconsequential and leave us only to consider the sufficient means, no specific CQ on them is provided here.

---

25   See the previous footnote.

The critical question associated to this scheme, then, addresses the sufficient means.

Critical Questions for Argumentation Scheme 12G

CQ1: How does carrying out at least one of SMi, where SMi belongs to {SM1, …, SMn} (in addition to {NM1, …, NMn}, if there are such) guarantee the realization of G*?
How might G* remain unrealized in spite of carrying out all of these means?

All of the remarks made in Topic 4 about that version of AS12 are naturally applicable to AS12G and so we will not repeat them.

Further, because any number of new sufficient conditions may have been introduced in this Topic, AS4b applies here as well.

If the concrete argumentation that instantiates AS12G is negatively concluded, that means that there are no sufficient means to realize G* and the reasoning/argumentation stops here.

If the concrete argumentation instantiating AS12G is positively concluded, that means that there is more than one means M1, …, Mn (that is sufficient) for x realizing G*, i.e. there several possible means for x realizing G*.[26] In this case x may choose the one that he considers *the best means*. As expected, the choice of the best means is a process subjected to argumentation. This takes us to Topics 6 and 7.

Stage 4. Topic 6
Topic 6 involves an argumentative process aimed at founding an answer (positive or negative) to the question: 'Are the selected means, M1, …, Mn, compatible with the objectives G1, …, Gn, that x has or should have?' In this sense, the Argumentation Schemes and respective Critical Questions to be used *are exactly the same as* the ones proposed for Topic 2, as is immediately perceivable if we replace, in the formulation of the problem, 'the selected means, M1, …, Mn' by 'the objective G*?' (as formulated in Topic 2). In line with what was stated above, we consider

---

26 Of course, if there are also necessary (but not sufficient) means, it will be necessary to combine them through a distribution of conjunction over disjunction. Thus: {NM1, …, NMn} and SM1, or {NM1, …, NMn} and …, or {NM1, …, NMn} and SMn.

that the burden of proof is on the side of the opponent here as well (even if this is x with 'second thoughts', before going into action, on whether he should or should not use the means M1, …, Mn, to realize G*).

We thus consider the schemes AS5 to AS9 to be reproduced here, along with their respective CQs and what we stated in their regard in Topic 2. The only difference is that the schemes and critical questions here employ the M* side of the G*/M* option where M* indicates 'set of selected means'. We now simply need to add an illustration and an explanation.

It suffices to recall our presidential 'murderer example' (mentioned in Topic 4): there may not be any incompatibility between the objective 'To be President of the Portuguese Republic' and *all the other* objectives that x has or should have. But if, at a given time, the means chosen by x to realize this objective is 'To get his most direct rival candidate killed', then this means will surely clash, no matter how efficient it is, with *several other* objectives or aims x has or *should* have.

Explanation: Three cases to be considered.
Case 1. If, among M1, …, Mn, only necessary means are to be found, then the conclusion that one of those means is incompatible with G1, …, Gn, immediately leads us to the argumentative process taking place in Topic 7.

Case 2. If, among M1, …, Mn, several sufficient means are to be found (thus resulting from a list of alternative means corresponding to the affirmative answer to the question of Topic 5), then if *some* of those alternative means, *but not all*, are considered incompatible with G1, …, Gn, through the argumentative process taking place in Topic 6, that may permit us to select only the compatible ones (given that, hypothetically, any one of them is sufficient to realize G*) which immediately leads us to Stage 5 and a decision to act.

Case 3. In the case of the sufficient means, only if all of them (that is, all possible means) are considered incompatible with G1, …, Gn, will we be directed to the argumentative process of Topic 7.

Also in regard to Topic 6, it will be possible to try a procedure of negotiation like the one described in the Excursus and associated with AS9. With this supplementary proviso: the potential modifications to be introduced into M1, …, Mn, cannot remove the effectiveness of any of these means making them *no longer sufficient* to realize G*.

Stage 4. Topic 7

Topic 7 involves an argumentative process aiming at founding a (positive or negative) answer to the question: 'Are the means M1….Mm, preferable to the goals !G1, ….!Gn, with which they are incompatible?'. AS10 (and variants) with their respective Critical Questions can also be used here, as can be immediately perceived if we replace the occurrences of 'M1, …, Mn' by 'G*' in the formulation of the problem. Likewise, the comments we associated to the presentation of AS10 and its CQ in Topic 3 are applicable here with the same caveat that the M* option is to be used in this Stage. Hence, nothing else needs to be added at this time.

Stage 5. Decision

In Stage 5, the final stage, there is not exactly a problem to be dealt with and to be answered, so this stage does not contain a Topic. It is only the matter of capitalizing on the reasoning process and on the argumentative course realized in the previous Topics, whether one has gone through all the Topics or just some of them. Obviously, the process may be stopped at any time, simply by answering 'No' to Topic 1, or from then on answering 'No' two consecutive times. If that happens, the agent may decide either not to act, or to make an appropriate modification to the goal or means (depending on where the process was stopped) and begin again with the modification in place. If the process is not stopped, however, and we have arrived at Stage 5, then it is now just the matter of making a decision (practical reasoning) or recommending the action (practical argumentation) founded on all the process, or courses, which started at Topic 1. Therefore, if we consider Γ as the best formulation of the argumentative thread that started at Topic 1 and ended in Stage 5, we can propose:

- For practical reasoning: 'Given that I accept Γ, I justifiably (do not) intend to carry out M1, …, Mn, to realize G*'.
- For practical argumentation: 'Γ being given, the recommendation that x carries out M1, …, Mn to realize G* is (not) justified'.

Conclusion

To conclude, given the detail in the explanations above, we will start by risking a limited amount of repetition for the sake of clarity. The above presentation constitutes our efforts to contribute a new integrated model of practical reasoning and argumentation to the field. Imagining a human agent in any given circumstance, the model begins by asking if the agent has a reason to alter the current state of the world. With the aim of providing a model extending beyond mere instrumental reasoning, we have included consideration of the motivations for aiming at a goal as the first step in explicit practical reasoning and argumentation.

The complete model is composed of five stages: Stage one addresses the agent's motivation for action; Stage two is concerned with the proposed goal and other goals; Stage three concerns available means for achieving the proposed goal; Stage four deals with the relation between the means and other goals; and Stage five is simply the decision to act or not. If the agent progresses through all five Stages, they will have reasonable grounds for deciding to act. If they are stopped at any stage, they will then have reasonable grounds for not acting.

In order to license moving from one Stage to the next, the agent must answer one or two 'Yes' or 'No' Topic questions. In any case, an affirmative answer results in a 'green light' to move to the next Stage. In some cases, a negative answer or 'red light' will lead to another Topic and thus a second chance to move to the next Stage. In other cases, a negative answer leads straight to a conclusion not to act.

Each Topic questions an aspect of the general theme of the Stage. Answering 'Yes' or 'No' to the Topics is not, however, based merely on the free thinking or intuitions of the agent. In order to reasonably answer the Topic questions, the agent must have reasons supporting their answer. Those reasons can be specified using an appropriate argument scheme. The model indicates what we consider to be the basic, necessary schemes to justify an answer to each one of the Topics, though in practice an agent may of course use schemes over and above the provided list.

Importantly, the reasons which emerge from the schemes are to be considered *pro tanto*, or contributory reasons, in the way that Jonathan

Dancy (2004) has characterized them. This consideration is important because of two major implications it carries through the reasoning. First, it means a reason on one side is not, by itself, enough to license moving to a conclusion to act or not act. The questions and schemes are set up in oppositional fashion so that contributing reasons from both sides can be weighed. For example, an agent using the schemes associated with Topic 1 could come up with four reasons to pursue the goal. Rather than jumping straight to a conclusion to pursue it, however, Topic 2 is aimed at finding reasons not to pursue it. Only after both reasons for and reasons against have been addressed is the agent free to look for means.

Second, reasons being contributory also means that one reason may outweigh all opposing reasons. In other words, the number of reasons and weight provided to one side of the 'Yes' or 'No' answer are not in a strict relationship. Thus, even though there may be four reasons for accepting the proposed means and one reason against, that one reason may outweigh the other four.

We have also aimed to improve the way moral considerations are included in the model in a few important ways. First, we have provided a way to include consequentialist and other moral considerations. While consequences are addressed during the selection of both the goal and the means, any given moral principle can be used in our Argumentation Scheme 10, which performs an overall weighing between a proposed goal or means and incompatible alternatives an agent has or should have. Rather than stipulate a moral authority for these decisions, a Critical Question for the Scheme asks the user to justify why the moral standard they have chosen is best, and to account for the exclusion of others. In the absence of a universally agreed authoritative moral theory, we think that the best that can be done at present is to argue for the selection of the chosen standard in rational use at any given time.

Second, we have included moral evaluation of the means and not just the goal. Such an inclusion may have a much bigger impact than at first appears as was partially demonstrated through our example of wanting to become president but killing to do so. While there may be nothing wrong with having a goal to become president, if one overlooks the moral component of killing to get there, an important check on the decision-making process has been overlooked.

Finally, when filled out, the use of such a model provides many opportunities for the evaluation of the reasoning or argumentation. While a full theory of evaluation is better suited for another paper, at this point it suffices to point out that evaluation can take the form of pointing to Argumentation Schemes that were not used when they could have been, pointing to poor usage of the Argumentation Schemes that we employed, assessing the quality of the answers to the Critical Questions, and acknowledging the inclusion or lack thereof of goals an agent should have. In all of these cases, an evaluator can pinpoint an exact component in the process and recommend a systematic solution for improvement.

References

Audi, R. (2006). *Practical Reasoning and Ethical Decision.* New York: Routledge.

Baumtrog, M. D. (2015). 'Improving Practical Reasoning and Argumentation'. *Dissertation*. NOVA University of Lisbon.

____(2017). 'Reasoning and Arguing, Dialectically and Dialogically, Among Individual and Multiple Participants'. *Argumentation*. Online first: DOI 10.1007/s10503-017-9420-3.

Broome, J. (2001). 'Are Intentions Reasons?', in C. Morris & A. Ripstein (eds.), *Practical Rationality and Preference: Essays for David Gauthier*. Cambridge: Cambridge University Press: 98–120.

____ (2002). 'Practical Reasoning'. In J. Bermúdez & A. Millar (eds.), *Reason and Nature: Essays in the Theory of Rationality*. Oxford: Oxford University Press: 85–111.

____ (2004). 'Reasons'. In R. J. Wallace, M. Smith, S. Scheffler, & P. Pettit (eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. Oxford: Oxford University Press: 28–55.

____ (2013). *Rationality Through Reasoning.* Chichester: Wiley Blackwell.

Dancy, J. (2004). *Ethics Without Principles.* Oxford: Oxford University Press.

Fairclough, I., & Fairclough, N. (2012). *Political Discourse Analysis: A Method for Advanced Students.* London: Routledge.

Hitchcock, D. (2011). 'Instrumental Rationality'. In P. McBurney, I. Rahwan, & S. Parsons (eds.), *Argumentation in Multi-Agent Systems: 7th International Workshop, ArgMAS 2010 Toronto, ON, Canada, May 10, 2010 Revised, Selected and Invited Papers*. Berlin, Heidelberg: Springer: 1–11.

Lord, E., & Maguire, B. (eds.). (2016). *Weighing Reasons.* New York: Oxford University Press.

Mizrahi, M. (2014). 'The Problem of Unconceived Objections'. *Argumentation* 28: 425–436.

Scanlon, T. M. (1998). *What We Owe to Each Other.* Cambridge: Harvard University Press.

Searle, J. (1995). *The Construction of Social Reality.* New York: The Free Press.

_____ (2001). *Rationality in Action.* Cambridge: MIT Press.

Smith, M. N. (2010). 'Practical Imagination and its Limits'. *Philosophers' Imprint* 10(3): 1–20.

Tversky, A., & Kahneman, D. (1974). 'Judgment Under Uncertainty: Heuristics and Biases'. *Science New Series* 185(4157): 1124–1131.

van Eemeren, F. H., Grootendorst, R., & Snoeck Henkemans, A. F. (2002). *Argumentation: Analysis, Evaluation, Presentation.* Mahwah: Lawrence Erlbaum Associates.

Walton, D. (1990). 'What is Reasoning? What is an Argument?' *The Journal of Philosophy* 87: 399–419.

_____ (2007). 'Evaluating Practical Reasoning'. *Synthese* 157: 197–240.

_____ (2013). *Methods of Argumentation.* Cambridge: Cambridge University Press.

_____ (2013). 'Value-Based Practical Reasoning'. In K. Atkinson, H. Prakken, & A. Wyner (eds.), *From Knowledge Representation to Argumentation in AI, Law and Policy Making: a Festschrift in Honour of Trevor Bench-Capon*. London: College Publications: 259–282. Retrieved from http://www.dougwalton.ca/papers%20 in%20pdf/13TrevVBPR.pdf.

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation Schemes.* Cambridge: Cambridge University Press.

Arguing, bargaining and persuading in constituent processes[1]

GIOVANNI DAMELE
FRANCESCO PALLANTE

1. Introduction

In this essay we will outline the constituent process that took place in Italy between 25 July 1943 (the fall of Fascism) and 1 January 1948 (the effective promulgation date of the Republican Constitution), paying special attention to the debate that arose during the sessions of the Constituent Assembly between 2 June 1946 and 27 December 1947.

This essay is based on Jon Elster's paper, *Arguing and bargaining in two constituent assemblies*[2], which focuses on the study of constituent processes and, above all, constituent debates. The theoretical issues raised by the concept of constituent power[3] will remain in the background for reasons of space.

---

1   Giovanni Damele is the author of section 8; Francesco Pallante is the author of sections 3–7 and 9. Sections 1, 2 and 10 are the fruit of a shared reflection.
2   ELSTER 2000 (the paper was based on a conference held at Yale University in 1991).
3   In Italian constitutionalist doctrine, see at least: BARILE 1966, DOGLIANI 1986 and 1990, MORTATI 1972, GRASSO 1985, PACE 1997, RESCIGNO 1996. In international literature, the following writers have recently worked on standardizing constituent processes with regulatory intents: GINSBURG-ELKINS-BLOUNT 2009, ELKINS-GINSBURG-MELTON 2009; GINSBURG 2012.

2. The Elster diagram

The summary indications in Elster's essay can be broken down into seven main steps of constituent processes:

1) convocation of the constituent assembly
2) choice of the delegate selection procedure
3) definition of the mandate of the assembly and of the delegates
4) verification of the delegates' credentials
5) choice of the decision-making procedure to be adopted during the assembly
6) discussion and approval of the constitution by the assembly
7) ratification of the constitution

The Norwegian scholar first distinguishes between the first two steps and the following ones, claiming that the convocation of the constituent assembly and the choice of the delegate selection procedure are performed by independent authorities who are not part of the assembly, while all of the other steps refer to assembly decisions (under penalty of establishing a puppet-body that simply enacts the will of others)[4].

This point is pivotal: only a process that is possibly "self-founded" can truly be considered as "constituent", whereas a process which follows from a decision by a pre-existing body[5] should be considered as being "constituted".

---

4   ELSTER 2000: 358–59 (in addition to 361–62, 364 and 366).
5   RESCIGNO 1996: 34 *et seq*.

3. The convocation of the constituent assembly

A) Starting from the first step, according to Elster, the authority convening the constituent assembly may be[6]:

–   the constitution itself, if it provides for periodic constituent assemblies (Thomas Jefferson's idea). The idea that the constitution provides for its total review is similar (see Article 193 of the Constitution of the Helvetic Confederation; something similar happened in Spain in 1978)[7].
–   an authority different from that of the State to which the constitution will apply: for example, an occupying power (as in Western Germany and Japan after the Second World War). Today, the hypothesis that a non-State authority has an international character prevails (consider Kosovo and Afghanistan).
–   a provisional government resulting from a revolution (as in France with the 1789 and 1848 revolutions) or a coup (the case of the Ghana Constitution of 1992). Naturally, a revolution or coup does not always involve the use of violence (since the constituent assemblies convened by bodies belonging to the previous constitutional organization – like the French National Convention which approved the 1793 Constitution – would allegedly fall under this hypothesis).
–   a mixed government or a seat of concertation between the old regime and the opposition (which is what happened in Poland in 1989 with the so-called Round Table Agreement).

B) In the Italian case, the convocation of the Constituent Assembly may be formally traced back to two documents:

–   Law Decree no. 151/1944, which provides for the election of a Constituent Assembly to pass resolutions on all aspects (including the institutional form: monarchy and republic?) of the "new State Constitution" (the so-called first provisional Constitution).

---

6   ELSTER 2000: 358, note 64.
7   *Contra* PACE 1997: 8 *et seq*., in his opinion these hypotheses amount to the practice of a constituted power.

–   Delegated Decree no. 98/1946, which redefines the powers of the Constituent Assembly establishing that the institutional issue must be solved by popular *referendum* to be held at the same time as the election of the assembly itself (so-called second provisional Constitution).

Both documents were adopted by provisional governments of national unity, hence by external authorities according to the Elster diagram: the former was enacted by the Bonomi government, the latter by the first De Gasperi government. In both cases, there were also influences from the monarchy and the Allies.

Elster notes that, since it is not clear whether the new or old rules apply at the time of constituent processes, the relationships between the old and new regimes often give rise to a logic paradox by virtue of the framers' attempt to give their actions formal legitimacy based on pre-existing legal arrangements[8].

In the Italian case, a clear symptom of these problems is the different format taken by the two decrees: one is a law decree, the other is a delegated decree. The difference is explained by the fact that the document dated 1944 – adopted under Article 3 of Law no. 100/1926 – establishes the delegation of legislative powers to the government; consequently, only subsequent documents may be delegated documents. However, since law decrees are temporary, there is the problem of the conversion of Law Decree no. 151/1944 into law: the fifteenth transitory and final disposition of the Constitution did so, however, terming it a … "delegated decree".

4. The delegate selection procedure

A) As concerns the selection method of assembly members, Elster stresses that the deciding authority should be different from the one convening the assembly; however this would create a "puppet assembly"

---

8   ELSTER 2000: 360.

since the delegates would allegedly be selected based on their loyalty to the convening authority[9]. The Norwegian political scientist does not, however, linger on subjects regarding the selection of the delegates[10].

B) In the Italian case, the document establishing the methods of formation of the Constituent Assembly is Delegated Decree no. 74/1946, which introduced a strictly proportional electoral law. The selection of the assembly members was referred to the people, convened on 2 June 1946 to vote by universal suffrage for the first time. It is clear, however, that the definition of the electoral system was extremely important and was largely affected by the fact that the parties were not familiar with their electoral "weight" (since they opted, Rawls-style, for the choice that would have secured them in case of defeat).

In the Italian case, two independent authorities were involved: the provisional government, which established the electoral system, and the Italian people, who elected the assembly delegates. This partly proves the Elster diagram wrong since the provisional government intervened in both phases – convocation and selection of the members – which preceded the formation of the Assembly.

5. The definition of the mandate

A) As concerns the definition of the mandate, the question is whether there is a constraint. On the practical side, it is easy – as Elster writes – for the authority convening the constituent assembly (or, more rarely, the one selecting its members) to try to influence the outcome of the work by constraining the mandate of the delegates. It is, however, equally easy for the constituent process to get away from its "creator" (which is what happened to Louis XVI)[11].

---

9   ELSTER 2000: 359, in particular note 65, which gives the example of "the body of 66 men convened in China by Yuan Shikai in 1914 to give his rule a semblance of legality through a 'constitutional compact'".
10   ELSTER 2000: 359.
11   ELSTER 2000: 361–62.

Aside from its practical infeasibility, the Norwegian scholar does not give a completely negative opinion of putting constraints on the mandate because it can be useful in order to strengthen the threats made during the bargaining[12].

B) In the Italian case, it could be believed that a type of constraint was applied to the mandate of the Constituent Assembly members by not allowing them to decide on the institutional question.

As already mentioned, Law Decree no. 151/1944 initially stated that the selection was the responsibility of the Assembly, then Delegated Decree no. 98/1946 re-examined the issue, referring the decision to the people by referendum. This is one of the key steps of the entire Italian constituent process. Despite the fact that the 1944 decree resulted from a comprehensive agreement involving the Allies, CLN (National Liberation Committee) parties and the monarchy, Lieutenant Umberto broke the understanding, asking the people to decide on the institutional question. Along the same lines, the (mainly monarchist) Italian Liberal Party and the Christian Democracy (DC) were concerned about the gap between the Party's positions and those of its electorate (considered more inclined to institutional continuity). De Gasperi also convinced the Allies to support the referendum as a solution. On the contrary, the left-wing parties and the Actionists favoured abiding by the original provisions of Law Decree no. 151/1944[13].

The issue was first raised by De Gasperi on 10 October 1945 under the Parri government. A complex debate arose which became intertwined with the two additional issues of whether or not the Constituent Assembly should also act as a law-making body and whether the Constituent Assembly should pass resolutions by a simple or a qualified majority. The Council of Ministers (and the so-called Cabinet, a selected committee made up of the ministries representing the CLN parties) was involved in these debates during the sessions held between 19 February and 2 March 1946 until the secretaries of the DC, PSI and PCI agreed to approve Delegated Decree no. 98/1946 under the De Gasperi government.

---

12   ELSTER 2000: 363.
13   The event is carefully re-enacted in RICCI 1996: 449–459.

6. The verification of credentials

A) Elster considers the verification of the credentials of the constituent assembly delegates a logic paradox (in addition to the paradox on the legal origin of the constituent assembly): the assembly cannot verify the credentials of the delegates without taking office, but – at the same time – it cannot take office without having first verified the credentials of its member[14]. An independent audit would be necessary, but this would undermine the independence of the assembly. In France, the issue was hotly debated during the Estates General convened by Louis XVI, and a solution was found on the basis of the following consideration: "It is impossible to believe that the majority of those who present themselves as delegates should not have valid credentials"[15].

B) In the Italian case, the Constituent Assembly established a Committee responsible for verifying the credentials of elected members, which it did, making some replacements.

   The issue raised no specific debates because the Assembly worked on the basis of the parliamentary procedures of the pre-Fascist Lower House (*Camera dei Deputati*), which had already provided that the Council was competent for the election.

7. The choice of the decision-making procedure

A) Elster mentions the following problems regarding the choice of the procedures to be adopted during the assembly to discuss and approve the constitution[16]:

---

14   ELSTER 2000: 366.
15   ELSTER 2000: 366.
16   ELSTER 2000: 367 and 404–405.

1) the duration of the sessions
2) whether the constituent assembly should also act as a law-making body
3) how to decide on the allocation of the time between law-making and constitution-making
4) the possibility to establish constitution-drafting or problem-solving committees
5) whether to proceed in closed sessions or open the debates to the public. (Elster believes that closed sessions encourage bargaining and arguing because it is easier to change opinion.)[17]
6) the quorum and the voting method (by "person" or by group[18]? And, if by "person", by roll call, show of hands, division of the assembly etc.?)
7) the procedure of transforming votes into decisions

B) In the Italian case, we must refer to Article 4 of Delegated Decree no. 98/1946 that applied to the Assembly the procedures promulgated for the Lower House in July 1900, as repeatedly amended until 1922. The Constituent Assembly itself made some "additions". In short, the decision on the procedures was made by an independent authority (the provisional government, although with the Assembly's tacit consent) in lieu, as the Elster diagram provides, of the Constituent Assembly.

Let us now look more closely at the individual profiles identified by Elster. Starting with the duration, Delegated Decree no. 98/1946 established an eight-month deadline from the first session (held on 25 June 1946). This deadline could be extended by no more than four months. After using such an extension, the Assembly applied for an additional six-month extension (plus a few days, to set the deadline at 31 December 1947) since the duration was decided partly by an independent authority (the provisional government), and partly by the Constituent Assembly itself.

As concerns law-making powers, the CLN was internally divided between the DC and PLI, which wanted, with the support of the Allies,

---

17   ELSTER 2000: 410–411 (on the usefulness that the involved parties may change their ideas, also see p. 385).
18   The question was hotly debated during the French Constitutional Assembly (ELSTER 2000: 367–368).

to limit the competences of the Assembly to constitution-making top-
ics (leaving ordinary law-making powers to the government) and left-
wing parties, which thought it preferable to refer the decision to the
Constituent Assembly. The question had to do with the fear that the
left-wing parties, if they won the election, might have exploited the As-
sembly's powers in order to establish "revolutionary" legislation. The
problem was addressed together with those of the subject in charge of
making the institutional choice and the quorum of the Constituent As-
sembly. In this case, the solution is also found in Delegated Decree no.
98/1946 (Art. 3). Although the provision ratified the victory of moder-
ate parties, the Assembly was permitted to indicate bills that, though
not part of its law-making competence, were to be submitted for its
resolution[19].

Moving on to the time of the sessions, the division between consti-
tution-making and "ordinary" activity[20] was decided with the planning
of the sessions by the Constituent Assembly itself: 375 public sessions
were held, 173 of which focused on the discussion and approval of the
new Constitution.

Concerning the establishment of constitution-drafting committees,
a Constitutional Committee was appointed under the presidency of
Meuccio Ruini in order to prepare the Constitution draft. The 75-mem-
ber Committee was split into three sub-committees: (1) citizens' rights
and duties; (2) constitutional organization of the State, (which was then
split into two branches: one on executive power, one on judicial power
and the Constitutional Court; a selected committee was also established
for the regional system); and (3) economic and social relationships. The
topics of the first and the third sub-committees partially overlapped, so
a Coordinating Committee was established to unify their work. Even-
tually, a Drafting Committee (with 18 members) prepared the text of
the final draft, coordinating and harmonizing the work of the three
sub-committees.

---

19   For a re-enactment of the event, see RICCI 1996: 449–459.
20   In addition to the opinions on the draft legislative decrees, the Assembly's main
     non-constituent activities were the vote of confidence for the De Gasperi Govern-
     ments II, III and IV; the approval of the budget laws for 1947 and 1948; and the
     ratification of the peace treaties signed in Paris on 10 February 1947.

As concerns whether the sessions were public or closed, the Constituent Assembly's activity was public, but that of the Committee and its various sub-committees was not since the Italian people were not permitted to directly attend the entire constitution-making process[21]. The debate was covered by the press[22] and the Ministry for the Constituent Assembly also provided extensive information.

The quorum and voting method were governed by the pre-fascist procedures of the Lower House[23]. Voting (by "person") was by ballot for the final approval of the bills and by sitting and standing in all other cases (unless ten members asked for voting by division, fifteen by roll call, and twenty by ballot). The quorum was the majority of participants. Under these rules, the Constitution was approved on 22 December 1947 with voting by ballot by roll call (out of 515 participants, 453 votes in favour, 62 against).

As concerns the "procedure for transforming votes into decisions"[24], once approved, the Italian Constitution was enacted by the provisional Head of State, Enrico De Nicola, on 27 December 1947 and published immediately in an extraordinary edition of the Official Gazette no. 298 of the same day. It came into effect on 1 January 1948.

8. The discussion

A) Elster believes that constitution-making projects represent a "paradigmatic case" useful to highlight two types of dialogue: *arguing and bargaining*. Those "types" are exhibited in "their most striking forms"[25] in constituent assemblies, which are more polarized than ordinary law-making bodies and oscillate between "higher law-making"

---

21   PALDIN 2004: 46 (and 48–49) writes, on the other hand, that the sessions were often "quite confidential".
22   POMBENI 1995: 93–96.
23   RICCI 1996: 449–459.
24   ELSTER 2000: 367.
25   ELSTER 2000: 347.

and "sheer appeal to force". Elster also introduces a third type of (in Elster's words) "speech act": "*rhetorical* statements aiming at *persuasion*", though he is uncertain about its proper analytical characterization. However, it seems that the distinction between the three types is a question of "motives". Not those of the speakers, which Elster distinguishes as "reason", "passion" and "interest", but the motives that the speakers ascribe to their audience. Rhetoric may perhaps be defined "by the feature that its practitioners appeal to the passions of their audience rather than to their reason or self-interest" since "in some debates reason speaks to reason; in others, interest to interest; in still others, passion to passion"[26]. Somehow, Elster's triadic model seems to reflect Aristotle's three-part division of persuasion modes, where *logos* can easily be matched with arguing, *ethos* can refer to bargaining (where the criterion of credibility is key) and *pathos* could match "rhetoric" within the meaning given by Elster, i.e. an appeal to the passions of the audience. However, Elster does not develop this parallelism, nor does he develop the analysis of the more genuinely pathetic components of assembly discussions.

A.1. (*Arguing and bargaining*). For Elster, rational arguing is subject to criteria of *validity*, and promises or threats to criteria of *credibility*[27]. The former recalls Habermas's theory of communicative action, binding a speaker aiming at *understanding* and not sheer *success* to "three validity claims: propositional truth, normative rightness, and truthfulness". Even speakers who are not "genuinely moved by impartial considerations of the common good", but whose concerns are "purely self-interested", may still be forced or induced "to substitute the language of impartial argument for the language of self-interest"[28]. This substitution would be the fruit of the *civilizing force of hypocrisy*, thanks to which "a speaker who wants to *appear* as aiming at understanding must also appear to be committed to these claims"[29]. Consequently, "one need not always oppose impartiality and self-interest" since "one may offer an argument *from* self-interest *for* impartiality". Such a typical argument is the

---

26   ELSTER 2000: 371, no. 116.
27   ELSTER 2000: 372.
28   ELSTER 2000: 349.
29   ELSTER 2000: 373.

so-called "veil of ignorance"[30], or, in general, represents all those cases in which "apparently self-interested behaviour" may actually be guided "by impartial concerns"[31]. Authenticity, or sincerity, on the other hand may be traced back – at least in one of its versions – to consistency, which is not incompatible with what could be defined as *argumentative hypocrisy*[32]. It would be important to distinguish between authentic changes of opinion, which *per se* would not reveal an inconsistent argument, and actual opportunism[33].

As concerns arguing, Elster distinguishes arguments as "tending to be" either consequentialist or deontological. Roughly speaking: appealing "to overall efficiency" or "to individual rights". The latter, as well as those "based on the public good" and those "which rely on some version of utilitarianism", are considered impartial because of their generalizability[34]. Starting from these arguments, the framers would somehow prove that they are motivated by impartial *reason*, despite being permeable to self-interested considerations, as we have noted. Elster qualifies framers as *imperfectly rational*[35].

While rational discussion is supposed to be based only on the "power of the better argument", constitutional bargaining, by contrast, rests on "resources that can be used to make threats (and promises) credible". Such resources may be extra-political or intra-political. The latter include the exchange of concessions[36]. The use of these resources is strictly dependent on the framers' ability to make them credible: the framers' credibility affects the credibility of their threats and/or their promises.

A.2. (*Pure and impure types*). In the analysis of the actual arguments of the two assemblies, the two types seem to translate in a sequence of "mixed" or "impure" cases. On the one hand, a strategic use of (apparently) non-strategic arguments is not only possible but common, and in some cases desirable: in these cases, "self-interested actors often try",

---

30   ELSTER 2000: 374.
31   ELSTER 2000: 388.
32   ELSTER 2000: 413.
33   ELSTER 2000: 377.
34   ELSTER 2000: 378–379.
35   ELSTER 2000: 380.
36   ELSTER 2000: 392.

in their own self-interest, "to ground their claims in principle". On the other hand "bargainers often try to present their threats as warnings"[37]. The difference between threats and warnings would lie in the fact that the former "are statements about what the speaker *will do*", whereas the latter are "statements about what *will* (*or may*) *happen*, independently of any actions taken by the speaker"[38]. In the former case, "self-interested actors" appeal to an *impartial equivalent*. In the latter case, bargainers substitute a *factual equivalent* of a threat.

From the point of view of arguing, what happens in reality is not actually a "*perfect fit* between partial interest and impartial arguments" but a "*maximal* fit"[39]. The reasons for this substitution between partisan arguments and impartial arguments may vary. First, "if others believe that one is truly arguing from principle, they may be more willing to back down". Second, "legislative coalitions tend to use public-regarding language as a 'subterfuge' for what is in reality a deal among special interests"[40]. Third, "by citing a general reason one might actually be able to persuade others"[41].

In any case, thanks to the *civilizing force of hypocrisy*, arguing "tends to yield more equitable outcomes than bargaining", even when it is purely strategic and based on self-interest, because it will prevent "the strong from using their bargaining power to the hilt". In this case, "the optimal impartial equivalent", able to "yield more equitable outcomes", will be the one that "dilutes" the self-interest of the strong by "taking some account of the interest of the weak"[42].

B) Elster believes that "the most important requirement" of a bargaining theory should be "that we are able to specify what will happen during a temporary breakdown of cooperation"[43]. In short, how the constituents can get out of an impasse caused by a non-cooperative – even

---

37   ELSTER 2000: 405–406.
38   ELSTER 2000: 415.
39   ELSTER 2000: 406.
40   ELSTER 2000: 408.
41   ELSTER 2000: 408.
42   ELSTER 2000: 413.
43   ELSTER 2000: 398.

if temporary – situation. An example is the debate which led to the final version of Article 29 of the Constitution[44].

It is known that this article was the result of a difficult writing process that significantly affected its text (and its subsequent interpretations). This is due, on the one hand, to the relevance of the topics (family and marriage) for the members of the DC Party and, in general more for Catholics, and on the other, to the difficult position of the Communist Party, which was not willing to be confined to markedly secular positions in view of the future political election (in which Catholic votes would have been critical) and could not ignore that a significant portion of its electorate (and PCI members) held positions very similar to DC Party members on some aspects of the matter, and divorce in particular.

The drafting of Article 29 was marked by a series of "temporary breakdowns of cooperation" that were overcome through exchanges of concessions, recourse to "intra-political" resources (for example a strategic placement of available votes of a certain parliamentary group) and the strategic use of procedures. A first impasse caused by the opposing positions of the sub-committee on the indissolubility of marriage and the definition of family as a "natural and fundamental unit of society" was overcome by Moro and Togliatti through an exchange of concessions which translated into a new version that divided the theme into two articles, included a (more ambiguous) definition of "natural society" in the article on the family and a (more vague) reference to "unity" (in lieu of indissolubility) in the article on marriage. The second breakdown of cooperation occurred because of Togliatti and the PCI members' failure to take strategic recourse to the "intra-political" resource of voting. In the Constitutional Committee, in keeping with the compromise reached with Moro, Togliatti led his group to approve the formula "family as natural society". However, when the vote was cast for the second article, the amendments against the clause of "indissolubility of marriage" were

---

44   CAPORRELLA 2010. The documents of the Constituent Assembly are available online in the "Previous legislatures" section of the website of the Lower House ("Constituent Assembly" section) or on the "Birth of the Constitution" website, by Fabrizio Calzaretti: www.nascitacostituzione.it. The debates leading to the final drafting of Article 29 were held on 26 July, 13 September, 30 October, 5–7, 12–13 and 15 November 1946 and 15 January 1947, as concerns the I Sub-committee; 4–8, 10–11 and 17 March, 15, 17–19, 21–24 April 1947, as concerns the Assembly.

not approved. Therefore, the result could be re-balanced only in the last voting session through the strategic use of the Assembly's procedures. The request for secret voting by twenty Assembly members allowed the compromise to be re-established, leading to the approval by three votes of an amendment by Socialist Party member Grilli to remove the reference to "indissolubility". As for bargaining and the recourse to warnings, it is interesting to consider the speech by Lelio Basso in the 7 November 1946 session of the I Sub-committee, during which the Socialist Party member warned that the "categorical request for the indissolubility of marriage" might have "led to a break-up of the Sub-committee".

As concerns the arguments used, the trend was not so much to argue on principles, but to present consequentialist arguments. In both cases, the objective was to present a general and impartial point of view. This is especially clear in the speeches by DC Party framers, whose main concern was to prove that the need for inclusion of "indissolubility" in the constitutional text went beyond compliance with certain religious principles. Corsanego declared that "the authentic Italian population, even in its humbler classes, has clear, well-defined and tangible arguments" on the family, recalling the authority of "common sense". He also resorted to a consequentialist argument, noting that "divorce represents the dissolution of the family and a poisonous germ for its establishment, as is proved in all the countries where it is accepted". La Pira also claimed to insist on permanence because he had been persuaded by "an increasingly determined confirmation in the scientific field of the indissolubility of marriage considered as a structural element of the family". He then stressed that the DC Party members wanted to include indissolubility in the Constitution because it should concern marriage as such, and not as a sacrament (hence, also civil marriage). Thus, he considered important "to overcome the question of the parties, so that the claim made is not the claim of the DC Party, but of the entire Italian population".

The intention of the DC Party members to present a theme from an impartial point of view is clear. Such a theme, they admitted, was essential for their political and religious position. The position of those opposing inclusion of the phrase "natural society" with respect to the family in the Constitution, and those opposing the introduction of the indissolubility of marriage appeared more delicate. Following the

compromise between Moro and Togliatti, the PCI Party members opposed the latter but not the former. On the other hand, the Socialists and some Liberals opposed both. However, neither the Socialists nor Togliatti posed the question of divorce. In a way, because of the necessary tactic in view of the future election, the position of the left-wing parties appeared more "defensive" and ambiguous, which exposed them to the accusation of inconsistency and opportunism.

Concerning the "exchange of concessions" between Moro and Togliatti, we should note the recourse to the strategic use of ambiguity, which allowed for an agreement on a formulation sufficiently ambiguous to provide different interpretations that were more or less directly consistent with the actors' different points of view. For Moro, keeping the expression "natural society" in the article allowed him to overcome a merely confessional position and affirm the "natural rights" of the family, while the definition of "natural society" had no legal effect for Togliatti and did not imply *per se* the conclusions that La Pira wanted to draw (i.e. "the indissolubility of the bond").

9. Modes of ratification

A) The modes of ratification – Elster states – are necessary to confer "downstream" legitimacy on the constitutional document approved by the constituent assembly.

The following possible modes can be identified[45]:

– right of veto of the independent convening authority (but how can a constituted power influence the constituent power?)
– ratification by the people through *referendum* or an *ad hoc* convention
– no additional ratification to the final vote of the constituent assembly.

B) In the Italian case, the approval by the Constituent Assembly was sufficient. However, it should be considered that there had already been

---

45   ELSTER 2000: 371.

popular involvement because of the institutional *referendum* in whose wake the Constituent Assembly operated.

10. Conclusion

As concerns the execution of the constituent proceedings, we can say that the Elster diagram shows actual endurance capacity as applied to the Italian constituent process. Some steps may be overestimated (such as the verification of delegates' credentials), others underestimated (like the definition of the assembly's powers). However, as a whole, the steps appear to match and the most critical profiles identified by the Norwegian scholar are present in the Italian case.

The conclusion from the arguing theory perspective appears more critical. The analysis of the debate on Article 29 highlights some weaknesses of Elster's model. The first is the role played by rhetoric and more generally the concept of rhetoric he refers to. The second problem is the articulation of the debates according to the arguing/bargaining opposition and the analytical utility of such an opposition. A third problem could be the strategic use of ambiguity. As concerns the latter two, we can supplement Elster's model with analytical instruments from arguing theory and bargaining theory.

Qualifying bargaining and arguing as "types of dialogue" appears to provide a better description of their characteristics and especially their co-presence within the same dialectic interactions through the concept introduced by Walton of the "*dialectical shift*". In this case, it is not just a "combination" of types of dialogue but a more-or-less gradual transition from one type of dialogue to the next. It may be a legal or illegal transition, and in the first case, the second type of dialogue is included in the first, thereby further developing, constructively, a dialectical shift[46]. Thus, bargaining can be transformed (more or less accidentally) into a persuasive dialogue. These transitions then make it

---

46   WALTON 1992: 138 and MACAGNO 2011: 106.

possible to assess the context within which a certain argument may be fallacious or not, replacing a rigid concept of fallacy with a dynamic one linked to the use of an argument in a set type of dialogue. Considering the Elster concept of bargaining, for example, and the central role played by threats or warnings within it, it may be interesting to recall Walton's analysis of the argument *ad baculum*, according to which the criterion distinguishing between a fallacious and a non-fallacious use of the threat is exactly the type of dialogue since it may be legitimate in bargaining but not in a persuasive dialogue[47].

Strategic use of ambiguity plays a key role in overcoming "temporary breakdowns of cooperation". As Eric M. Eisenberg noted, strategic ambiguity favours agreement on an abstraction without committing the bargainers on their potential future interpretations[48]. This is even more important when considering that arguers (or bargainers) in a deliberative setting may have multiple objectives, some of which may even be (partly) contrasting. One can see that this is very similar to what Cass Sunstein defined as "*incompletely theorised agreements*"[49] in a juridical setting, i.e. a communicative strategy that does not minimize but manages ambiguity[50].

The role of rhetoric remains to be defined. Elster's dyadic model appears to be a triadic model, which is missing a component: "*rhetorical* statements aiming at *persuasion*". This is because Elster considers the term "rhetoric" to essentially have a negative meaning, i.e. the common meaning of manipulation, appeal to passions (and not to reason) and demagogy. However, it is reductive to confine rhetoric to a sheer appeal to passions because it is actually the reference theoretical framework of any persuasive discourse. In a deliberative setting, the persuasive purpose combines both arguing and bargaining and eventually the appeal to emotions. As a technique of persuasive discourse, rhetoric expands its scope well beyond a mere appeal to the audience's passions. The point is not even the strategic use of arguments but, more generally, strategic arguing as a discourse technique whose purpose is to persuade

---

47   WALTON, MACAGNO 2007: 75. More in general on the topic, see WALTON 2002.
48   EISENBERG 1984: 231.
49   SUNSTEIN 2007.
50   EISENBERG 1984: 238.

the audience. Therefore, Elster's hierarchy should be overturned since both strategic arguments and bargaining could develop in a deliberative framework and resort to persuasive (rhetorical) discourses.

References

BARILE, P. (1966), *Potere costituente*, "Novissimo digesto italiano", Torino, Utet, 13: 443–450.

CAPORRELLA, V. (2010), *La famiglia nella Costituzione italiana. La genesi dell'articolo 29 e il dibattito della Costituente*, "Storicamente", 6. (http://www.storicamente.org/07_dossier/famiglia/famiglia_costituzione_italiana.htm).

DOGLIANI, M. (1986), *Potere costituente*, Torino, Giappichelli.

DOGLIANI, M. (1990), *Costituente (potere)*, "Digesto delle discipline pubblicistiche", Torino, Utet, 4: 281–287.

ELKINS, Z., GINSBURG, T., MELTON, J. (2009), *The Endurance of National Constitutions*, New York, Cambridge University Press.

ELSTER, J. (2000), *Arguing and Bargaining in Two Constituent Assemblies*, "University of Pennsylvania Journal of Constitutional Law", 2: 345–421.

EISENBERG, E.M. (1984), *Ambiguity as Strategy in organizational communication*, "Communications Monographs", 51: 227–242.

GINSBURG, T., ELKINS, Z., BLOUNT, J. (2009), *Does the Process of Constitution-Making Matter?*, "Annual Review of Law and Social Sciences", 5: 1–23.

GINSBURG, T. (ed.) (2012), *Comparative Constitutional Design*, New York, Cambridge University Press.

GRASSO, P.G. (1985), *Potere costituente*, "Enciclopedia del diritto", Milano, Giuffrè, 34: 642–670.

MACAGNO, F. (2011), *Types of Dialogue, Dialectical Relevance, and Textual Congruity*, "Anthropology & Philosophy",1–2: 101–121.

MORTATI, C. (1972), *La Costituente. La teoria, la storia, il problema italiano*, now in Id., *Raccolta di scritti*, vol. 1, *Studi sul potere costituente e sulla riforma costituzionale dello Stato*, Milano, Giuffrè: 7–343.

PACE, A. (1997), *L'instaurazione di una nuova Costituzione. Profili di teoria costituzionale*, "Quaderni costituzionali", no. 1, pages 7–50.

PALADIN, L. (2004), *Per una storia costituzionale dell'Italia repubblicana*, Bologna, il Mulino.

POMBENI, P. (1995), *La Costituente. Un problema storico-politico*, Bologna, il Mulino.

RESCIGNO, G.U. (1996), *La discussione nell'Assemblea costituente del 1946 intorno ai suoi poteri, ovvero del potere costituente, delle Assemblee costituenti, dei processi costituenti*, "Diritto pubblico", 1: 1–41.

RICCI, A.G. (1996), *Aspettando la Repubblica: i governi della transizione 1943–1946*, Roma, Donzelli.

SUNSTEIN, C.R. (2007), *Incompletely Theorized Agreements in Constitutional Law*, "Social Research: An International Quarterly", 74 (1):1–24.

WALTON, D. (1992), *Types of Dialogue, Dialectical Shifts and Fallacies,* F. H. van Eemeren et al. (eds.), *Argumentation Illuminated*, Amsterdam, SICSAT: 133–147

WALTON, D. (2002), *Scare Tactics*, Dordrecht, Kluwer.

WALTON, D., MACAGNO, F. (2007), *The Fallaciousness of Threats: Character and* Ad Baculum, "Argumentation", 21: 63–81.

III – Ethics and Aesthetics:
Implicit and Explicit Conections

'I'm beginning to know myself. I don't exist.' The Interval of Álvaro de Campos

Bartholomew Ryan

> Começo a conhecer-me. Não existo.
> Sou o intervalo entre o que desejo ser e os outros me fizeram,
> ou metade desse intervalo, porque também há vida …
> Sou isso, enfim …
> Apague a luz, feche a porta e deixe de ter barulhos de chinelos no corredor.
> Fique eu no quarto só com o grande sossego de mim mesmo.
> É um universo barato.[1]

Introduction: Navigating between Philosophy and Poetry

As both an exercise in navigating between philosophy and poetry and what can be gained from doing so, I analyse Álvaro de Campos' seven-line poem displayed above. This poem not only highlights Fernando Pessoa's classic treatment of tedium, masks, the issue of the self and identity, insomnia and the theatre of existence, but also emphasises the concept of the 'interval'. I argue that the interval here has three interconnected aspects when philosophy and poetry come together: the motifs of 'between', the ruin and the creative act – which together locate the nomadic writer. As much a fragment as it is a poem, these lines by Pessoa's heteronym Campos have no title or date. It was published for the first time by the Ática publishers in Lisbon in 1944, nine years after

---

1 Pessoa (2002: 433; 1998: 200). Translation: 'I'm beginning to know myself. I don't exist. / I'm the gap between what I'd like to be and what others have made me, / Or half of this gap, since there's also life … / That's me. Period. / Turn off the light, shut the door, and get rid of the slipper noise in the hallway. / Leave me alone in my room with the vast peace of myself. / It's a shoddy universe.'

The poem also captures the mystery and welcoming puzzle that still remains in the massive body of work after all the confessions, deceptions and masks created in the endless shedding of the self.

The interval can be called a concept as it is a multifaceted idea that runs through Pessoa's work that helps articulate key elements in his artistic strategy as well as his psychology, philosophical perspectives, spirituality, and the fractured cosmology that is at play. Thus, it signifies many things, and – as in so much in Pessoa – the unity of its meaning lies in its multiplicity. Agreeing with Richard Zenith, I see Pessoa's aesthetic and spiritual pursuits as one and the same quest (Pessoa 2016: 14), which is also very much in tune with James Joyce's 'eternal affirmation of the spirit of man in literature' (Joyce 2008: 620). The interval is used so frequently in the Pessoa corpus that we might see it now without surprise, without thinking, and as something even trivial. But it would be unwise to overlook this word as it is a key to the heart of Pessoa's writing.

In exploring the interval, we could dive into various passages scattered across the *Livro de Desassossego*, where the word turns up over thirty times, some passages beginning with the title 'Intervalo Doloroso' or simply 'Intervalo'[4]. There are also poems from Pessoa which explicitly confront the interval in both his Portuguese and English poems – most obviously the poem 'Intervalo' (Pessoa 2006a: 385) in Portuguese from 1935, and 'Meantime' (Pessoa 1994: 202–203) in English (translated as 'Intervalo' by Jorge de Sena) which was his only publication in England in *The Athenaeum* in 1920, and another English poem called "The King of Gaps" from his magical, pantheistic collection under the title *The Mad Fiddler*. The interval was also of interest to Pessoa's closest friend and fellow poet Mário de Sá-Carneiro who is important for the formation of Pessoa's art, especially in his poem "Inter-Sonho" from 1913. Sá-Carneiro ends the poem with explicit reference to the interval: 'Pressinto um grande intervalo, / Deliro todas as cores, / Vivo em roxo e morro em som…'[5]. In this poem we see the obvious connection that the interval has not only to dreams but also to music, and the role that it has

---

4   See especially Pessoa (2012: 76, 98, 113, 196, 199, 227, 286, 347, 373 / 39, 62, 78, 163, 166, 194, 255, 318, 345).

5   Sá-Carneiro (1978: 57) 'I envision a great interval / I go wild in all the colours / I live in purple and die in sound' (my translation).

for the structures in music which is important to the interval for literary usage and dreamscapes in *Livro de Desassossego*.

   There are other poems from Campos which incorporate the interval such as 'Saudação a Walt Whitman' ('Não quero intervalos no mundo!'[6]); an untitled poem with no date where he writes: 'E é no intervalo que existo'[7]; and in one of his greatest poems 'Lisbon Revisited (1926)': 'Compreendo a intervalos desconexos; / Escrevo por lapsos de cansaço; / E um tédio que é até do tédio arroja-me à praia.'[8] Even earlier in the epic 'Ode Marítima', the interval is implicit throughout the poem, beginning with the dramatic opening verses in the image of the gap between the wharf (*cais*) and the ship (*navio*) that set the scene and allusions for his epic, imaginative adventure in piracy, homoeroticism and the reckless life of a seafarer. Campos' seven-line poem appears on the page like a philosophical shipwreck long after his greatest ode has been published in Portugal's greatest modernist moment and magazine *Orpheu*.

1. Aspects of the Interval in the Poem

What is the interval? There are five key components and/or words all turning up in different texts of Pessoa that belong to the family of the interval: *interlúdio*, *intervalo*, *lacuna*, *entreacto*, and English word *gap*. We can see that the *interlúdio* derives from 'inter' signifying 'between' and 'ludus' – signifying 'play'; 'gap' comes from gape, which can be a hole in a wall or ledge and which can also signify the yawn; 'lacuna' from Latin, which means hole or pit; and the *entreacto*, as intermission, literally 'between' the 'act' in the theatre. The interval itself stems from the Latin *intervallum* – 'between' the 'vallum' which signifies 'rampart' or 'wall'. There can be three simple definitions of the interval: as the space between two points or two objects; as the space of time between

---

6   Pessoa (2002: 170): 'I want a world without gaps!'
7   Pessoa (2002: 265): 'It is in the interval that I exist' (my translation).
8   Pessoa (2002: 300 / 218): 'At intermittent intervals I understand; / I write in respites from my weariness; And a boredom bored even of itself casts me ashore.'

events, artistic shows, dates or epochs; and even as a temporary messianic interruption found most powerfully in Western thought in Saint Paul's *rhipé* [twinkling of the eye], Kierkegaard's *Øieblikket* [the moment] and Luther and Heidegger's *Augenblik* [moment of vision].[9]

As mentioned above, I would like to introduce three aspects of the interval in this particular poem. First, the interval includes the standard motif of 'between' – as 'inter', which points to the space between waking and sleeping (insomnia), between sexualities especially in the case of Campos who is inclined towards both sexes, between physical space made metaphorical – such as the water between a wharf and a ship, and in the space between philosophy and poetry.

Second, the interval contains the motif of the 'ruin' – as fragment, disaster, garret, collapse or cracked cosmos. It is that 'man from Porlock'[10] who interrupted one of English literature's greatest unfinished poems 'Kubla Khan'. The interval in this poem is that which is not written, what was going to be written. Instead the reader only has the beginning and end of this magical poem, and the middle part – the interval – is now only in our imagination to discover or recover. It is no wonder that Pessoa obsessed over Coleridge's poem. Like other philosophers and poets, the ruin becomes Campos' place for spiritual renewal, as a 'ruin of all space' (Joyce 2008: 24) to begin again. There is a word coined by Joyce in *Finnegans Wake* that might well fit Campos which is the paradoxical 'chaosmos' (Joyce 1992: 118) – combining chaos and cosmos, encompassing disorder and order, or 'thisorder' (Joyce 1992: 540). The literary theorist Kuberski defines chaosmos as such:

> a unitary and yet untotalized, a chiasmic concept of the world as a field of mutual and simultaneous interference and convergence, an interanimation of the subjective and objective, an endless realm of chance which nevertheless displays a persistent tendency toward pattern and order. (Kuberski 1994: 3)

---

9   Kierkegaard (1980: 82–90); The Bible (1997: ii. 221) [I Corinthians 15:52]; Heidegger (1963: 338).

10 'A man from Porlock' is the title of a piece that Pessoa wrote and which was published on the 15th February 1935 in *Fradique* in Lisbon. It discusses the enigmatic figure who supposedly interrupted the writing of Coleridge's 'Kubla Khan'. Pessoa calls this man from Porlock 'the unknown interrupter' who showed up and obstructed a communication between the abyss and life' [*esse interruptor incógnito, a estorvar uma comunicação entre o abismo e a vida*]. Pessoa (2000: 491; 2005: 54).

Pessoa comes close to expressing this idea of ruin, cosmos and chaos in a poem in 1934 one year before he dies: 'No intervalo cresce o mundo / Com sóis e estrelas sem fim'[11]. Campos is the 'man in the garret [*o da mansarda*]' (Pessoa 2002: 322; 1998: 175) who lives in the symbolic ruin of things, as a man on the margins, the defeated one who keeps going, whose life is a shipwreck but which allows for new beginnings.

The third aspect is the act of creativity. It is in the interval where creative energies can be unleashed. The creative, self-conscious writer thus is locating him or herself in the interval – that space that is vacated, forgotten, or seemingly frivolous. Thus, the writer of the interval can be an ironic nobody writing for no one, acting as the immature child or supplementary writer in the basement or attic. For Whitman, a great inspiration for Campos, the great poets emerge from the intervals and he describes them thus: 'How they are provided for upon the earth, (appearing at intervals,) / How dear and dreadful they are to the earth'.[12]

With these three aspects, the interval implies something physical (in insomnia, sexuality, material space), interdisciplinary (between philosophy and poetry), spiritual and cosmic (in the ruin), and elusively aesthetic and creative (as transitional, marginal, ironic and iconoclastic).

2. A Philosophical Reading of the Poem

Let us read the poem. A few years earlier, Campos wrote that it is in the interval that he exists.[13] Now he goes one step further. The first line 'Começo a conhecer-me. Não existo [I'm beginning to know myself. I don't exist]' relates to the third aspect from above – in the heightened self-consciousness and an awakening and self-knowledge. It is also a self tired of itself which is brilliantly expressed in a poem later in the

---

11   Pessoa (2006a: 235; 2006b: 329): 'In that gap is born the universe / With suns and stars past counting.'

12   Whitman (2004: 44). This poem, called 'Beginners', was first published in the third edition of *Leaves of Grass* in 1860.

13   See note 7.

twentieth century by Patrick Kavanagh in 'The Self-slaved': 'Me I will throw away / Me sufficient for the day' (Kavanagh 2005: 227). Campos' opening sentence is a sudden spark of self-awareness and a perfect way to start this concise poem full of melody for a poem without much rhyme. We have the modernist figure of the interval declaring himself as not existing. Bernardo Soares asked the questions: '… a quem assisto? Quantos sou? Quem é eu? O que é este intervalo que há entre mim e mim?'[14] Here Campos tries to tell it straight. As a follower of Caeiro, knowing oneself is to no longer exist. If the ancient Sanskrit epic – *The Mahabharata* – views the self as one's opposite, then Campos has come to understand not only this but that the beginning of knowledge is to be nothing. And his reality in knowing is that he does not exist. Then Campos tries to unravel this teasing declaration in the first two sentences with the second and third line: 'Sou o intervalo entre o que desejo ser e os outros me fizeram / ou metade desse intervalo, porque também há vida … [I'm the gap between what I'd like to be and what others have made me / Or half of this gap, since there's also life]'. 'I am' (*Sou*) or his being is *the* interval, but an interval that is between what he desires to be and what others have made him. These 'others' leave open for the reader and spectator to see them as Campos' own creator Pessoa, or in the narrative itself of those who see and meet Campos, or us as the readers of today. In this extreme self-reflection, there is the obvious nod to heteronymity but also to the concept of 'inter' itself and the process of creativity. The second part of this sentence in the third line deepens the mystery. He could be half of this interval because there is also life. This is the present and future cosmos, revealing the indifference of the cosmos, the stars or the universe to Campos and man's plight. He can't even be all of this interval. Kant's dark night of cognition grows longer and the stars more distant.

'Sou isso, enfim …[That's me. Period].' This line is the first of four attempts to conclude the opening declaration, but of course (like his much longer, unfinished poems such as *Saudação a Walt Whitman* and *A Passagem das Horas*) he cannot and will not conclude so easily – being that interval, the between, the transition. Even the three dots that

---

14   Pessoa (2012: 222; 2015: 189): '… who am I watching? How many am I? Who is I? What is this gap between me and myself?'

follow 'enfim' already show the uncertainty of finishing. Pessoa and Campos cannot remain silent; instead the repetition unfolds where he tries to write the same poem again and again. And so in not being able to be silent (and Campos is always talking), he gives the order: 'Apague a luz, feche a porta e deixe de ter barulhos de chinelos no corridor [Turn off the light, shut the door, and get rid of the slipper noise in the hallway]'. This is a very vivid image of solitude and call for closure. There is the desire for a return to darkness which alludes to points two (the ruin) and three (creation). He is ordering someone to close the door, to finish it for him, to make the closure for him, and to get rid of that irritating noise in the empty hallway. This hallway is the image of a lonely passageway to nowhere, and the noise of the slippers would be enough to drive one mad. Who is he speaking to? And why the order, or is this an order at all? Is he trying to convince himself of something or to motivate himself? Or is it the ghost of Pessoa – Campos' creator – who speaks? Or is Campos speaking to Pessoa? Or perhaps the order is to everyone he has ever met, and all the sensations that have emerged from these encounters, the ghosts from the past, or perhaps to us the readers. The starry firmament 'no longer lights the solitary wanderer's path,' (Lukács, 1971: 36) and a man in this new world is to be solitary here.

The full stop between lines five and six, between the two orders, reveals the pause, the breath, the silence, and the loneliness and solitude. There is an attempt to end the poem again, which could end here. But he continues, with the order to leave him alone in the room with the vast peace of himself ('Fique eu no quarto só com o grande sossego de mim mesmo'). In the imperative of the verb *ficar*, again we may ask to whom is he speaking? To which we, as readers, may answer with another question: 'Who am I reading?' The question – 'To whom am I speaking?' – has a profound significance for the poet. It is the reflection on the posterity of the word, writing for the dead and for those in the future with the abiding hope that the writing lives on with future readers. With Campos' declaration of isolation and the ambiguity of whom he is speaking to, the motif of writing for nobody which I will mention again in my conclusion comes to the fore and is linked to the self-appointed 'man in the garret'. This is the chosen isolation and Campos' repetitive call and cry for attention as the poet who cannot keep quiet or remain still. The 'grande sossego de mim mesmo' could be related to the sudden access he has been given or given himself to the knowledge

that he does not exist. The cosmos, which signified order, is now for the modern mind paradoxically full of chaos indicating the ruin of all things where these lines of Campos reside. And what 'room' is he in? The beauty of this line is its openness to prod the reader's imagination. Each line is an attempt to end and to start, like a warming up to the main part which never comes. Again he is finished, reminding me of Beckett's last words in his formidable prose trilogy: 'I can't go on, I'll go on' (Beckett 1994: 418). Beckett, Campos and Pessoa do go on, continuing again – as is the calling of the one from the *mansarda*, that one from the margins, like Nietzsche's posthumous writer[15] or Kierkegaard's *Extraskriver* (Kierkegaard 1983: 7); in sum, the one at the interval.

The last line makes another concise, complete declaration: 'É um universo barato [It's a shoddy universe].' This reveals Campos's pessimism and realism, but also the ruin and *chaosmos* of all things. In our modern world and life, anything can now be had, anything can be bought and sold, everything has a price, which devalues the whole universe. And, as Campos told himself and the world a few years previously with the publication of 'Tabacaria', he failed in everything.[16] And yet precisely in this state of failure and ruin, and even paralysis and petrification, in his chaos of feeling, the affirmation of art emerges in the creation of form in the attempt to express sensations. The sad, confessional poem under scrutiny here also reveals the continuity and repetition in the face of disaster and not even existing. And at the same time, finally, this is Campo's spiritual renewal, his survival and his conviction – which is never giving up on the creative force.

Conclusion: The Nomadic Life

In conclusion, Campos – as writer of the interval – is always repeating himself, refusing to settle, refusing to be tamed and to be pinned down, and always revealing the crisis of human communication – much as

---

15   Nietzsche (1976: 566): 'Some are born posthumously.'
16   Pessoa (2002: 320; 1998: 174): 'Falhei em tudo [I failed in everything].'

Kierkegaard and Nietzsche have done for philosophy. He is a nomadic
writer[17], floating between two worlds, exemplified in this poem: 'I'm
beginning to know myself. I don't exist.' It is no coincidence that Sar-
amago later makes Pessoa show up from the grave in his great novel *O
Ano da Morte de Ricardo Reis* to tell Ricardo Reis that he is 'floating,
in other words, in midocean, neither here nor there. Like the rest of the
Portuguese [*você anda a flutuar no meio do Atlântico, nem lá, nem cá,
Como todos os portugueses*]' (Saramago, 1999: 312; 1998: 353). Dur-
ing his own myth-making and propaganda writings for *Orpheu*, Pessoa
wrote a letter in English (which was never sent) describing the Portu-
guese as having an 'indefiniteness of soul' and a 'temperamental nonre-
gionalism' (Pessoa 1966: 143).

Campos' poem brings up the figure of the nomadic writer who
claims to be writing for nobody. This is something that Campos has in
common with a particular kind of modern European writer of the inter-
val that continues to unleash an uneasy feeling, who creates mask upon
mask displacing identities, and whose wanderings are unfolding rather
than progressing with purpose.[18] Nomadic writers of the interlude such
as Kierkegaard, Nietzsche, Benjamin, Kleist, Kafka and Beckett are all
attempting to write for that 'nobody' – as the invisible, disrupting other
who may not fit in with the narrative that history would like to write.
These are the most compelling and penetrating philosophers and poets

---

17   Zbigniew Kotowicz wrote a monograph on Pessoa presenting him as a nomadic
     soul. See Kotowicz (2008).
18   I think of Pessoa's writings as a method that Benjamin describes: 'Method is a
     digression. Representation as digression – such is the methodological nature of
     the treatise. The absence of an uninterrupted purposeful structure is its primary
     characteristic. Tirelessly the process of thinking makes new beginnings, returning
     in a roundabout way to its original object. This continual pausing for breath is the
     mode most proper to the process of contemplation. For by pursuing different levels
     of meaning in its examination of one single object it receives both the incentive to
     begin again and the justification for its irregular rhythm. Just as mosaics preserve
     their majesty despite their fragmentation into capricious particles, so philosophical
     contemplation is not lacking in momentum. Both are made up of the distinct and
     the disparate; and nothing could bear more powerful testimony to the transcendent
     force of the sacred image and the truth itself. The value of fragments of thought
     is all the greater the less direct their relationship to the underlying idea, and the
     brilliance of the representation depends as much on this value as the brilliance of
     the mosaic does on the quality of the glass paste' (Benjamin, 2003, 28).

of modernism. They are always at the interval between what has been lost and what may be discovered or rediscovered in a transformed way in the future. Thus, they are inviting the reader to spend more time at the frontiers, which, as the poet well knows, is the key to transformation. Deleuze and Guattari follow the nomad as the writer of the interval disrupting the *nomos* of appropriation, distribution and production and that which designates the law: 'A path is always between two points, but the in-between has taken on all the consistency and enjoys both an autonomy and a direction of its own. The life of the nomad is the intermezzo […] The nomad knows how to wait, he has infinite patience. Immobility and speed, catatonia and rush […]' (Deleuze and Guattari 1987: 380). Thus, Deleuze and Guattari state at the very beginning of their work *A Thousand Plateaus* that their 'rhizome' – which grounds their whole thesis of nomadology – is in the interval: 'A rhizome has no beginning or end; it is always in the middle, between things, interbeing, *intermezzo* […] Kleist, Lenz and Büchner have another way of traveling and moving: proceeding from the middle, through the middle, coming and going rather than starting and finishing. […] The middle is by no means an average; on the contrary, it is where things pick up speed' (Deleuze and Guattari 1987: 25).

I end here with a revealing explanation by Pessoa on what a poem is and on his own aesthetic and spiritual position and journey. The paragraph presents the written word as an activity and the creative mind that is always in flux, elusive and plural, flying by the nets of convention, authority and temporality. Written in English probably in 1916, this paragraph provides more clues to understanding the *chaosmos* at play and the ever-shifting dissolution and plurality of the self that is evident in the seven-line poem that I chose for this short essay:

> I sometimes hold that a poem […] is a person, a living human being, belongs in bodily presence and real fleshly existence to another world, into which our imagination throws him, his aspect to us, as we read *him* in this world, being no more than the imperfect shadow of that reality of beauty which is divine elsewhere. I hope some Day, after death, I shall meet in their real presences the few children of these I have as yet created and I hope I shall find them beautiful in their dewy immortality. You may perhaps wonder that one who declares himself a pagan should subscribe to these imaginations. I was a pagan, however, two paragraphs above. I am one no longer as I write this. At the end of this letter I hope to be already something else. I carry into practice as far as I can that spiritual disintegration I preach. If I am ever coherent, it is only as an incoherence from incoherence. […] (Pessoa 1966: 133)

References

Badiou, Alain (2005). *Handbook of Inaesthetics*, translated by Alberto Toscano. Stanford: Stanford University Press.

Beckett, Samuel (1994). *Molly, Malone Dies, The Unnamable*. London: Calder Publications.

Benjamin, Walter (2003). *The Origin of German Tragic Drama*, translated by John Osbourne, London: Verso.

The Bible (1997). *Authorised King James Version with Apocrypha*. Oxford: Oxford World's Classics.

Deleuze, Gilles, and Félix Guattari (1987). *A Thousand Plateaus: Capitalism and Schizophrenia*, translated by Brian Massumi, Minneapolis: University of Minnesota Press.

Joyce, James (1992). *Finnegans Wake*. London: Penguin Books.

_____ (2008). *Ulysses*. Oxford: Oxford University Press.

Heidegger, Martin (1963). *Sein und Zeit*. Tübingen: Max Niemeyer Verlag.

Kavanagh, Patrick (2005). *Collected Poems*. London: Penguin Classics.

Kotowicz, Zbigniew (2008). *Fernando Pessoa: Voices of a Nomadic Soul*. Exeter: Sherman Books.

Kuberski, Philip (1994). *Chaosmos: Literature, Science and Theory*. New York: State University of New York Press.

Kierkegaard, Søren (1980). *The Concept of Anxiety*. Princeton: Princeton University Press.

_____ (1983). *Fear and Trembling*, edited and translated by Howard V. Hong and Edna H. Hong. Princeton: Princeton University.

_____ (1993). *Upbuilding Discourses in Various Spirits*, edited and translated by Howard V. Hong and Edna H. Hong. Princeton: Princeton University Press.

Lukács, Georg, *The Theory of the Novel* (1971). Translated by Anna Bostock. Cambridge Massachusetts: The MIT Press.

Nietzsche, Friedrich (1954). *The Antichrist*, in *The Portable Nietzsche*, edited and translated by Walter Kaufmann, London: Penguin Books.

Pessoa, Fernando (1966). *Páginas Íntimas e de Auto-Interpretação*, edição Jacinto de Prado Coelho e Georg Rudolf Lind. Lisboa: Ática.

____ (1994). *Poemas Inglesas*, 4.ª edição. Lisboa: Edições Ática.

____ (1998). *Fernando Pessoa & Co.: Selected Poems*, edited and translated by Richard Zenith. New York: Grove Press.

____ (2000). *Crítica: Ensaios, Artigos e Entrevistas*, edição Fernando Cabral Martins. Lisboa: Assírio & Alvim.

____ (2002). *Poesia de Álvaro de Campos*, edição Teresa Rita Lopes. Lisboa: Assírio & Alvim.

____ (2005). *The Education of the Stoic*, edited and translated by Richard Zenith, Cambridge: Exact Change.

____ (2006a). *Poesia 1931–1935*, edição Manuela Parreira da Silva, Ana Maria Freitas e Madalena Dine. Lisboa: Assírio & Alvim.

____ (2006b). *Fernando Pessoa. A Little Larger than the Entire Universe: Selected Poems*, edited and translated by Richard Zenith. London: Penguin Books.

____ (2012). *Livro do Desassossego*, edição Richard Zenith (10ª edição), Lisboa: Assírio & Alvim.

____ (2015). *The Book of Disquiet*, edited and translated by Richard Zenith, London: Penguin Classics.

____ (2016). *Fernando Pessoa: English Poetry*, selected and introduced by Richard Zenith. Lisboa: Assírio & Alvim.

Sá-Carneiro, Mário de (1978). *Obras Completas de Mário de Sá-Carneiro II: Poesias*, edição João Gaspar Simões. Lisboa: Edições Ática.

Saramago, José (1998), *O Ano da Morte de Ricardo Reis*. Lisboa: Editorial Caminho.

____ (1999). *The Year of the Death of Ricardo Reis*, translated by Giovanni Pontiero. London: The Harvill Press.

Whitman, Walt (2004). *The Complete Poems*. London: Penguin Books.

Rational Landscape: Spatial justice, politics and aesthetics: The city of Lisbon as a case study

Diana Soeiro

In recent years, a new paradigm has emerged to counterpoise the old paradigm where the social and the historical are privileged. In this new paradigm social, historical and spatial perspectives are in balance, 'with no one of the three ways of looking at and interpreting the world inherently privileged over the others' (Soja 2010: 3). This movement, which is still in its formative years, we call *spatial turn* (Schlitte et al. 2014) and this chapter adopts its methodology which we will now briefly describe.

If in the last century historical and social aspects were highlighted, it is now claimed that spatial perspective has to come into play. This spatial turn was brought forward by geographers and quickly embraced by several other research fields with an impact on human sciences (Soja 2010: 3). In the words of American geographer Edward W. Soja (1940–2015), the shift is that instead of giving 'greater stress to how social processes such as class formation, social stratification, or racist or masculinist practices shape geographies', we now focus on 'how geographies actively affect these social processes and forms' (Soja 2010: 4). This means going beyond a conservative understanding of space where it is seen as a receptacle, where things happen to it and in it.

Adopting a critical spatial perspective, we will dwell on the relation between philosophy and urban planning, taking both a theoretical and practical approach, strengthening the symmetry between social and spatial explanation. As Ferrão reminds us (2012: 67), in urban planning, the academic domain (basic training) and professional domain (spatial planning practice) do not coincide, requiring the contribution of several other disciplines. The city of Lisbon (Portugal) will be used as a case study relying mainly on the latest report issued by the Lisbon City Council (CML) this year. This chapter therefore gives a description of the method, which highlights spatial dynamics, acknowledging it as an active element.

By putting forward the explanatory power of spatial thinking, we can better assess both affectivity and the impact of decision-making on spatial dynamics. History and sociology reflect on space and spatial dynamics as a consequence of what happens historically and socially, making space a secondary element, an aftermath. A critical spatial perspective understands space as an element that can be the cause of social and historical transformations. Space is acknowledged as an active component and not exclusively as passive (where social and historical transformations are simply reflected).

Adopting a negative description, says Soja, space is actively involved in sustaining 'inequality, injustice, economic exploitation, racism, sexism, and other forms of oppression and discrimination.' (Soja 2010: 4) The positive formulation of this statement is that the way we shape space can play an active role to help us achieve a more just society. In order to help us achieve that, urban planning and the highest principles by which the city is sustained play a leading role. The exercise of identifying these, which includes detecting specific elements that lead to specific urban planning measures and their implementation, is what Soja called spatial justice.

1. From politics and aesthetics to spatial justice

When it comes to spatial planning, if we adopt a critical spatial perspective, this enables us to throw a new light on the relation between two traditional areas – politics and aesthetics. The relation between the two is known to be strong with politics frequently making use of aesthetics in order to affirm and sustain power (Adorno et al. 2006; Rancière 2013; Benjamin 2002[1]). In order to move on from this dialogue between politics and aesthetics (whether claiming that the relation is beneficial,

---

1   According to Walter Benjamin (1892–1940), the first to significantly reflect on the relation between politics and aesthetics, right-wing fascism practised the aestheticizing of politics while communism politicized art (Benjamin 2002: 122). On this topic, see also Hillach et al. (1979).

dreadful or inevitable, or claiming that they both should be independent and follow separate ways), spatial justice is a useful approach. Through the methodology of spatial justice, justice is highlighted as the main foundation on which the city stands, allowing us to find a new way of rationalizing the urban landscape. To the power of politics we counterpoise the power of place (see De Blij 2010).

Cities are growing at a very fast rate, rapidly increasing their complexity and therefore demanding not only material but also contemplative resources in order to better help us plan the future.

In 2014, for the first time more than half of the world's population became urban. It is expected that in 2050, 66 per cent of the world population will live in urban areas. In developed regions the number will reach 85 per cent. In Europe, Latin America and the Caribbean, North America and Oceania the growth is expected to be minimal since a large number of people already live in urban areas – in 2050, many countries are expected to reach almost 100 percent. The fastest growth rate will take place in Africa (rising from 40 per cent in 2014 to 56 per cent in 2050) and Asia (from 48 to 64 per cent). In 2010–2015, the highest average growth rates in Africa were in Burundi, Ethiopia, Rwanda, Uganda, Tanzania, Angola, Lesotho, Namibia, Burkina Faso, Mali and Nigeria. In Asia: in China, Bangladesh, Bhutan, Maldives, Nepal, Lao People's Democratic Republic, Thailand, East Timor and Vietnam.

By 2050, most countries in Africa will have reached between 60 and 70 per cent living in urban areas, with Ethiopia (38 %), Kenya (44 %), Madagascar (30 %) Angola (64 %), Congo (60 %), Morocco (74 %), South Africa (77 %) and Nigeria (67 %) being the countries most responsible for this due to a combination of the size of their population and the growth rate. In Asia, already 100 per cent of the population of Hong Kong and Macau live in urban areas and the outlook for Japan is that by 2050 the figure will reach 98 per cent. The countries that will have a bigger impact between 2014 and 2050 by creating a prevailing urban population as a result of their large population and the growth rate are Southern Asia (from 34% to 52 %), India (from 32 % to 50 %), Indonesia (from 53 % to 71 %) and China (from 54 % to 76 %) (United Nations 2014, 20–25). In Portugal, the urban population was 48 per cent in 1990, 63 per cent in 2014 and it is expected to rise to 77 per cent in 2050. In absolute numbers, this means that around 7 564 million people

will live in an urban environment and 2 279 million in a rural environment (United Nations 2014: 23).

Looking at these numbers, we quote Lewis Mumford, who back in 1961 sagely predicted: 'Will the city disappear or will the whole planet turn into a vast urban hive? – which would be another mode of disappearance.' (1966: 3) The question is: if we are on the verge of globally becoming fully urban, what sense can we make of 'the city' as we know it?

The prospect of our planet becoming a global village is not a futurist scenario anymore and the likelihood of the whole world becoming 'a city', literally, is closer than we perhaps anticipated. Assuming that sustainability is always a relevant criteria to apply to existing cities, as it is a concept that embodies the effort to continue to promote growth while assuring the population a quality living standard that allows them to prosper, the criteria truly becomes indispensible in new cities, particularly in those countries which are experiencing an accelerated growth rate.

But what does a country's population becoming increasingly urban means to a country? Construction is one of the most significant actions to increase a country's GDP. This means that if on one hand the creation of cities and larger cities that progressively attract more people reflects positively and immediately on a country's economy, on the other hand, they may not be assuring sustainability because they are being created rapidly, which reflects negatively on the population in the short term and on the economy in the mid and long term. We highlight three main reasons that contribute t a lack of sustainability in the urbanization process: (1) economically and politically, urban rehabilitation proves to be less profitable in the short term than new construction, which may hinder access to investment; (2) countries where urban areas are already extensive will have to have more robust economic strategies in order to be sustainable since they are not able to rely on new construction to increase their GDP; (3) countries that offer the possibility for new construction due to easy access to investment, which immediately benefits the country's GDP but does not assure sustainability, can experience faster growth than is desirable. Though sustainability has become an over-used

word that may sometimes have a blurred meaning, a significant criterion to assess what is urban sustainability is spatial justice.

We present here a brief description of what spatial justice may refer to: 'On the level of culture and society, there are four broad categories of spatial strategies of power: (1) the construction of hierarchies, (2) segregation, (3) marginalization, and (4) long-term, large-scale mechanisms of spatial transformation like apartheid, colonialism and globalization. Each of these has a particular paradigm of operation, and each impacts at various scales of physical space. From the scale of the body, up through the scale of buildings and cities to the scale of the landscape, power exercises explicit and implicit control over the shaping and occupation of space' (Findley 2005: 7). In developing countries these strategies are more visible but this does not mean that they do not exist in developed countries. They do exist but are usually more subtle.

Considering what has been said perceptively, we observe that all over the world cities are being created, renovated and expanded. It is not our goal to discuss here how aesthetically appealing the options being taken in each case are because that would be to express an understanding of aesthetics in its most superficial sense. In adopting a higher sense of what is at stake in the aesthetic experience, spatial justice can be useful and it does concern aesthetics because the fact is that, perceptively, cities (taken as an object able to be conceived, moulded and experienced by the human mind) are appearing at a fast pace before our eyes, changing the way we move around space, constricting the way we move our body, the way we relate with our family, with the work place, the way we eat and our access to food, and so on. It is not only the case that cities look different, we experience them differently. As much as we shape them, they shape us. Space is a topic that simultaneously concerns aesthetics and ontology and the two are truly inseparable when we discuss spatial planning.

As we have seen, an increase in built area benefits a country's economy, which at first sight may be considered as something having a positive political impact. But only if we reflect on how, why and under what principles these cities are being built, will we be able to then truly assess if their political and economic impact is positive.

Being the state of affairs the one we have described above, adopting a critical spatial perspective and acknowledging the methodology

of spatial justice as an effective approach strategy, we ask: how can we better rationalize landscape? We will take Lisbon as a case study.

2. Spatial strategies of power: on borders

We tend to dislike the word 'power' – though some tend to like it precisely for the exact same characteristics we will describe below. Behind it seem to lie obscure secrets, blackmail, bribery, deceit, betrayal and corruption. But power is not always a bad thing. Someone will always have power and, as Plato said in the *Republic*, each one of us exerts power in a specific realm[2] – whether we like it or not, whether we do it well or badly.

Power exists and the question is how to deal with it and manage it better in order for it to be as beneficial as possible for the majority of people. Spatially, one of the main elements that allows us to define power strategies are borders. From the Great Wall of China (206 BC) to the current European refugee crisis (2015) and Donald Trump's wish to build a wall across the Mexican border (2016), there are many examples of how borders are one of the most basic elements used to sustain power. It is a way of preserving territory.

As Paquot tell us: 'Animals' territory does not correspond to a perfectly delimited portion, protected and protective. It is mobile, elastic in its outline, variable according to 'seasons', hours, activities and dangers. … It is mostly at the moment of reproduction that animals delimit their territory.' (Paquot 2009: 14, 15) But among men, territory easily becomes about exact delimitation. Sometimes this delimitation is spatially visible; sometimes the delimitation is artificially created on paper with no sense of discontinuity existing in the landscape. Either way, delimitation is always about distance. In Deleuze's words: 'The territory is

---

2 That is why he then describes how can we exert power in the best way possible, which implies understanding what are the characterises of a virtuous city, among them justice which in its turn can only be found in the city if, and only if, it can be found in each of its citizens. (Plato 1930, 1935)

first of all the critical distance between two beings of the same species: Mark your distance. What is mine is first of all my distance; I possess only distances.' (Deleuze & Guattari 1987: 319)

In 2012, a new administrative reform of Lisbon's spatial structure policy was implemented through Law 56/2012 (November 8) setting out a new city map. The Mayor of Lisbon at the time, António Costa[3], implemented the new plan which had started being drafted in 2008 by Augusto Mateus and João Seixas. (CML 2012) The new map consisted of establishing new boundaries and merging several parishes. Instead of the 53 that had existed since 1959, there were now 24. According to the new administrative plan, as Lisbon is a coastal city with a river waterfront to the south, the delimitation of parishes does not coincide with the land margin itself but instead with one of the river's thalwegs (one of the deepest parts of a river), thereby adding 15.7 km2 to the area of Lisbon. Out of the 24 parishes, ten are located on the riverfront (Pais 2016: 6, 7, 10).

This new administrative reform means that when the time for local council elections comes (every four years), a team is elected to represent each parish. This may seem an obvious statement but it is particularly relevant to mention it because, parallel to the plan that established new boundaries for the parishes, a second map with different boundaries called the Operational Units for Planning and Management (UOPG) co-exists. The UOPG divides Lisbon into nine administrative areas and it is part of the 2012 Municipal Director Plan (PDM) that established the boundaries. Each area is managed according to a specific program defined and regulated by the PDM which translates into specific execution and financing strategies (Pais 2016: 12).

This means that two administrative reforms took place in parallel and that both established different boundaries on paper. One is legally binding when it comes to electing representatives, and has some autonomy (24 parishes), while the other is legally binding when it comes to bigger management options and distributing financing (nine operational units). There are units which include two or more parishes and there are units that cut through parishes, thus making one parish belong to two different units. It is clearly assumed in the latest report on Lisbon

---

3   Currently Prime Minister of Portugal, in office since 26 November 2015.

that these two systems co-exist and that their boundaries do not coincide (Pais 2016: 13).

In an effort to get closer to the population by practising 'good governance' (Pais 2016: 14), Lisbon City Council decided to create five management areas called Territorial Intervention Units (UIT), thereby establishing a third set of boundaries in Lisbon. Each UIT is managed by a multidisciplinary team that helps identify and solve problems related to public space and public equipment, thus creating an intermediate structure between parishes and City Council. In the case of UITs, which were established in 2011 and have been officially working since 2015, the boundaries coincide with the 2012 parish boundaries (and not with the PDM's boundaries). Each UIT includes between three and six parishes (Pais 2016: 14).

When it comes to mainland Portugal, there are a total of 18 districts, among them Lisbon. Setúbal, however, is also an important district. When we look at the boundaries of the Lisbon Metropolitan Area (AML), which includes an area both north and south of the river, the north area lies within the district of Lisbon but the area south of the river is part of the district of Setúbal (according to the country's districts map). As a result, this means that there is a fourth set of boundaries that comes into play when we consider Lisbon's spatial planning.

The co-existence of these four sets of boundaries brings an additional challenge to synchronized action among all parishes and between the districts of Lisbon and Setúbal, disempowering the spatial dynamics and consequently their representatives and citizens. Why and how this power structure came to be is not relevant at the moment, but the fact is that two different systems co-existed, a third was created in order to build a bridge between parishes and the City (which may perhaps have been an attempt to solve the conflict between the two previously existing ones which had different boundaries), and a national spatial planning framework established a fourth one. This leads to excessive bureaucracy, population disempowerment and bad governance since it becomes hard to understand how the system works (and therefore how to access and manage it), making it harder to understand how a specific budget is managed, who manages it, and consequently to identify whose responsibility it is to assure specific structures or services.

Boundaries can have a positive effect, bringing about positive social outcomes, but to simultaneously use four different sets of boundaries seems excessive, confusing and detrimental to promoting a simple, intelligible, straightforward system that creates a fluid and easily adaptable spatial dynamic. The excessive superimposed sets of delimitation are administratively too heavy and they promote lack of flexibility and mobility to the city's inhabitants – the lack of an elastic outline that Paquot referred that is natural to animals. This spatial dynamics makes Lisbon, territorially, vulnerable to have social and economic fissures. The current system does not promote social justice, neither for the citizens nor for their representatives.

3. Design and decision support systems in architecture and urban planning

An understanding of what governance actually means is sometimes confusing. But whenever the word is used it is meant to imply a close connection with democracy, also referring to the wide participation of a majority (Fung & Wright 2003). When applied in the context of urban planning, governance therefore aims to stress democracy's essence by highlighting the relevance of participation in the spatial dynamics so that space can be a true democratic experience, i.e. a shared exercise of power. With no community involvement there is no participation and consequently no governance at play (Haus et al. 2004).

In Lisbon, participation in public discussion on urban-related issues (Territorial Management Instruments (IGT)) is still very low (Pais 2016: 362). Since 2008, Lisbon City Council has implemented a Participatory Budget, an instrument where part of the budget available is spent on projects that citizens themselves propose and, subsequently, vote for. The ones with more votes are the ones going forward. For the last five years now, we can observe that the number of proposals has been slowly decreasing though we can see that the number of voters has been increasing. In 2015, the Participatory Budget had around 36 000

voters (Pais 2016: 364, 365). As Lisbon has around 500 000 inhabitants, we can conclude that only around 7.2 per cent of the population votes.

This number is not very expressive though it can be argued that the instrument is recent. However, since both proposal submission and the voting process take place online, the population that is being targeted has to be IT savvy and have access to a computer and the internet. Furthermore, the way in which the Participatory Budget is publicized (mostly online and on very specific websites) only reaches a small number of people and, above all, a very specific demographic.

Even excluding the obligatory computer and internet use of the Participatory Budget, if we look at the participation in public meetings, we conclude that there is some alienation on the population's behalf when it comes to actively participating in the decision-making process as a result of low attendance. Participation though increases as we move away from the historic centre towards the periphery (Pais 2016: 368). The population and the elderly follow a similar pattern, with the centre having a lower population but a high elderly rate. We can therefore conclude that those over 65 years old, although they represent around 23.9 per cent of the population, have little or no participation (Pais 2016: 25). Since 2009, the majority of proposals have shown a greater concern and need for intervention in public and green spaces followed by mobility issues (Pais 2016: 364).

When it comes to explain low citizen participation in urban planning, we think that more than forty years of dictatorship (1926–1974) still weighs on free speech and publicly taking a stand. In addition, in an effort to make a positive contribution to help the population heal this cultural scar, community and proximity structures should be encouraged to play a more active role, thereby allowing more segments of the population to be represented in the participation process.

4. On scale

In 2011, around 2.8 million people lived in the Lisbon Metropolitan Area (AML) out of which around 500 000 in the city of Lisbon (Pais 2016: 17). Portugal has a total area of 92 200 km2, the AML has around

3 000 km2 and Lisbon 84.97 km2 (OECD 2011: 23; AML 2016). This means that out of the overall urban population in Portugal, around 37 per cent live in the Lisbon Metropolitan Area (which represents 3.3 percent of the country's territory), out of which 6.6 per cent in Lisbon (representing 0.09 percent of the country's territory).

If we consider that mainland Portugal has a total of 18 districts, each with their respective district capital, we have around 63 per cent of the urban population living in 17 districts. (Let us not forget that when referring to the Lisbon district we are only referring to the area north of the river.) Portugal is therefore, a macrocephalic country, having one big city, densely populated when compared to any other city of the rest of the country, where a large majority of the population is concentrated. As in many other countries, the primacy of capital cities is not perceived as a positive phenomenon. A macrocephalic growth of capital cities indicates 'a high concentration of elites and a monopoly of the institutions of modernization' (Aveline-Dubach et al. 2014: 28).

Such a high concentration of population in such a small area also exposes it to higher risks when it comes to natural hazards, which are unpredictable events. In Portugal it is well known that there are several seismic flaws. Back in 1755 the country suffered a violent earthquake (of an estimated 8.5–9.5 magnitude on the Richter scale) and ever since then we believe we are particularly vulnerable to a similar kind of event. If that were to happen, there is a 39 per cent likelihood of Lisbon crumbling, with all its parishes affected. Knowing that seismic prediction can be a difficult science (Hough 2009), it is positive to know that a project (Geo SIG) is underway to produce a better cartographic map of Lisbon. By looking at a recent map that assesses seismic risk, we can observe that the river coastal area, the historic centre and the eastern part of the city are the most susceptible areas (Pais 2016: 92–95). It is also known that 85 per cent of Lisbon is extremely vulnerable to earthquakes, thereby exposing 68 per cent of the population to high risk and potentially, almost certainly, destroying 57 per cent of existing buildings.

The situation sounds dramatic and it is even more so considering that the biggest concentration of empty houses is precisely in the historic centre (over 35 per cent), then in the area surrounding it with between 15 and 35 per cent, and then in the outer area of the city with less than 15 per cent (Pais 2016: 116). On one hand, we can assess this as positive

since fewer people would be hurt but the historic centre is also where the majority of people work (Pais 2016: 109, 265).

Another area where a massive macrocephalic imbalance occurs is in culture. Most museums and cultural activities take place in Lisbon. Yet when it comes to art galleries and temporary exhibition spaces, there is a tendency for these spaces to decrease in Lisbon but to increase in the rest of the country (Pais 2016, 177–180). This is perhaps one of the few areas that demonstrates a resistance to the excessive concentration of resources and activities in Lisbon.

From an Economy perspective, which also confirms the country's macrocephaly, 37 per cent of GDP in 2014 was generated in the AML, which hosts 29 per cent of all jobs (Pais 2016: 254, 256).

The primacy of capital cities should be discouraged at a national policy level and Lisbon should invest in prospering instead of growing (Jackson 2011).

5. Urban planning and cultural identity

When it comes to the existence of green spaces in Lisbon, it is hard to understand how their area is calculated although a lot of them are iden- tified as having a quite extensive 'area of influence' (Pais 2016: 40–46). There is a general feeling of the lack of green spaces in the proximity and these, as we have seen, are among the main concerns when it comes to citizen participation. Green areas are essential to ensure a high qual- ity living standard as they have a positive effect on stress, mental and physical health, and air quality (Gilbert 2016).

Different types of green areas are urban farms, which have only recently started to be promoted around the city by the City Council. Just like the chapter on investment in solar power, whose high potential is acknowledged (Pais 2016: 81), the chapter on urban farms is still too short (Pais 2016: 48–51). What is more, the use of renewable energies is a key element to promote a sustainable city, especially in a city where almost 50 per cent of energy certificates are C or G in services and around 55 per cent of dwellings are C and D (Pais 2016: 83, 84). There

is a still long way to go in planning more green spaces and creating a culture that envisions its future as mostly relying on renewable energy sources.

Culturally, although the risk of an earthquake (and tsunami) is high, Lisbon is characterized by old buildings with more than 20 per cent built before 1919, 19 per cent between 1919 and 1945, 25 per cent between 1946 and 1960, 21 per cent between 1961 and 1980 and 15 per cent between 1981 and 2011 (Pais 2016: 106, 107). Most new buildings have been built in the north and northeast of the city. The centre and the historic centre contain 63 per cent of buildings considered to be structurally 'bad' or 'very bad' (Pais 2016: 109). Curiously enough, these are also the areas where most businesses are located (60 per cent) and where most people in Lisbon work (65 per cent) (Pais 2016: 265). Even co-working spaces and start-ups are located in these areas (Pais 2016: 274, 275).

Speaking of businesses, in 2015 the number of companies formed exceeded the number of bankruptcies for the first time since 2008 (Pais 2016: 274). However, this number should not cause immediate optimism since less than one in three business start-ups survive for five years, half may last two to three years and 20 per cent close within a year (Stokes & Wilson 2010: 86). Still, it is noticeable that there is a will to create new business. Is that will being supported and encouraged by national and city-level measures? Healthy, small and medium sized business are vital for a sustainable economy and a carefully strategy should be in place to support these activities, short, medium and long-term. Only then, this recent entrepreneurship impulse will fructify.

Family wise, following the impact of the 2008 crisis, in 2015 each woman living in the AML had 1.56 children and Lisbon had the highest rate in the AML with 2.03 children (Pordata 2016). It is therefore significant if we add the information that the average number of residents per house in 2011 was 2.3 (Pais 2016: 121). This means that either there are a lot of single parents or there are many couples with no children. It is common knowledge that the government wants to change this situation by encouraging families to have more children (Faria 2014) as traditionally people in Portugal used to have large families. But what does that mean when it comes to house typologies? If families have more children, this means that they need to have more space. Is Lisbon's real

estate market ready to offer more square metres and, more importantly, can families afford it? If not, that means that more people will move from the city centre to the outer crown of the city, or even perhaps to the AML or a different district (which we can observe has been a growing tendency in the last few decades). This brings us to the topic of mobility and transports. If people move away from the city centre: 1) do they, necessarily, have to get into the city daily?; 2) if so, how will they get to the city?

Since 1991 the use of individual transportation (car) has tripled, parking spaces are rarely sufficient and there is a great discrepancy between night and daytime use of parking spaces (lack of spaces during the day and a large majority empty at night). Most buildings inside the city do not have private parking and so those that do are usually prized (causing a great impact on rents or sale prices). Consequently, it comes as no surprise that collective transports (metro, bus and boat) have experienced a decrease in the number of passengers (Pais 2016: 204, 220, 221–224). An alternative option was created recently and, since 2008, there has been a 500 per cent increase in cyclable areas. To date, 79 km are already open and there is a total of 148 km of cyclable routes (Pais 2016: 225).

The culture of car use seems to be on the rise and despite measures being taken to discourage it, these do not seem to be effective. Bike lanes are a positive investment but because the city has many hills it is not an option that is possible for everybody, particularly for the elderly. However, car culture is not fully responsible for the rise of individual transportation. While urban planning discourages car use, reducing and limiting circulation space in particular parishes, it is known that when compared with 2010, buses (Carris), metro (Metropolitano de Lisboa) and boat (Grupo Transtejo) have reduced their offer in 25%-30%. Since 2010, Carris lost 26.5% of its workers and Metropolitano de Lisboa, 18.2%. (Saraiva 2016) The metro system has little or no maintenance and the material is either in bad shape or in need of replacement. Most trains are from the 1990s and are in need of a deep revision so they continue to run in the next twenty years. Still, due to budget restrictions imposed both by the Ministry of Environment and Ministry of Finances that is due to take place. Maintenance engineers, conductors and administration are at a breaking point and the situation leads, inevitably,

to a bad service. (Cipriano 2016) As for Carris, the current City Council President, Fernando Medina, has recently admitted that he expects that a solution will be effective in ten years, due to the current chaotic state of the company. There is a strategy being outlined, that aims to integrate all available means of transportation in order to bring back mobility to Lisbon, and part of the solution encompasses that the Carris management starts being done by the City Hall itself. In Medina's own words, both Carris and Metropolitano de Lisboa do not currently provide a public service due to its quality standards. (Boaventura 2016)

We now understand better why car use has increased but cannot help wondering how the disappearing space for car circulation, being currently implemented around city, will articulate with a bad transportation service in the next ten years (assuming that that will be the gap until the situation is solved). It is expected that the city will experience a low mobility capacity during these years, with heavy traffic associated to a poor quality public transportation service, unless either a wide mass of people start using the bicycle lanes (which due to the city's topography is not for everybody) or move away from the city centre (which is contradictory with the plan that has among its goals to be able to attract more people to live in Lisbon). A large part of the population uses the car to move around the city due to poor public transportation service but also those who come to work in Lisbon, living in the outskirts, are a vast majority because Lisbon's centre is macrocephalic when compared to AML or its district. It would be beneficial if, also in this sense, the macrocephaly would be smoothed allowing to revitalize not only the city centre but also its outskirts. If livelihood in the outskirts was more sustainable, citizens would not need to go to the city centre and many would choose to move from the city centre to the outskirts.

Tourism seems to be an activity that many identify with Lisbon's future due to the city's inviting weather, varied landscape and friendly people. The city was awarded Best Tourist Destination for Cruises by the World Travel Awards (2014), 2nd Best European Destination 2013 and 9th Best World City for Company Events (ICCA 2010) (Pais 2016: 298). Lisbon offers mostly 4-star hotels clustered in the centre (Pais 2016: 302) but alternatives like Airbnb, which have become increasingly popular, are not mentioned in Lisbon's latest 2015 report. Since 2013 there has been an increase of 53 per cent in Lisbon's hosting capacity

(beds) (Pais 2016: 298). This massive commitment to tourism is a sharp reaction to the 2008 financial crisis by allowing many businesses to access quick money in order to survive. How sustainable is it? Again, we wonder. How many hotels does a city need? Truly that is a question that requires an urgent answer. We may end up with many hotels on our hands, particularly in the centre and historic centre which is the area the Lisbon City Council considers to be the most promising to revitalize by increasing the population living there (Pais 2016: 22). But if hotels continue to be built, where will people live? Is the goal to attract temporary inhabitants or permanent residents to these areas?

A key element in traditional construction and a significant part of our culture are *azulejo* tiles, which are found on many buildings around Lisbon. In order to preserve and identify them properly, a program called the Lisbon Program for Tile Research and Conservation (PISAL) has recently been created (Pais 2016: 190, 191). The value of this building material is not exclusively aesthetic. It is an object that reminds us of the Arab presence and influence in Lisbon and lacking a very important function in cooling buildings by protecting them from excessive heat. There should be incentives given to use this traditional element more often in architecture instead of making it just a vague memory, shown exclusively in museums instead of on the streets.[4]

6. On inequality

Out of all 18 municipalities that are part of the AML, Lisbon is the municipality with the highest Ageing Index, with 185.8 elderly people for every 100 youngsters (under age 14). From 1990 a tendency can be seen for the number of adults and youngsters to decrease while the number of elderly people (above age 65) remains steady (Pais 2016: 25). This

---

4   Portuguese architect Álvaro Siza Veira (*b.*1933), Pritzker Prize honouree in 1992, used tiles in the metro station in the heart of the historic centre, Baixa/Chiado, in 1998 (featuring art work by Ângelo de Sousa (1938–2011)) and also in Terraços de Bragança, also located in Lisbon's historic centre in Rua do Alecrim (2004).

makes us predict that infrastructures for the elderly are urgent since the population is getting older. Moreover, elderly tourists have been identified as one of the segments who are most attracted to Lisbon (Pais 2016: 298). How is this segment of the population being integrated in the city's planning? How can the city be experienced by the elderly?

It is therefore puzzling when we look at the social equipments that were being proposed for the city in 2015. We can see that many daycare centres are planned, with land already assigned, but only four assisted residences in the centre and historic centre and two more in the outer crown (Pais 2016: 169). Let us remember the following: most house typologies in Lisbon have two to four rooms. As a result of high rents or the desire to have more children (which the government encourages), people move out of the city. Yet the number of daycare centres in Lisbon will drastically increase. Why? It is known that mainly in the outer crown of the city a large number of children attending daycare centres do not live in Lisbon (Pais 2016: 151, 152). We can then understand that what happens is that people use the car to go into Lisbon, drop off their children at the daycare centre in one of the outer parishes and then go to work in the centre. Yet overall, when we look more closely at the capacity of daycare centres, homes, home support services, residences and meeting centres, it is underwhelming, with some parishes having close to zero capacity (Pais 2016: 174). This is an area where much needs to be done.

Education is also an area that shows inequalities. Portugal's illiteracy rate now stands at 5.2 per cent and in Lisbon it is 3.2 per cent. Although there is not much difference between these figures, when it comes to publishing, 55 per cent of books published are done so outside Lisbon but in terms of reading habits, 70 per cent of newspapers and magazines are sold in the AML (out of which 48 per cent in Lisbon) (Pais 2016: 181, 182).

In the AML, out of its 18 municipalities only three have a population with a college degree, between 20 and 27 per cent. And as for number of years of schooling, Lisbon has 42 per cent with 9 years schooling, 32 per cent with a college degree, 16 per cent with 12 years of schooling and 7 per cent with no schooling at all (Pais 2016: 33).

Numbers of Erasmus students, particularly from Italy, Spain and Germany, are on the rise but the majority are from Portuguese-speaking

African countries (PALOPs). Foreign students comprise 10 per cent of the total of students (around 13 000) (Pais 2016: 283–286) and numbers seem to be on the rise showing a growth potential that can be relevant to embrace in order to promote a sustainable city growth.

Education is therefore an area that offers a huge potential for the Portuguese, with room to create a more appealing offer to those who wish to attain twelve years of schooling. Also, in universities, there is a move to attract both Portuguese and international students that should be utilised. Hopefully, this double call will not lead to changes that somehow increase inequality between national and foreign students.

Acoustics can also be an element of inequality. For the first time in 2012, Lisbon created a Noise Map that aimed to implement legal measures to reduce noise that had been compulsory since 2007. If we look at a General Noise Map of the city made in 2010, we can easily conclude that more than half the population is exposed to levels higher than 60 Lden (Pais 2016: 57, 58). In order to tackle this problem, a strategy is being implemented between 2014 and 2029, representing an investment of 9 million euros, to replace the city's road pavements in order to provide better noise absorption, to lower the speed limit in some areas and to erect acoustic barriers (Pais 2016: 61). Even so, when it comes to areas with an active nightlife or which are frequently used for temporary events that emit loud noise, although regulations are being slowly introduced, better fiscalization methods need to be implemented. Local residents should be entitled to their rest and a city that is 'alive' does not always imply being a 'loud city'.

Having considered acoustics and the impact it can have on the population, we may wonder why if a specific population is bothered by excessive noise in their area, they do not move. The fact is that in Lisbon ownership accounts for the main occupation of houses (50.9 per cent), which makes moving a less viable option. Renting is an option for 42.3 per cent and only 0.8 per cent are in a co-operative or collective property regime (Pais 2016, 119). Though renting is a more flexible system when it comes to facilitating relocation, high rents, associated to low salaries and a precarious job market contribute to a lower flexibility than expected. The average salary is 1383 euros, 300 euros more than in the rest of the country, the unemployment rate in 2014 at 14.9 per cent, associated to a house market highly subjected to speculation,

confirm this (Pais 2016: 263, 264). Taking into account the real estate market and salaries, co-ops and collective property regimes should be encouraged.

As for access to healthcare in Lisbon, at best one could say that it is undergoing a transition phase. We can see that several proximity structures are being planned, but five central hospitals are closing and the situation is confusing. What we know, from observing the current map of health facilities we can conclude that the centre and historic centre are currently deprived of proximity health structures (Pais 2016: 142).

Conclusion

Spatial justice, proposed by Soja, is an extension of the concept of 'right to the city' first proposed by Henri Lefebvre (in 1968) and defended nowadays by one of the most renowned geographers, David Harvey (Harvey 2013). As Harvey says: 'The right to the city is far more than the individual liberty to access urban resources: it is a right to change ourselves by changing the city. It is, moreover, a common rather than an individual right since this transformation inevitably depends upon the exercise of a collective power to reshape the processes of urbanization. The freedom to make and remake our cities and ourselves is, I want to argue, one of the most precious yet most neglected of our human rights' (Harvey 2008: 23).

Here, space as an indispensible relation between aesthetics and ontology is clear. The city, just like Plato said, shapes us but we are the ones shaping it. It is also relevant at this point to remember Husserl and the phenomenological school for whom there only exists lived space and never 'pure' space in the sense that 'we are space'. In that sense, 'right to the city' and 'spatial justice' connect, philosophically, with a phenomenological approach.

We have presented a detailed analysis of the data available in the latest report on Lisbon (Pais 2016) bearing in mind the right to the city and spatial justice. The report was made in the aftermath of the 2008 financial crisis, a situation that always represents a time when money

and properties change hands. In Lisbon, one of the best examples to illustrate this was the sale of a state fund called 'Sete Colinas'[5] to a German investor. The fund consisted of several historical buildings and large plots of land inside the city (70 thousand square metres and 72 million euros) (Antunes 2016).

Even before that though, as João Ferrão[6] shows in his eloquent evaluation of Portugal's spatial planning policy (2012), there were already problems which he analyses in detail. His conclusion is that when it comes to spatial planning in Portugal, the main problem is cultural. There is a lack of *territorial culture* in society, common to both citizens and institutions, and also a prevailing *spatial planning culture* that is stuck in the old modern paradigm where bureaucracy is excessive and the administrative process too complex (Ferrão 2012: 117).

In order to change this situation it is necessary to reveal society's shared beliefs and values as these are the key to linking institutions and citizens (Ferrão 2012: 125). For this to happen, the spatial planning community should take more initiatives and play a more active role in discussing the country's issues. Decision-makers should make spatial planning a priority, invest in further training and encourage the participation of the many intervening actors in society that lead to successful urban planning; and citizens should be taught to demand more from spatial planning and take a more active role (Ferrão 2012: 131–134).

As much as it is aesthetically appealing to see investment take place in the city of Lisbon, thus making the city look better, the question is: who will this investment benefit? Is the investment creating stronger foundations for the population living there to become more secure and sustained? According to what values is Lisbon's landscape being rationalized?

From the analysis we have just carried out, we conclude that the investment orientation, supported by public policies, has created a spatial

---

5   Among the properties included in the Sete Colinas Fund were the Almada-Carvalhais Palace, a national monument since 1920 (Largo do Conde Barão), the Portugália building (Avenida Almirante Reis), the São Paulo building, several buildings in Praça de São Paulo and Avenida 24 de Julho, and a building in Portas de Santo Antão (next to the Coliseu dos Recreios theatre) (Antunes 2016).

6   João Ferrão (*b.*1952) is a Portuguese geographer who was Secretary of State of Spatial Planning and Cities in Portugal's 17th Constitutional Government (2005–9).

dynamic that encourages an easy way in and an easy way out of the city, where cruise passengers are on the rise especially since 2008 (Pais 2016: 234), where the airport increased the number of passengers by 71 per cent between 2002 and 2013 (Pais 2016: 388), where accommodation is widely available and Erasmus students make up 10 per cent of university students. Lisbon is a good city for temporary visitors and in recent years the city seems to have worked hard to make them feel welcome. That has been achieved.

But what about Lisbon's inhabitants? There has been a loss of jobs, a loss of inhabitants, a GDP decrease and a large loss of businesses. We can say that this has happened as a result of the 2008 financial crisis. But the question is whether the recent investments being made will enable us to recover from it better, creating a more resilient economic structure that allows our growth to be sustainable. How is *territorial culture* being promoted by recent investments and construction work in the city? Is a dialogue with the city's actors taking place? What are society's shared values so that the city we envision is a shared, democratic exertion of power?

It is important to reflect on the fact that Lisbon is not Portugal and that macrocephaly should not be encouraged, meaning that Portugal should make an effort to stop conceiving its territory as 'Lisbon, and the rest of the country', which is culturally a much embedded attitude that strengthens the country's macrocephaly. This cultural attitude becomes a major obstacle to changing one of the most visible national spatial problems, the primacy of its capital. Ultimatley, it also affects negatively the capital itself putting excessive weight on social equipments, urban structures and administration, lowering living standards and expectations for the population.

A slow, carefully planned and progressive effort in the direction of distributing power and resources to other districts should take place. But for this to happen as successfully as possible, not only Lisbon but also the rest of the country would need to take significant steps to change the way they conceive spatial planning (and acknowledge its usefulness), which implies the cultural change that Ferrão rightfully calls for. It would also require coherent articulation among all the districts where all society's actors would have to be aware of their role and contribution towards the successful planning and implementation of a national spatial planning

policy. Cultural changes are usually slower than we would wish but this would be a beneficial move in the right direction.

Coming back to where we stand now, when it comes to the city of Lisbon, we consider that an excessive focus on tourism, along with the recent changes in the main arteries of the city that aesthetically made the city look better, can create an illusion of prosperity for those who visit the city but not for those who inhabit it. The question is: to whom is the city being shaped for? How is it shaping its inhabitants? Moreover, if the idea is that Lisbon should 'hang on to' tourism (as if it were the city's only resource for keeping the economy going), and if the number of beds available is rapidly on the rise, such cultural attitude not only promotes an economy with low sustainability (highly dependent on the tourist flow) but greatly contributes to cultural disempowerment by turning the city's inhabitants into hosts with no life of their own, living in the limbo of those who come and go, experiencing space as a living paradox where they see themselves excluded from the place they inhabit. Lisbon demands more from all of us and an adequate spatial strategy, politically endorsed, is key in order to shape the city's social and economic success. Only then the city's landscape will be rational embodying simultaneously, spatial justice, politics and aesthetics.

References

Adorno, Theodor, Walter Benjamin, Ernst Bloch, Bertolt Brecht, & Georg Lukács (2006). *Aesthetics and Politics* (afterword by F. Jameson). London and New York: Verso.

AML (2016), Website da Área Metropolitana de Lisboa. http://www.aml.pt/aml/territorio/. Accessed August 15, 2016.

Antunes, Marisa (2016). 'Investidor alemão compra Fundo Sete Colinas'. *Expresso*, April 2.

Aveline-Dubach, Natacha, Sue-Ching Jou, & Hsin-Huang Michael Hsiao (2014). *Globalization and New Intra-Urban Dynamics in Asian Cities*. Taiwan: National Taiwan University Press.

Benjamin, Walter (2002). 'The Work of Art in the Age of Mechanical Reproduction.' *Selected Writings*, vol. 3, 1935–1938. Cambridge (MA): Harvard University Press.

Boaventura, Inês (2016). 'Medina avisa que reverter a "degradação" dos transportes públicos vai levar uma década', *Público*, September 20. https://www.publico.pt/local/noticia/medina-avisa-que-reverter-a-de gradacao-dos-transportes-publicos-vai-levar-uma-decada-1744673. Accessed October 9, 2016.

Cipriano, Carlos (2016). 'Por que é que o Metro de Lisboa funciona tão mal?', *Público*, September 17. https://www.publico.pt/local/ noticia/por-que-e-que-o-metro-de-lisboa-funciona-tao-mal- 1744294?page=-1. Accessed October 9, 2016.

CML (2012). 'Lisboa Administrative Reform Published', CML Official Website. http://www.cm-lisboa.pt/en/news/detail/article/lisboa-ad ministrative-reform-published. Accessed August 15, 2016.

De Blij, Harm (2010). *The Power of Place: Geography, destiny, and glo- balization's rough landscape*. New York: Oxford University Press.

Deleuze, Gilles, & Félix Guattari (1987). *A thousand plateaus: capi- talism and schizophrenia*. Minneapolis and London: University of Minnesota Press.

Faria, Natália (2014). 'Portugal já é o país da União Europeia onde nascem menos bebés'. *Público*, July 10. https://www.publico.pt/ sociedade/noticia/portugal-ja-e-o-pais-da-uniao-europeia-onde- nascem-menos-bebes-1662352. Accessed August 15, 2016.

Findley, Lisa (2005). *Building Change: Architecture, Politics and Cul- tural Agency*. London and New York: Routledge.

Fung, Archon, & Erik Olin Wright (2003). 'Thinking about empowered participatory governance', in Archon Fung & Erik Olin Wright (eds.), *Deepening Democracy: Institutional Innovations in Em- powered Participatory*. London and New York: Verso: 3–42.

Gilbert, Natasha (2016). 'Green space: A natural high'. *Nature* 531, S56–S57 (March 17). doi: 10.1038/531S56a.

Harvey, David (2008). 'The right to the city'. *New Left Review* II (53) (September–October): 23–40.

____ (2013). *Rebel Cities: From the Right to the City to the Urban Rev- olution*. New York: Verso.

Haus, Michael, Hubert Heinelt, & Murray Stewart (eds.) (2004). *Urban Governance and Democracy: Leadership and Community Involvement*. London and New York: Routledge.

Hillach, Ansgar, Jerold Wikoff, & Ulf Zimmerman (1979). 'The Aesthetics of Politics: Walter Benjamin's 'Theories of German Fascism'. *New German Critique* (17) Special Walter Benjamin Issue (Spring): 99–119. doi: 10.2307/488012.

Hough, Susan (ed.) (2009). *Predicting the Unpredictable: The Tumultuous Science of Earthquake Prediction*. Princeton: Princeton University Press.

Jackson, Tim (2011). *Prosperity Without Growth: Economics for a Finite Planet*. London: Earthscan.

Mumford, Lewis (1966). *The City in History: Its origins, its transformations, and its prospects*. London: Secher & Warburg.

OECD (2011), *OECD Environmental Performance Reviews: Portugal 2011*. OECD Publishing, http://dx.doi.org/10.1787/10.1787/9789264097896-en. Accessed August 15, 2016.

Pais, Paulo Prazeres (ed.) (2016). *REOT 2015: Relatório do estado do ordenamento do território*, volume I, Câmara Municipal de Lisboa, DMU, Departamento de Planeamento, Março 2016.

Paquot, T. (2009). 'Qu'appelle-t-on un territoire?', in Paquot, T., and C. Younès (eds.), *Le territoire des philosophes: Lieu et espace dans la pensée au XXe siècle*. Paris: La Découverte: 9–27.

Plato (1930). *The Republic*, Books 1–5 (translated by P. Shorey). Cambridge, MA: Loeb.

Plato (1935). *The Republic*, Books 6–10 (translated by P. Shorey). Cambridge, MA: Loeb.

Pordata (2016). 'Índice sintético de fecundidade'. Base de dados Pordata, http://www.pordata.pt/Municipios/Índice+sintético+de+fecundidade-739. Accessed August 15, 2016.

Pordata (2016). 'Taxa de analfabetismo segundo os Censos: total e por sexo', http://www.pordata.pt/Municipios/Taxa+de+analfabetismo+segundo+os+Censos+total+e+por+sexo-721. Accessed August 15 2016.

Rancière, Jacques (2013). *The Politics of Aesthetics*. London: Bloomsbury.

Saraiva, Hermínia (2016). 'Transportes de Lisboa perderam 59,7 milhões de passageiros', *Dinheiro Vivo*, September 17. https://www.

dinheirovivo.pt/empresas/transportes-de-lisboa-perderam-597-mil hoes-de-passageiros/. Accessed September 17, 2016.

Schlitte, Annika, Thomas Hünefeldt, Daniel Romic, & Joost van Loon (eds.) (2014). *Philosophie des Ortes: Reflexionen zum Spatial Turn in den Sozial und Kulturwissenschaften*. Bielefeld: transcript.

Soja, Edward W. (2010). *Seeking Spatial Justice*. Minneapolis: University of Minnesota Press.

Stokes, David, & Nick Wilson (2006). *Small Business Management and Entrepreneurship*. Boston: Cengage.

United Nations, Department of Economic and Social Affairs, Population Division (2014). *World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352)*. http://esa.un.org/unpd/wup/highlights/wup2014-highlights.pdf. Accessed August 15, 2016.

Technology and urban space: on the relation between the historical approach and the transformation of aesthetic values in Walter Benjamin

Nélio Conceição

Introduction

The aim of this text is to bring to light some features of Walter Benjamin's philosophy that could contribute to a discussion on aesthetic values. This contribution can be regarded from a direct and an indirect point of view: direct in relation to the texts where he explicitly tackles the concept of *value*; indirect in relation to the historical and critical background of his thought, which forms a constant *evaluation* of the transformations occurring in the conditions of human experience in modernity. These two elements often interweave. For instance, in the well-known essay "The Work of Art in the Age of its Technological Reproducibility", the historical perspective – the understanding of the changes in our relation to aesthetic phenomena resulting from the increasing reproducibility of works of art – is grounded on the distinction between cult value and exhibition value. We can also say that the second element, the historical and critical task of Benjamin's thought, is scattered throughout his entire work. I will focus my attention on two themes that involve this interweaving of aesthetic values and historical thinking, technology and urban space. By doing this, I have no intention of being exhaustive, but I am just bringing up fundamental features and interrogations in order to prepare further developments in regard to aesthetic values and their relationship to artistic, political and ethical questions.

1. History, art and technology

Prolonging the critical approach on modernity inherited from Nietzsche, Max Weber and Georg Simmel, and confronting this approach with aspects deriving from Messianism, German Romanticism, Marxism or *avant-garde* artistic practices, Benjamin's texts thoroughly evaluate the transformations, often contradictory, that occur in modern capitalist societies. The weakening of tradition as a condition for the formation of human experience (*Erfahrung*) and its substitution by new forms of lived experience (*Erlebnis*) is a central thread of this evaluation. This pair of concepts is fundamental in the texts written on Baudelaire, in "The Storyteller", in "Experience and Poverty" and, explicitly and implicitly, in the majority of his texts that combine critique of art and social and political analysis. They perfectly illustrate the distinctive working of Benjamin's dialectics by combining destructive and constructive elements, thus maintaining a historical tension between that what is lost and that what is gained in the periods of transformation.

Benjamin is one of the first authors to have studied the implications of the new technologies of recording/reproduction, such as photography and cinema, for the understanding of the artistic transformations in modernity. Nevertheless, his insightful remarks do not constitute a philosophy of technology with clear-cut principles; on the contrary, they are disseminated in a manifold of texts, often as remarks focusing on concrete objects and particular historical situations. If we understand the aesthetic values as axes around which the production and critique of aesthetic phenomena circulate, giving them meaning and feeding their practices, we can therefore investigate the role technology has had in the transformation and creation of artistic values. It is in this sense that Benjamin's thought reveals all its relevance. But defending its actuality in a straightforward way is perhaps less important than clarifying its complexity and making the effort to understand why today, while reading his texts, we can still have a sense of contemporaneity and why his concepts still prove to be fertile – the more we are able to detour them in order to answer the demands of our present time, the more fertile they are.

In "On Some Motifs in Baudelaire", published in 1940, Benjamin states that Baudelaire was the first poet to embrace the experience of shock as a core element of his artistic work.[1] This embracement was not merely an individual stance; on the contrary, it corresponded to the experience of his readers, a trait that justifies the successful reception of his work at the time. More forcefully than ever, lyric poetry was integrating social elements directly related to the transformations occurring in urban societies, with the shock effect being the most important of them all. The shock effect is closely tied to the perceptual, bodily and mnemonic transformations taking place in urban life, characterized by immersion in the crowd and the new sensory stimuli arriving from different objects and situations, often related to the new technologies and an increasingly mechanized world.[2] Also photography, states Benjamin, while fixating an event for an undetermined period of time, achieves it by imposing a shock on the event, freezing it and giving it a posthumous character. This idea is developed in the context of the analysis of the small-scale gestures unfolding a complex process which involves haptic and optical experiences. The perception conditioned by shock, on the other hand, is for Benjamin a formal principle of cinema:

> Thus, technology has subjected the human sensorium to a complex kind of training. There came a day when a new and urgent need for stimuli was met by film. In a film, perception conditioned by shock [*chockförmige Wahrnehmung*] was established as a formal principle. What determines the rhythm of production on a conveyor belt is the same thing that underlies the rhythm of reception in the film[3].

---

1   Walter Benjamin, "On Some Motifs in Baudelaire", in *Selected Writings*, Vol. 4, Cambridge/Massachusetts/London, The Belknap Press of Harvard University Press, 2003, p. 319.

2   In his interpretation of the urban element in Benjamin's writings, Graeme Gilloch sums up this idea as follows: "The hallmark of modern experience is 'shock'. This in turn engenders forgetfulness and a distinctive form of memory, the *mémoire involontaire*. In addition, the accelerated tempo and new, machine-based rhythms of metropolitan life lead to a distinctively modern temporal sensibility rooted in the commodification of time (equation of time and money) and repetition (fetishism and fashion)". Graeme Gilloch, *Myth and Metropolis. Walter Benjamin and the City*, Cambridge, Polity Press, 1996, p. 8.

3   Walter Benjamin, "On Some Motifs in Baudelaire", p. 328.

This line of thought, affirming the worthlessness of approaching the experience of shock from a strict negative perspective and suggesting that the best attitude is to find a way of adapting to it, turning it into something productive, giving it a social and political task, runs through many of Benjamin's texts. For instance, in the various versions of the "Work of art" essay, the shock experience encompasses not only film and photography, but also Dadaism, a movement which destroys the traditional forms of relating to painting, such as contemplative immersion (*contemplative Versenkung*).[4] It is also the persistence of tradition that is questioned in the destructive gesture of the Dadaists, particularly in regard to the transmission of former experiences with art and, if we want to add, of enduring aesthetic values.

The experience of shock is thus a transversal notion running through the strictly poetic elements (both in the sense of production and reception), the new technologies and also the sensorial and psychological factors. Psychoanalysis and the question of trauma, memory and remembrance are also part of Benjamin's reflection on remembrance, and they can be read alongside the texts where he combines the experience of the city with the experience of his own childhood, as in *Berlin Childhood around 1900*.

The issue of technology is also a gateway to Benjamin's particular historical materialism. The text "Eduard Fuchs, collector and historian" develops one of the most detailed presentations of the task he ascribes to the historical materialist. Though focusing on the nineteenth century, some of his remarks also clarify his perspective regarding technology. These remarks emerge in the context of his critique to "the bungled reception of technology"[5], a process sustained by a series of enthusiastic efforts incapable of confronting the fact that technology is intrinsically attached to the production of commodities, therefore serving the principles of capitalism. "Technology, however, is obviously not a purely scientific development. It is at the same time a historical one."[6] According

---

4   For the third version, see Walter Benjamin, "The Work of Art in the Age of Its Technological Reproducibility", in *Selected Writings*, Vol. 4, p. 266.

5   Walter Benjamin, "Eduard Fuchs, Collector and Historian", in *Selected Writings*, Vol. 3, Cambridge/Massachusetts/London, The Belknap Press of Harvard University Press, 2002, p. 266.

6   *Idem*, *ibidem*.

to Benjamin, the historical constitution of technology should drive us to question the positivistic and undialectical separation between natural sciences and humanities. It should force us to examine its destructive elements and by the same token rescue it from the narratives based on the sheer idea of progress. This means: opening up the possibility of blasting the historical continuum by pinpointing the destructive character of its energies – when adopted, for instance, by war and propaganda.

In what follows, Benjamin adds some critical remarks to the history of culture conceived as a discipline. His critique focuses on the attempt to see historical materialism as a history of culture, grounded on the possibility of presenting its content from a non-binding distance, by throwing them into relief. This model is based on illusion and false consciousness. It neglects not only the fact that there is no such thing as a dialectical historical approach without a dialectic between the present and the past, but also the fact that "the products of art and science owe their existence not merely to the effort of the great geniuses who created them, but also, in one degree or another, to the anonymous toil of their contemporaries. There is no document of culture which is not at the same time a document of barbarism. No cultural history has yet done justice to this fundamental state of affairs, and it can hardly hope to do so."[7] This means that the task of the historian is also to give voice to the anonymous of society, those forgotten ones who are the invisible counterpart of humanity's great cultural achievements.[8]

---

7   *Idem*, p. 267. In *Das Passagen-Werk*, the relation between barbarism and culture is mentioned again, in this case with a reference to the way it manifests in the survival of values: "Barbarism lurks in the very concept of culture – as the concept of a fund of values which is considered independent not, indeed, of the production process in which these values originated, but of the one in which they survive. In this way they serve the apotheosis of the latter <word uncertain>, barbaric as it may be." Walter Benjamin, *The Arcades Project*, Cambridge/Massachusetts/London, The Belknap Press of University Press, 1999, [N5a,7], pp. 467–468.

8   We can also say that while examining the city Benjamin pays attention to peripheral figures like the prostitute or the rag-picker. Alongside "marginal" figures belonging to the bourgeoisie, like the *flâneur* or the collector, they are part of the micrological analysis of modernity which reveal the contradictions of progress and the phantasmagorical character of the narratives, often mythological, sustaining the capitalist world. By giving them voice, he is putting in motion his own critical thinking and allowing for a different, often detoured, manner of understanding historical time.

Without disregarding the weaknesses of Eduard Fuchs' ideas, Benjamin nevertheless recognizes that he was capable of disrupting the principles of history of culture, coming closer to the dialectical task. Driven by his character of collector, he stepped into fields on the edge where the traditional concepts of art could only fail. "The order of values which determined the consideration of art for Goethe and Winckelmann has lost all influence in the work of Fuchs"[9]. It is in this context that Benjamin refers to the three main dialectic elements in Fuchs' oeuvre: the interpretation of iconography, the contemplation of mass art, the examination of the techniques of reproduction. These three motifs have in common the fact that "they refer to forms of knowledge which could only prove destructive to traditional conceptions of art"[10]. They accentuate the importance of reception in art and, within certain limits, also contribute to correct the process of reification occurring in a work of art, obliging a revision of the concept of genius and to prevent the excesses of formalism.

We can transpose these remarks, which directly articulate the changes in aesthetic values with technology and mass art, to the critique and the history of reception of other objects, whether from literature, cinema, photography or, more recently, digital art. They are part of the historical materialist approach and therefore they are apt to reveal new tasks and new social and political functions in art. It is important to mention that the ideas raised by Benjamin regarding the relation between art and technology should not be reduced to the question of reproduction/reproducibility, a question that was already examined by Eduard Fuchs and that guides one of Benjamin's most well-known and discussed essays. The connection between art, technology and history is all the more fertile when it encompasses a model of thought capable of weighing the processes of transformation, the destructive and constructive elements of culture, the dialectical relation between the What-has-been (*Gewesen*) and the Now (*Jetzt*) that forms the constellations showing new possibilities of reading.[11] And such notions as optical unconscious and innervation, which belong to a wider understanding of

---

9   Walter Benjamin, "Eduard Fuchs, Collector and Historian", p. 269.
10   *Idem*, *ibidem*.
11   Walter Benjamin, *The Arcades Project*, [N3, 1], p. 463.

the relation between perception, human body and technology, prove not only the complexity of his ideas, but also the capacity of his thought to envisage a productive task in the midst of the historical tensions.

2. Cult value and exhibition value

The "Work of Art" essay focuses on the manifold consequences originated by the mechanical reproduction of works of art. It analyses its impact on artistic processes, on the function of art and on its relation to society and politics; at the same time, it reflects upon the existence of forms of art such as photography and cinema which can reach the masses in an unprecedented manner. One important idea behind the essay is the transition from a historical period in art defined by the cult value to a period defined by the exhibition value of works of art. Cult value is based on authenticity and uniqueness. These are attached to the notion of the original, to the here and now of the original. By liberating the hand and allowing infinite and fast multiplications, mechanical reproduction dissolves the importance previously ascribed to the original and weakens its authority.

> The reason is twofold. First, technological reproduction is more independent of the original than is manual reproduction. For example, in photography it can bring out aspects of the original that are accessible only to the lens (which is adjustable and can easily change viewpoint) but not to the human eye; or it can use certain processes, such as enlargement or slow motion, to record images which escape natural optics altogether. This is the first reason. Second, technological reproduction can place the copy of the original in situations which the original itself cannot attain. Above all, it enables the original to meet the recipient halfway, whether in the form of a photograph or in that of a gramophone record.[12]

In order to characterize this loss of authenticity, Benjamin uses the term aura, previously introduced in his "Little History of Photography", and

---

12   Walter Benjamin, "The Work of Art in the Age of Its Technological Reproducibility" (third version), p. 254.

links it to a general process concerning important changes in the value of tradition engendered by photography and film.[13] He then recalls the fact that the earliest works of art were at the service of cult and rituals – first magic, then religious. In this regard, even the secular worship of beauty which started in the Renaissance and lasted three centuries is said to have a cult character. "With the emancipation of specific artistic practices from the service of ritual, the opportunities for exhibiting their products increase"[14]. This liberation accentuating the exhibition value, exponentially increased by mechanical reproduction, allows works of art to acquire different functions, especially political.

Cult value is thus correlated with aura; consequently, with the vanishing of the cultic function the aura of works of art tends to disappear. We can of course put this idea into question by asking if our contemporary experiences with art are completely freed from these "tissues of space and time" that maintain an unapproachability. Didi-Huberman, for instance, challenging the canonical readings, brings the notion of aura and its constitutive distance to a secular context in art in order to understand the minimalist movement.[15] In the case of photography – and it is not by chance that the decay of aura is associated with the pioneering work of Atget, for whom the human figure lost its privilege – the cult value resides quite obviously in the human face, particularly in the cult of remembrance which is so characteristic of the familial milieu.

---

13   The aura is firstly defined according to spatiotemporal categories: "the unique apparition of a distance, however near it may be" (*Idem*, p. 255). In a note, Benjamin explains that this distance is also the property of cult value: "The *essentially* distant is the unapproachable. Unapproachability is, indeed, a primary quality of the cult image" (*Idem*, p. 272). This is not the place to go deeper into the characterization of the aura (also in other texts) nor to question the reach of Benjamin's proposal, but it is nevertheless important to add that the usage of the term aura is far from knowing a stable meaning in his texts; instead, it seems to lead to a complex network of problems regarding human perception, our experience with objects and the mutable understanding of works of art.

14   *Idem*, p. 257.

15   Georges Didi-Huberman, *Ce que nous voyons, ce qui nous regarde*, Les Éditions de Minuit, Paris, 1992, pp. 103–123.

3. Notes on urban issues

The question of aesthetic values in Benjamin, as in several of his contemporaries who analysed the new technical means like photography and cinema, is closely tied to the study of mediated perception and the consequences resulting from changes in the conditions of perception. As mentioned before, the experience of shock, increased by the urban modern life, comprises an obvious perceptual dimension, often related to bodily-rooted reactions to technological stimuli. Used as a principle for understanding cinema, as Benjamin sometimes suggests, it might lead to an excessively narrow approach, but regarded as a conflation of key elements accounting for the changes of human experience in modern societies, it can still prove its fecundity.

In *Malerei, Fotografie, Film*, published in 1925, Moholy-Nagy, while commenting upon the new perspectives allowed by photography, and particularly upon a photograph where the distortion of the human figure is an appeal to experiment with different ways of looking at the picture (and at reality), talks about an "invitation to re-evaluate our way of seeing [*Aufforderung zur Umwertung des Sehens*]. This picture can be turned round. It always produces new vistas".[16] In fact, in this re-valuation (or transvaluation, to recall Nietzsche's motto of the "transvaluation of all values") we can detect a bond between perception and value, nourished by technology. Seeing differently as a possibility of revaluating phenomena is a sort of perspectivism, naturally not a rigorous Nietzschean one, but one that nevertheless breaks the fixed and longstanding images of the world and accentuates the multiplicity of points of view.[17] In this sense, it is also noteworthy that many of the avant-garde artists who explored photography and film as technologies

---

16 Lázsló Moholy-Nagy, *Painting, Photography, Film*, Cambridge, Massachusetts, The MIT Press, 1987, p. 61.

17 Is this just a metaphor? Is there any connection between the changes in perception mobilized by the new techniques of reproduction and the questioning of true and one-sided objectivities operated conceptually in the philosophical realm? We can perhaps talk about a simultaneous process here, a process of fragmentation which also occurred in other fields of knowledge: in science, in literature, in art.

capable of creating a new vision of the world, such as the circle around Bauhaus, Rodchenko or Dziga Vertov, did it by constantly addressing the urban landscape, exploring perspectives and rhythms, recreating reality by means of montage.[18] In the case of Moholy-Nagy, there is a clear relationship between the changes in the visual realm and social change, and throughout his practice, his teaching and his writings, we can discern a utopian element dealing with the complex political situation of the time.[19]

According to Benjamin, seeing differently has also a destructive element. In the second version of the "Work of Art" essay, the one discussed with Adorno and Horkheimer, the one that served as a basis for the French translation, Benjamin introduces the element of play as a counterpart to the decay of cult value. Play belongs to the cultural context of what Benjamin calls the second technology and is closely tied to the notion of experimentation. The distinction between first and second technology unfolds, respectively, the distinction between cult and exhibition value. The passage from the first to the second technology implicates the decreasing of beautiful semblance (of the "object *in* its veil" – in this second version, an equivalent to aura). "That what is lost in the withering of semblance and the decay of the aura in the works of art is matched by a huge gain in the scope for play [*Spiel-raum*]"[20]. According to this analysis, which is also a utopian projection on the future of art, cinema is in a privileged situation. Before describing the characteristics of the optical unconscious, as well as its technical components – close-up, slow motion or enlargement – Benjamin makes use of the dialectic ideas of destruction and *Spielraum*:

> On the one hand, [cinema] furthers insight into the necessities governing our lives by its use of close-ups, by its accentuation of hidden details in familiar objects, and by its exploration of commonplace milieu through the ingenious guidance of the camera; on the other hand, it manages to assure us of a vast and unsuspected field of action [*Spielraum*].

---

18   The last chapter of *Malerei, Fotografie, Film* is exactly the presentation of a project entitled "Dynamic of the Metropolis", dating from 1921–22, a purely visual experiment with filmed events in space and time. *Idem*, pp. 122–137.

19   See Victor Margolin, *The Struggle for Utopia. Rodchenko, Lissitzky, Moholy-Nagy, 1917–1946*, Chicago, London, The University of Chicago Press, 1977.

20   Walter Benjamin, "The Work of Art in the Age of Its Technological Reproducibility" (second version), n. 22, p. 127.

> Our bars and city streets, our offices and furnished rooms, our railroad stations and our factories seemed to close relentlessly around us. Then came film and exploded this prison-world with the dynamite of the split second, so that now we can set off calmly on journeys of adventure among its far-flung debris.[21]

Not only does this section describe a "revaluation of seeing", but it also adds that this description can promote the destruction of the images encapsulating our cities. The process of destruction leads to the creation of a free scope, which is however not an absolute emptiness but a utopian social space characteristic of the second technology. In order to become revolutionary, the goals of this space should be sustained by a collective innervation.[22]

Benjamin soon understood the pertinence of the connection between photography, cinema and the city. Without belonging to any avant-garde movement, Atget is nevertheless a good example of this connection and of the capacity of photography to clear the stereotyped atmosphere of urban space, in this case, of Paris. Nowadays, the *Spielraum* is necessarily different. In fact, more than to the creation of utopias, many photographers, film and video makers dedicate themselves to the study of architectonic and urban utopias set in motion a long time ago. In the work of Portuguese artists such as Paulo Catrica or Nuno Cera, photography and film (video) are used in order to deepen this study. They can be said to evaluate either the post-war utopias, or the complex tissues of modern cities, their peripheral areas and contradictions. And in both cases the evaluation is less concerned with final judgements and more with the issue of visibility, of presenting cities in a rigorous but manifold way, thereby allowing comparisons. This is perhaps the best method of dealing with the complexity of contemporary urban life.

Walter Benjamin's thought about the city encompasses several interests and approaches: 1) urban "pen-pictures", such as his descriptions

---

21   *Idem*, *ibidem*, p. 117.
22   Walter Benjamin, "The Work of Art in the Age of Its Technological Reproducibility" (second version)", p. 124. For a reading of the relation between technology and architecture in the context of this version of the essay, see Diane Morgan, "*Spielraum* et *Greifbarkeit*: un acheminement vers une architecture utopique", in Libero Andreotti, *Spielraum: W. Benjamin et L'architecture*, pp. 291–301.

of Naples and Moscow; 2) descriptions and analysis of the urban setting in the *Passagenarbeit* (the Paris writings); 3) autobiographical texts: "Berlin Chronicle" and "Berlin Childhood around 1900"; 4) radio broadcasts; 5) reviews of books dealing with the city (e.g. Franz Hessel's *On Foot in Berlin*).[23] If we agree on the constellational nature of the knowledge of the city enabled by his texts[24], we can find interesting parallels with the functioning of technological apparatus. In this sense, Graeme Gilloch argues that

> the shifting vantage-point of the film camera is also important. Benjamin's concern with the depiction of the urban is interwoven with a conscious refusal of or resistance to the presentation of an overarching, integrated, coherent view of the city as a whole. The imagistic approach highlights the fleeting, fluid character of modern metropolitan existence. It denies a systematic, stable perspective.[25]

Hence, besides the idea of constellation, technology increased an element which is fundamental in Benjamin's relationship with the city: physiognomy and, implicitly, the idea of decipherment. The city as a monad can be conceived as an "entity that encapsulates the characteristic features of modern social and economic structures, and is thus the site for their most precise and unambiguous interpretation."[26] Therefore, not only is it important to see differently, but also to be able to decipher.

---

23 Graeme Gilloch, *Myth and Metropolis. Walter Benjamin and the City*, pp. 1–20.

24 While trying to circumscribe Walter Benjamin's pertinence to urban studies, Michael Keith focuses not only upon some of his constellatory thematics, but also upon the analytical and political value these thematics still have today. They run as follows: the culture of money and the cultural production of economic value; problematizing the real and the production of space and time; the city as text and emblem; aura, distance and closeness and the problem of the city view; authenticity and urbanism as corporeal experience. Michael Keith, "Walter Benjamin, Urban Studies and the Narrative of Everyday Life", in *A Companion to the City*, (ed. Gary Bridge and Sophie Watson), Malden, MA, Blackwell, 2000, pp. 410–429.

25 Graeme Gilloch, *Myth and Metropolis. Walter Benjamin and the City*, p. 18.

26 *Idem*, pp. 5–6. On this topic, see also Willi Bolle, *Fisiognomia da Metrópole Moderna. Representação da História em Walter Benjamin*, São Paulo, Edusp, 1994.

Final remarks

The ideas outlined allow us to consider the city – understood as a re-search field ranging from philosophy to the arts – as a privileged space where the values of the present are discernible, where the historical tensions with the past can be measured, where we can project the ex-pectations and challenges of the future. Nowadays, the way we live and experience urban space is certainly different from the nineteenth century way brought forth by the quotations and observations of *Das Passagen-Werk*. It is necessarily different from the political and cultural context of the Weimar Republic and the years that culminated in the Second World War. Nonetheless, the purpose of this text is to stress the fact that the complexity of Benjamin's historical thought, the interweav-ing of political, technological or aesthetical questions, may provide us with "methods", concepts and insights that prove to be contemporary. Ethical and political questions arise constantly while reading his texts. Though in a detoured and disguised manner, we are heirs of his evalu-ation of modernity, of the difficult relation to tradition, of the ever-in-creasing blind discourse on technological progress, often unacquainted with the destructive counterpart of the new. What is interesting in Ben-jamin, as in contemporaries such as Moholy-Nagy or Kracauer, is not a plain belief in technological progress and modern urban life, but the ability to simultaneously understand the destructive and constructive el-ements of the transformations they encompass, which confront us with axiological questions.

References

Benjamin, Walter (1999). *The Arcades Project*. Cambridge (MA): Harvard University Press.
_____ (2002a). 'Eduard Fuchs, Collector and Historian', in *Selected Writings, Vol. 3*. Cambridge (MA): Harvard University Press.

_____ (2002b). 'The Work of Art in the Age of Its Technological Reproducibility' (second version), in *Selected Writings*, *Vol. 3*. Cambridge (MA): Harvard University Press.

_____ (2003a). 'On Some Motifs in Baudelaire', in *Selected Writings, Vol. 4*. Cambridge (MA): Harvard University Press.

_____ (2003b). 'The Work of Art in the Age of Its Technological Reproducibility' (third version), in *Selected Writings*, *Vol. 4*. Cambridge (MA): Harvard University Press.

Bolle, Willi (1994). *Fisiognomia da Metrópole Moderna: Representação da História em Walter Benjamin*. São Paulo: Edusp.

Didi-Huberman, Georges (1992). *Ce que nous voyons, ce qui nous regarde*. Paris: Les Éditions de Minuit.

Gilloch, Graeme (1996). *Myth and Metropolis: Walter Benjamin and the City*. Cambridge: Polity Press.

Keith, Michael (2000). 'Walter Benjamin, Urban Studies and the Narrative of Everyday Life', in Gary Bridge & Sophie Watson (eds.), *A Companion to the City*. Oxford: Blackwell: 410–429.

Margolin, Victor (1977). *The Struggle for Utopia: Rodchenko, Lissitzky, Moholy-Nagy, 1917–1946*. Chicago: The University of Chicago Press.

Moholy-Nagy, Lázsló (1987). *Painting, Photography, Film*. Cambridge (MA): MIT Press.

Morgan, Diane (2011). '*Spielraum* et *Greifbarkeit*: un acheminement vers une architecture utopique', in Libero Andreotti (ed.), *Spielraum: W. Benjamin et L'architecture*. Paris: Éditions de la Villette: 291–304.

Aesthetic Values Before and Beyond the Evaluation of Artworks

Nuno Fonseca

Introductory note

Thinking about aesthetic values, as often happens when we think about aesthetic concepts, properties or experiences, gives us the opportunity to question the term *aesthetic*[1] that progressively entered philosophical discourse during the eighteenth century. What makes it difficult to answer, though, is the fact that its meaning has oscillated over time and generated various misconceptions and ambiguities. These have been very common even among professional philosophers and one of them results in the reduction of the subject area and confusion between one field – *Aesthetics* – and another – *Philosophy of Art*. The reasons for this simple, and to many philosophers almost unproblematic, confusion are linked not only to historical, sociological or cultural reasons, but also to strong conceptual affinities between the '*aesthetical*' and the '*artistic*' to the point that some, who are consciously aware of the problem, argumentatively admit its inevitability and legitimacy, particularly in respect to the question of values. Aesthetic value and artistic value could, according to some, be equivalent since the value of artworks could lie in their ability to produce aesthetic experiences, and symmetrically, for others the aesthetic qualities of an experience

---

1   And let us not forget that the term has been used over the years in several different ways, sometimes simultaneously, to designate a kind of object, a kind of attitude, a kind of experience, a kind of judgement, a kind of value.

of a non-artistic object – such as a natural landscape for instance – could derive from the fact that we look at it as if it were an artwork.[2]

One of the main purposes of this chapter is precisely, as an endeavour to determine the meaning and scope of the expression 'aesthetic value', to argue that aesthetic and artistic values are not exactly the same even though the artistic value of an artwork may result in part from its aesthetic value. Moreover, other types of values such as cognitive, ethical, political and social shall every so often be taken into account in the evaluation of artworks. And one of the consequences of that distinction – between the aesthetic and the artistic[3] – is the fact that the range of consideration of aesthetic values goes way beyond the evaluation of artworks insofar as aesthetic experience is not an exclusive business[4] of the artistic domain.

Another important aspect for the clarification of the notion of 'aesthetic value', mostly in the context of our main line of research, is tied to the concept of *value* and to the close affinities between aesthetic values, on one hand, and ethical and cognitive ones on the other. This is precisely where I must now begin.

---

2   Berys Gaut clearly defends this point of view in *Art, Emotion, and Ethics* (Gaut 2007: 35).

3   More recently, this separation of the aesthetic and the artistic has been gaining some philosophical consideration among contemporary authors and even becoming a standpoint for analysis of aesthetic experience as in the recent book by Jean-Marie Schaeffer, *L'Expérience Esthétique*. See Schaeffer (2015: 40–45), and, more generically, the concluding chapter of Talon-Hugon (2004).

4   And 'business' might even be a good word here since the aestheticization of our daily activities, our social exchanges, communication, even politics and economy, has become an overwhelming fact in contemporary society. Some authors, among whom the French theoretician of hypermodernity, Gilles Lipovetsky, and his associate aesthetician, gastronome and film critic, Jean Serroy, even identify a global process of 'aestheticization of the world', with aesthetic values informing most aspects of culture, be they leisure activities, entertainment, industry, commerce or lifestyle in general. See Lipovetsky & Serroy (2013).

1. Values: aesthetic, ethical and cognitive

Aesthetics as a philosophical discipline is relatively young – with no more than three centuries of academic existence – and closely linked to philosophical modernity. On the other hand, the modern theory of *value* and the philosophical discussion of 'values' is even younger, happening solely during the nineteenth century, eventually reassigned from the doctrines of classical economy. Nevertheless, as may easily become obvious, many of the topics that came to occupy modern aestheticians and also many of the main problems of a modern philosophical consideration of values are as timeworn as Philosophy itself. In fact, the modern themes of *Aesthetics* and *Axiology* partly inherited there 'preconceptions' – how could it be otherwise? – from the ancient discussion over Beauty, Goodness [and Justice] or Truth. This is not the place for a history of philosophy or a history of ideas, but it is not inappropriate to remember that those universal archetypes were inherited and transmitted by some medieval Neoplatonist authors and commentators, some calling them '*transcendentals*'[5], i.e. the metaphysical properties, the most general – wider – notions of 'Being'. In a well-known text by Pseudo-Dionysius the Areopagite (sixth century), *De divinis nominibus*, which was the focus of numerous commentators during the whole period of medieval thought, those notions are the names of God, His characteristic attributes which emanate in all Creation[6]. In this Christian conception of those archetypes, the universe was a manifestation of inexhaustible goodness, truth and divine supersubstantial beauty. In the end, each of these characteristic attributes was a different aspect of the same reality. The value of creatures derived from the greater or lesser closeness with the metaphysical reality of God.

---

5   To be accurate, we must say that the inclusion of Beauty in the medieval theory of transcendental is not consensual, appearing only episodically in authors of a more Neoplatonic inspiration (Hilduin or John the Saracen, for instance).

6   For a brief presentation of the appropriation and evolution of this idea in medieval authors, see chapter 2, 'Transcendental Beauty', in Umberto Eco's *Art and Beauty in Medieval Aesthetics* (Eco 2002: 17 ff).

But even devoid of its religious (Christian) connotations – or, at least, those connotations having been sublimated – the discussion of the Beautiful, Good and True nonetheless inherited in the dawn of modernity some metaphysical connotations which linked them together. That is certainly why in the meditations of the Francophone rationalists, such as in Jean-Pierre de Crousaz's *Traité du beau* (1715), we can still find a concept of beauty as a sensorial manifestation of the true and the good[7] or even why, in the considerations of British philosophers (Shaftesbury and Hutcheson) from the beginning of the eighteenth century, we also find so many references to the intimate relationships between the notions of taste, virtue, moral and beauty[8]. [We could also remember how the ideal of the "Honnête Homme" in seventeenth century France – the perfect courtier, the upright man, who should have refined manners, good taste and a virtuous moral character – may have been an influence on those British considerations.]

But, in fact, it is not only due to metaphysical hints still present in Enlightenment thought that we can spot topical discussions that oscillate between the aesthetical, the ethical and the cognitive. These close affinities in philosophical reflection on aesthetics have been present since its early origins owing to the ambiguous character of the aesthetic field. On the one hand, it has always had an epistemological ground where the ambition to rationally understand the domains of the sensitive, the sensible and the imagination[9] is unmistakably manifest. This means that there is a purpose to find in sensibility an *analogon rationis*, a way of knowing, through sensorial perception (αἴσθησις), the natural and the human world. On the other hand, there is an intimate connection between taste and the power of appreciating, contemplating and/or judging the sensible and the formal qualities of artifacts and/or

---

7   See Talon-Hugon (2004: 42).
8   We could simply read, for instance, Lord Shaftesbury's *Characteristicks of Men, Manners, Opinions, Times* (1711), or Francis Hutcheson's *Inquiry into the Origin of Our Ideas of Beauty and Virtue* (1725) to find these topics and interconnections.
9   '§1 AESTHETICS (as theory of the liberal arts, as gnoseology of the lower faculties, as the art of beautiful thinking, and as the art of thinking analogous to reason) is the science of sensate cognition.' This was the definition provided by Baumgarten in his famous 1750s work *Aesthetica*.

natural phenomena – to the point that Kant identified German *Aesthetics* (Baumgarten) with what other Europeans (mainly British) called *Critique of Taste*[10]. This means that *Aesthetics* have always been concerned with evaluation acts and processes.

'Aesthetic reason' would thus share with theoretical and practical reason some judicative and evaluative inclinations and skills towards the phenomenal field, which allows it to consider the world in all its heterogeneity and qualitative depth. Aesthetic values, like cognitive and ethical ones, allow us to differentiate the world's objects and states of affairs inasmuch as they are neither equivalent nor indifferent in respect to their established relations with the subject of perception (αἴσθησις). Here, the reference to the subject of perception is surely a trademark of modern philosophical aesthetics. As opposed to previous periods where the objective character of judgements and evaluations about beauty was warranted by a universal archetype, by a (metaphysical or divine) transcendental, manifesting itself in the sensorial world at different degrees, in philosophical modernity judgement about beauty becomes a problem, located between the contingent subjectivity of perception (and affection) and the necessary normativity of an objective verdict – eventually accompanied by a claim for universality (as in the aesthetic judgement according to Kant). To discuss aesthetic values or, to put it in a different way, to know what enables us in a particular circumstance to say that something is beautiful or not (ugly?) has apparently been one of the features of this field since its foundation to the point that, even today, we can find many (inevitably simplistic) definitions of *Aesthetics* as the science that investigates the nature of beauty.

For this understanding of Aesthetics, many references have been made since the early history of this philosophical field to the principles of beauty, to poetic beauty and to the fine arts [*Beaux-Arts*, *Schönen Künste*, *Belas-Artes*], topics that have filled numerous pages of aesthetic

---

10   'The Germans are the only ones who now [1781] employ the word "aesthetics" to designate that which others call the critique of taste. The ground for this is a false hope, held by the excellent analyst Baumgarten, of bringing the critical estimation of the beautiful under principles of reason, and elevating its rules to a science. But this effort is futile.' (Kant 1998: 156).

doctrines.[11] It is also true though that, from an early stage, many other aesthetic properties – beyond harmony, unity and formal balance which were present in early treaties on beauty – like intensity or excess have emerged as well as other possible values for philosophical meditation on aesthetic experiences, in particular the 'Sublime', to which Edmund Burke partially dedicated his famous [*A*] *Philosophical Enquiry into the Origin of our Ideas of the Sublime and Beautiful* (1757), followed by the unsurpassable 'Analytic of the Sublime' (§§23–29) in the *Critique of the Power of Judgement* (1790). Although both – the *beautiful* and the *sublime* – deal with aesthetic experiences, each presents distinct phenomenologies and produces different emotions: pleasure – resulting from the harmonious interplay between faculties (understanding and imagination) in the case of the beautiful, at least according to Kant – and a mix of pleasure and pain or even terror and respect in the case of the sublime – according to Kant and Burke, respectively. As regards value, that is, as far as we refer to what makes those experiences worthy of desire or esteem, it is not misplaced to talk here about different values. In fact, this difference would easily refute the tempting – but too quick – hedonistic account of value for aesthetic experiences (the value of each aesthetic experience would be assessed by the amount of pleasure that it could provide). Later, during the twentieth century, for instance in the seminal article by Frank Sibley on 'Aesthetic Concepts' (1959), a wider and much more varied palette of aesthetic properties was added[12] to enrich the discussion about pluralism regarding aesthetic values.

---

11   Even though the whole book by Annie Becq on the origins of French modern aesthetics (1680–1814) provides abundant demonstration of this, one might find the perfect illustration of this evolution in early European aesthetics by simply reading the first part of book III, 'Towards poetic reason', which deals with the concept of Beauty and how it dominates the French aesthetic writings of the period, becoming the fundamental basis for the aesthetic value of art. See Becq (1994: 513–646).

12   In this influential paper, the British philosopher lists several 'aesthetic terms' while discussing the specificity of aesthetic concepts and how they cannot be reduced to non-aesthetic ones. Although his focus is the specific character of aesthetic concepts, it is clear that they can be used in the appreciation and evaluation of aesthetic objects and experiences and thus serve as a reference for aesthetic values. For the list of 'aesthetic terms', see Sibley (1959: 421–3).

Nevertheless, and before going further in the discussion of the plurality of possible aesthetic experiences and the refusal of aesthetic and artistic values to amalgamate, something more shall be said concerning the distinction between aesthetic, ethical and cognitive values, as until now we have mainly pointed towards their affinities and crossings rather than to their differences.

Intuitively, it may sound as though it is an easy task to distinguish values according to their different fields of influence and the different purposes of each type of evaluation. Yet saying that types of values are distinguished from each other because they apply to different fields – aesthetical, theoretical or practical – and goals – aesthetic appreciation, cognitive analysis or adequate action – does not seem very informative nor interesting enough, and it definitely begs the question. Listing the values of each field can also end up in the discovery that the beautiful, the good and the true are equivalent insofar as we try to spot what is positive or negative in each evaluating process in such a way that the beautiful would be what is good or true in the aesthetic field, and thus the opposite of what is ugly, which would then be the bad or false in aesthetic terms. But, yet again, what distinguishes the aesthetic from the cognitive or the ethical is what remains unexplained.

With no further delay, we should then say that *ethical* values govern or justify actions or behaviours, the practical choices that guide each individual in his daily intercourse with states of affairs and other individuals; *cognitive* values are those that manage the possibility and validity of knowledge; and *aesthetic* values are those that condition the appreciation, contemplation and evaluation of aesthetic experiences, in other words, those which enable the association between sensible, expressive and formal – configurational and structural – properties and qualities of objects and states of affairs – situations or events – and the corresponding affective responses of the subject who experiences them.

Nonetheless – and just to make distinctions harder – an aesthetic experience, in this simplified description, probably also implies a form of knowledge acquisition insofar as any experience will, in one way or another, bring new data about the environing world and about each person, enriching her as an individual, and for that reason showing

how aesthetic experience may have a cognitive value. Similarly, that same experience, as long as it implies a vital existential kinship with the world and the others, may, eventually, produce some insightful intuitions about the best way for action or interaction of that same individual with her environment and, therefore, carry some important ethical value. This last point is of course arguable since one can always object that the eventual ethical value of aesthetic experience will turn out to be like an extra, an unexpected surplus relative to the main content of the aesthetic experience – an experience that, according to tradition, is allegedly disinterested.[13] With respect to the cognitive value of experience, I would risk saying that it belongs inherently to aesthetic experience in the sense that this is not simply an experience of fruition but always an experience of knowledge, even if resulting from confused perception and (un)clear knowledge – to use a Leibnizian reference[14] – but nonetheless some kind of knowledge. Notwithstanding, what has just been said concerning eventual ethical and cognitive values of aesthetic experiences does not dismiss the need and opportunity to distinguish it from specifically aesthetic values, which can only emerge from such experiences.

---

13   The idea of the disinterestedness of aesthetic experience – or, at least, of the pleasure derived from it – has a long past, starting maybe with the 3rd Earl of Shaftesbury (in his already mentioned *Characteristicks of Men, Manners, Opinions, Times*) and having its most famous pleading in Kant's *Critique of Pure Reason*. For more details on the origins and development of the idea of disinterestedness, see Stolnitz (1961).

14   In a short 1684 text on cognition, truth and ideas (*Meditationes de cognition, veritate et ideiis*), G. W. Leibniz made a distinction between clear and obscure but also between confused and distinct knowledge and then spoke of clear but confused knowledge comparing it to what painters understand when looking at pictures: '*Similiter videmus pictores aliosque artifices probe cognoscere, quid recte, quid vitiose factum sit, at judicii sui rationem reddere sæpe non posse, et quærenti dicere, se in re quæ displicet desiderare nescio quid.*' This '*nescio quid*', which he will later refer to in his own famous *Nouveaux Essais sur l'Entendement* (1704) as the '*je ne say quoy*' (an expression that was actually in use at the time by painters at the *Académie Royale de Peinture et Sculpture*), might be the epistemic specificity of aesthetic experience.

2. Aesthetic and artistic values

As was initially said, we should not accept the confusion between the aesthetic and the artistic. In spite of their obvious affinities and the relevance of the study and consideration of artistic phenomena to understand the aesthetic phenomena properly, to the extent that artworks and artistic practices intensify, question, excite, present and represent our aesthetic experiences of the natural and human world, it is undeniable that we have or may have many aesthetic experiences that go way beyond the strict spectrum of the *artworld*, a world that is culturally and historically determined and is not necessarily universal (despite the global tendencies to become so). This means that aesthetic values are relevant in many non-artistic fields of human experience, such as the experience and appreciation of natural phenomena – something that has actually been part of aesthetic discussions, at least since the eighteenth century. It would suffice to remember one of the most interesting and informative accounts of the aesthetic experience of nature in the incomparable 'Cinquième promenade' of Jean-Jacques Rousseau's *Reveries of a Solitary Walker* (1776–78)[15] but we could likewise notice, for instance, the importance of natural object collections in herbaria

---

15   Although Rousseau's autobiographical and elegant literary rendering of his 'rêveries', which were to some extent his aesthetic experiences while enjoying the charms of the beautiful natural environment near the Swiss lake of Bienne, became paradigmatic when discussing the aesthetic experience of nature, Edmund Burke had previously devoted, although not very systematically, several pages of his *Philosophical Enquiry into the Origin of Our Ideas of the Sublime and Beautiful* (1757) to the aesthetic appreciation of vegetables (plants and flowers) and animals (e.g. sections 5 and 20 of Part II or sections 2 and 3 of Part III) and to the aesthetic experience of the sublime generated by landscapes, mountains, abysses or atmospheric phenomena (lightning and thunder) (section 17 of Part II, sections 14 and 17 of Part IV, etc.). Later, Immanuel Kant would of course also contribute in a significant way to the discussion of the aesthetic experience of nature when he addressed the issue in his own *Critique of the Power of Judgement* (1790), yet he frames the aesthetic judgement of natural objects as if they were art objects (created by God). For a more detailed consideration of Kant's aesthetics of nature, see Budd (2002: 24–89) and for a systematic account of environmental aesthetics, see also Carlson (2000).

and 'cabinets of curiosities', which developed as from the end of the sixteenth century, to acknowledge our aesthetic attitude towards natural bodies, places, landscapes, and so on. But then again, we can easily find a use for aesthetic values and considerations even in intellectual, scientific, philosophical and mathematical experiences – which can be appreciated beyond their purely cognitive value in an aesthetic manner for their elegance, harmony, design, etc.[16] – or in interpersonal, social or even socio-economic experiences – like ritual, religious and celebratory gatherings, object and equipment design, urban and spatial planning, just to single out some obvious ones – and obviously in our everyday experiences. We may name just a few where there are evident aesthetic features, such as sports and leisure games, gastronomy, collecting, hobbies, tourism, etc. But then again let us not forget that the simple fruition of ephemeral everyday experiences may contain aesthetic judgements and rewards[17], something that had already been noticed in the

---

16   Once again, since the first modern aesthetic philosophers, we can find considerations of aesthetic features in intellectual experiences like mathematics and geometry. See, for instance, chapter 3 of Francis Hutcheson's *The origin of our ideas of beauty, order, harmony and design* (1725), unequivocally entitled 'The beauty of theorems'. For a contemporary consideration of the aesthetics of mathematics, see Montano (2014).

17   In recent years, several articles and books have been devoted to everyday aesthetics. For instance, in the book edited by Andrew Light and Jonathan M. Smith, *The Aesthetics of Everyday Life*, among other topics, the aesthetic experience of everyday environments, like the unplanned elements of our surroundings, but also the way we are affected and influenced aesthetically by the planned configurations of urbanism and architecture are some of the issues mentioned. A different topic that is addressed is the way contemporary sports can be viewed aesthetically or even as art: and here images of Leni Riefenstahl's *Olympia* (1938), the German documentary film on the 1936 Berlin Olympic Games, might immediately jump to our minds. But the fact is that many sports have this aesthetic dimension in the eye of the beholder, which can be fueled by the pleasure given to the spectators by some elegant gesture, by a high standard and skill displayed for the performance of acts requiring a certain prowess, by the presence of a certain 'disinterestedness' besides the goal of winning (if we abstract ourselves from the professionalization of some sports). See Light & Smith (2005). Yuriko Saito, who focuses in the same book on the aesthetics of daily weather, offers in her own 2007 book a theory on everyday aesthetics, considering as aesthetic different characteristics of ambiences, features of transience in day-to-day environments (associating to

nineteenth and early twentieth centuries with Baudelaire's famous *The Painter of Modern Life* or Walter Benjamin's accounts of urban *flânerie*. In all these experiences, we can pertinently consider aesthetic values and properties such as the beautiful, the sublime, the picturesque, the elegant, the stylish, the dexterous, the transient, etc. The fact that some intentional factors or purpose, semantic content or an identifiable creator may be absent does not prevent the existence of aesthetic properties that can be enjoyed or aesthetic values that should be taken into account as factors of consideration and evaluation determining affective responses, preferences and choices.

In other words, aesthetic values are only a portion of the values that are taken into account in the evaluation of artworks or artistic practices. And here it is maybe relevant to recall and emphasize that artistic experiences are, obviously, not reducible to the experiences of reception and critique of artworks since they also include creative activities and mixed activities of enjoyment and participation in artistic events and performances. In all these artistic experiences, performances or events but also in the critical acts of appreciation and evaluation, some kind of cognitive, ethical, political, historical, sociological or economic values frequently need to be considered. Art is a cultural phenomenon, sociologically and anthropologically complex, which can be evaluated according to several criteria, amongst which aesthetic values are not even the most important and sometimes are in fact almost absent. In the complex but ever-growing world of the art market and the stratospheric financial value of some artworks, the valuation of art is not so much an aesthetic process of evaluation but instead a multi-agent process of (mainly financial) worth estimation, which depends on market laws, scarcity or uniqueness, but also on cultural aura, trends, originality, authenticity, craftsmanship or virtuosity, celebrity and market worth of artists, reputation of previous owners and sometimes academic and institutional legitimacy given by scholars and experts.[18]

---

them aesthetic terms like 'neat', 'messy' or 'disorganized') or the feelings related to ageing processes. For more details, see Saito (2007).

18   Recently, a collective book written by several scholars, most of them working at the École des Hautes Études en Sciences Sociales in Paris and under the direction of sociologist and contemporary art expert Nathalie Heinich and the already mentioned philosophers of art, J.-M. Schaeffer and C. Talon-Hugon, analyses a set

Furthermore, when considering art history and contemporary artistic production, we can easily recognize an avant-gardist tendency – a tendency that came after and opposed to the aestheticism of 'art for art's sake' for which the value of art was essentially tied to its ability to produce aesthetic experiences – a tendency to annihilate the aesthetic conditions and the sensorial or expressive effects of art, a drive willing to "de-aestheticize" art. This happened initially with the *ready-mades* of Marcel Duchamp, an artist who would assume and himself theorize the "de-aestheticization" of his ready-made choices. Indeed, in a later text about the 'ready-mades', Duchamp wanted to state very clearly that '[their] choice was not dictated by any kind of aesthetic delectation', but was instead 'grounded on a reaction of *visual* indifference, supplemented by a total lack of good or bad taste… actually, a complete anaesthesia' (my translation).[19] And subsequently, in the 60s and 70s, this attitude was reassumed by the promoters and creators of conceptual art. For them, a conceptual artwork should be free of expressive purposes and of aesthetic or emotional properties.[20] It should be reduced down to its sole conception or to the mere transmission of its idea(s), eventually giving up any kind of object or performance that could embody it, any object which might become prone to appreciation or evaluation.

The value of conceptual art was –so it seemed at least – deliberately condensed in its cognitive aspect. To be fair, conceptual art has frequently focused on social and political criticism and, for that reason,

---

of art values 'beyond beauty and ugliness', as the title says, such as *authenticity*, *perenniality*, *expensiveness*, *universality* and *responsibility* among others, that give a wide and plural perspective on the value of art that goes way beyond the traditional view of aesthetic evaluation. See Heinich, Schaeffer & Talon-Hugon (2014).

19   See Duchamp (1961: 191).

20   In a text published in 1967, the conceptual artist Sol LeWitt wrote some paragraphs on conceptual art where he declared: 'In conceptual art the idea of concept is the most important aspect of the work. When an artist uses a conceptual form of art, it means that all of the planning and decisions are made beforehand and the execution is a perfunctory affair … It is usually free from the dependence on the skill of the artist as a craftsman. It is the objective of the artist who is concerned with conceptual art to make his work mentally interesting to the spectator, and therefore usually he would want it to become emotionally dry.' See Alberro & Stimson (1999: 12).

referred mostly to ethical, political or cultural values.[21] Therefore, it has often been praised (or instead blamed) for its political ideas or socio-cultural criticism rather than for its aesthetic values. Despite its deliberate disallowance of aesthetic values, it does not mean though that conceptual artworks have been totally exempt from those kinds of values. As Elisabeth Schellenkens[22] argued, even ideas – if conceptual art can actually be reduced solely to those[23] – can have aesthetic properties and be appreciated for the sake of that: in the same way we can ascribe aesthetic qualities to intellectual non-artistic phenomena, like an elegant philosophical argument or a harmonious mathematical proof, we can assign aesthetic values to conceptual art ideas even more. Moreover, the performative gesture of presentation of conceptual artworks gives them – even if allegedly against their author's will – an aesthetic effect that at least virtually modifies (raises?) the value of those ideas. In other words, presenting an idea, a concept, with an artwork or artistic performance is not the same as providing a propositional content, a scientific or even a philosophical argument in a non-artistic manner. Notwithstanding, this is not the same as saying that we can derive aesthetic value solely from the fact of its being an artwork – which would certainly be a concession to the *aesthetic theory of art*, a position that I repudiate here. Nonetheless and despite presenting itself often without object and willingly without the traditional aesthetic properties, which we are immediately ready to attribute to an artistic object, this is not enough to deprive it of an aesthetic value, nor to forbid the possibility of its evaluation, considering its eventual aesthetic values beyond the claimed cognitive, ethical or political ones.

---

21 To be fair, this socio-political dimension of conceptual art is frequently inherent to the goals of conceptual artists inasmuch as questioning art, what it is, what can count as art and what role art and, therefore, artists play in society are at the core of what contemporary conceptual artists are trying to do.

22 See the chapter 'The Aesthetic Value of Ideas' in Goldie & Schellekens (2007: 71–91).

23 A view that Sol LeWitt advocated in his 1969 *Sentences on Conceptual Art*: '10 – Ideas alone can be works of art; they are in a chain of development that may eventually find some form. All ideas need not be made physical.' Likewise, Joseph Kosuth, in his famous *Art after Philosophy* (1969), also presents the tautology: 'the 'art idea' (or 'work') and art are the same'. See Alberro & Stimson (1999: 107, 166).

3. Aesthetic values before and beyond the evaluation of artworks

At this point it is obvious that since aesthetic and artistic values do not coincide, we are allowed to speak of aesthetic values before and beyond the evaluation of artworks. *Before*: since, prior to the evaluation of an artistic object, we must already possess a notion of aesthetic values which eventually might be taken – and frequently are taken – into account when artworks are considered, appreciated or critiqued. Primitive notions, like 'beautiful' or 'ugly' but also 'harmonious', 'ordered', 'complex', 'integrated', 'elegant', etc.[24] – not to mention some perceptual notions like 'coarse', 'shiny', 'strident', 'insipid' or 'fetid' – certainly precede the notion of artistic object or even art, but furthermore the essence of the artworks is an intensification, problematisation, reflection, expression or even a representation/presentation of the aesthetic properties and values of natural or everyday experience. It is worth noticing that the aesthetic properties of nature have inspired musicians and painters (landscape painting in general being an obvious example, but also symphonic poems or imitative songs) and that even the most common and trivial properties of ordinary objects and everyday experiences have been filtered by the eyes and the bodies of painters (in still-lifes and genre painting), writers (in novels and other literature formats), filmmakers (and not just in documentary films) and also other performance and contemporary artists, who in the last few decades have often blurred the frontiers between art and everyday life and addressed its most unassuming yet meaningful aspects. *Beyond*: for the reason that, as announced beforehand, we can put aesthetic values to good use in very different and non-artistic domains of human experience (science, religion, philosophy, mathematics, tourism, gastronomy, interior design, sports, eroticism…).

But what specifically are those aesthetic values, where, when and how shall we take them into account? That is precisely what we must

---

24 Some of these 'aesthetic terms', but not all of them, might actually be already found in the article mentioned by Frank Sibley. See Sibley (1959: 421–3).

research from this point on whenever we are dealing with specific instances of aesthetic experience and once a clarification of the connotations and critical range of 'aesthetic values' – as I hope I have managed to provide – has been made.

References

Alberro, A., & Stimson, B. (eds.) (1999). *Conceptual Art: A critical anthology*. Cambridge (MA): The MIT Press.

Becq, A. (1994). *Genèse de l'esthétique française moderne 1680–1814*. Paris: Albin Michel.

Budd, M. (2002). *The Aesthetic Appreciation of Nature*. Oxford: Oxford University Press.

Carlson, A. (2000). *Aesthetics and the Environment: The Appreciation of Nature, Art and Architecture*. London: Routledge.

Dewey, J. (1934). *Art and Experience*. New York: Penguin.

Duchamp, M. (1961). 'A propos des "ready-mades"' in *Duchamp du Signe: Écrits* (1994). Organized and presented by Michel Sanouillet. New reviewed and augmented edition by Elmer Peterson. Paris: Champs-Flammarion.

Eco, U. (2002). *Art and Beauty in the Middle Ages*. Transl. Hugh Bredin. New Haven: Yale University Press.

Gaut, B. (2007). *Art, Emotion and Ethics*. Oxford: Oxford University Press.

Goldie, P. & Schellekens, E. (eds.) (2007). *Philosophy and Conceptual Art*. Oxford: Clarendon Press.

Heinich, N., Schaeffer, J.-M. & Talon-Hugon, C. (eds.) (2014). *Par-delà le beau et le laid: enquêtes sur les valeurs de l'art*. Rennes: Presses Universitaires de Rennes.

Light, A. & Smith, J. M. (ed.), (2005). *The Aesthetics of Everyday Life*. New York: Columbia University Press.

Lipovetsky, G. & Serroy, J. (2013). *L'Esthétisation du monde: vivre à l'âge du capitalisme artiste*. Paris: Gallimard.

Kant, I. (1998). *Critique of Pure Reason*, in *The Cambridge Edition of the Works of Immanuel Kant*. Ed. and transl. by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.

Kant, I. (2000). *Critique of the Power of Judgement*, in *The Cambridge Edition of the Works of Immanuel Kant*. Ed. by Paul Guyer, transl. by P. Guyer and Eric Matthews. Cambridge: Cambridge University Press.

Montano, U. (2014). *Explaining Beauty in Mathematics: An Aesthetic Theory of Mathematics*. Berlin: Springer.

Saito, Y. (2007). *Everyday Aesthetics*. Oxford: Oxford University Press.

Schaeffer, J.-M. (2015). *L'Expérience Esthétique*. Paris: Gallimard.

Sibley, F. (1959). 'Aesthetic Concepts'. *The Philosophical Review* 68 (4): 421–450.

Sibley, F. (2001). *Approach to Aesthetics: Collected Papers on Philosophical Aesthetics*. Oxford: Clarendon Press.

Stolnitz, J. (1961), 'On the Origins of "Aesthetic Disinterestedness"'. *The Journal of Aesthetics and Art Criticism* 20(2): 131–143.

Talon-Hugon, C. (2004). *L'Esthétique*. Paris: PUF.

Images and Values: a Husserlian perspective

Claudio Rozzoni

1. Works of art neither exist nor do not exist

Starting with the Husserlian notion of *image object* which is to be found in the *Husserliana XXIII*, this short text aims to shed some light on a concept of image that is both *perceptual* without being *real* and *iconic* without being a *copy*. I do think – and I will try here to sketch out *why* – that this very notion merits all the more attention in a moment when the reflection on art has to deal with so-called 'immaterial' or 'ephemeral' works and, from many points of view, is undergoing a 'crisis of values'. Indeed, a renewed inquiry on the Husserlian characterization of images can represent an important step to establishing the basis and a possibility of development for a phenomenological comprehension of the relationship between *facts* and *values* in aesthetics; more precisely, between those unreal – yet perceptive – 'facts' called artworks and values.

In order to proceed in an orderly fashion, it is very useful firstly to pay attention to the way in which the notion of *image object* is called into question in Husserl's famous lecture course, entitled *Phantasie und Bildbewußtsein* [*Phantasy and Image Consciousness*] (1904/05) and edited in *Hua XXIII*. On this occasion, the father of Phenomenology famously tries to describe the essential characteristics of image distinguishing three moments that are supposed to constitute it. That is to say, the 'image thing [*Bildding*]', namely the *image carrier*, the 'image object [*Bildobjekt*]' (Husserl 2005: 20) and the 'image subject [*Bildsujet*]', which can be considered as what the image object is a 'representant' for (Husserl 2005: 21).

Regarding the term *representant* [*Repräsentant*], it has to be specified that by writing that the image object is a representant for an image subject, Husserl is not at all saying that it is hereby meant to be a sign

of it. The reference lies *in* the image itself, and this *internal* reference is indeed an essential characteristic that distinguishes an image from signs and symbols, which, in turn, refer to something else externally (Husserl 2005: 31). In his text entitled *Artificial Presence*, Lambert Wiesing significantly had recourse to this very background pointing out that on the Husserlian basis we can distinguish between 'the depiction that becomes visible' (Wiesing 2010: 31), that is the *image object*, and 'the material that makes the depiction visible' (Wiesing 2009: 31), namely the 'image carrier' (Wiesing 2010: 30). This consideration of image regards the case of 'physical images' – distinguished from what Husserl in the same course calls 'phantasy images' (cf. Husserl 2005: 20) – then this subdivision also regards images that we find in works of arts such as paintings, photographs, films and theatrical plays as well. The reason why I am calling Wiesing's text into question is due to the way in which he characterizes the consistence of the image object, which is that of 'an immaterial, which means exclusively visible, object … brought into appearance [*Erscheinung*] by means of a material image carrier' (Wiesing 2010: 35). According to such a description, the image object, it could also be said, is not the *fact*, but, at least at first sight, the *appearance* of the fact. We will return to this point later.

Thus, it seems that in perceiving an image, a not-seeing is implied in this process in the sense that in order for the image object to appear [*erscheinen*], the physical image does not have to be seen – although it is apprehended. What we properly see in addressing the image does not appear to be the image as a 'thing', that is to say, the very image that, for example, could possibly be broken by, say, an iconoclast. What we see in a regime of image consciousness is the image object, and the kind of perception we undergo is directed to – or, we could add, is addressed by – an 'immaterial … object', 'exclusively visible', an object that is 'not subject to the laws of physics'. An iconoclast, as I said, might destroy the image thing but, properly speaking, cannot demolish the image object (although destroying its material condition *de facto* causes its disappearance). In this very sense, it can be stated that the factual canvas can be ripped into pieces but not the image object appearing on its material basis. As Wiesing puts it, the image object is 'an artificial presence' (Wiesing 2010: 20). Also taking into consideration Fiedler's and Sartre's lessons, Wiesing notes that 'the visibility of

a pictorially depicted thing is not attached to a substance that could also be perceived by other senses' (Wiesing 2010: 20).

Let us now return to Husserl in order to specifically address, in his phenomenological terms, the issue concerning artistic images. If we refer to the case of a work of art, it seems that in experiencing it, we are not interested in its existence, or rather in any existing at all. According to this, it can be said that the image object appears without existing ('it has no existence at all' [Husserl 2005: 23]), indeed as a *nothing*. It surely manifests itself – we can see it, even though only see it – but as something that cannot be said to be existent.

With regard to this aspect, a document of great significance is a Husserlian manuscript dating back to 1906[1]. It is extremely important for at least two reasons. First, here Husserl directly refers to aesthetics [Ästhetik], not only to the essence of images in general then, but also to the essence of 'works of art'. Secondly, here Husserl introduces the question of the *value* of a work of art, giving us a first clue to defining his use of the term *Wertnehmung* in relation to these objects that seem not to be considered as existing. Nevertheless, works of art are not – *nota bene* – without existence. They are not 'non-existent', which would presuppose a negation of their existence and consequently a position taking toward their existence.

2. Being disinterested

As regards this aspect, another very important document can profitably be brought into play: a famous letter Husserl writes to Hugo von Hofmannsthal only a year after this first part of the manuscript I have just referred to. Hofmannsthal visited Husserl in December 1906. The former was in Göttingen (where the latter was teaching at the time) to give a lecture, and on that occasion offered the philosopher a literary

---

1   Cf. MS A VI 1 [1906–1918]. Preserved at the Husserl Archive in Leuven. I would like to thank the Husserl Archives in Leuven for permission to quote from Husserl's *Nachlass*.

gift, 'presumably' his *Kleine Dramen* [*Short Dramas*]². Husserl's letter is the response to that present, but very soon, and very clearly, becomes the occasion for him to speak about similarities between two figures, that of the phenomenologist and that of the artist. What is the position – he asks – the two of them share? The answer to this question interestingly concerns the question of attitude, to be precise, a very particular 'attitude towards all forms of objectivity' (Husserl 2009: 2). As far as the artist is concerned, s/he suspends 'all attitudes relating to emotions and the will which presuppose such an existential attitude' (Husserl 2009: 2).

Furthermore, it is worth pointing out the fact that here Husserl *de facto* also speaks about another precise figure, still an artistic one, representing an essential part of the aesthetic experience, that is to say, that of the spectator. For the spectator too, no interest in the existence of the object observed in contemplating the work of art is involved. If that were the case, the spectator would lose the 'purely aesthetic' experience (Husserl 2009: 2). He would be involved in – he would be concerned about – the existence of what he is seeing, he could feel joy at the existence of something, he could want to desire something, and so on. Here, as has already been suggested by other scholars, we can hear a Kantian echo insofar as for Husserl 'art must exclude all influences from the intellect and the will'³ (Wallenstein 2009: 4).

Husserl develops this very issue by stating that in the aesthetic attitude we suspend judgment about the things before us. Undoubtedly, we can still have feelings before the things we see in this attitude. Nevertheless, these are feelings which do not depend on the existence of what we are perceiving. We are not interested in the existence of the work of art. Of course, what we feel in aesthetic experience can possibly depend on its materiality. However, even in this case, our feelings do not rest on the materiality conceived of as something existing, and this holds true also when viewing a monochrome artistic image: a Gerhard Richter grey monochrome painting in this sense becomes an image object, 'the grey presenting itself in the image and as an image is not posited

---

2   Cf. Husserl (2009: 2, translator's note 2).
3   Regarding this aspect, 'in many of his descriptions, Husserl appears to be retrieving the Kantian vocabulary of imagination and beauty in the third *Critique*, for instance when he determines phantasy as the domain of "disinterestedness" … , "purposelessness", and "play"' (Wallenstein 2009: 3).

as *real*' (Lotz 2010: 178). We do not perceive this object as a real colour insofar as we are able to see the *image object* that the work of art makes appear for us, this *image object* which is not 'subject to the laws of physics', even though it appears on the basis of something physical.

In fact, Husserl recognizes a similar attitude as peculiar also to the philosopher. The parallel he draws in the letter to Hofmannsthal lies in the fact that phenomenologist and artist – and spectator, we might add – see the world as appearance [*Erscheinung*]. They are not interested in the fact world, but in the *sense* which makes this fact possible, in its ways of appearance. Thus, what both are looking at is this very sense, which the phenomenologist tries to give back with 'concepts', and the artist, who must 'have genius' (another Kantian reminiscence), through 'intuitions' (Husserl 2009: 2).

In the very same manuscript of 1906, this particular suspension of existence is related to the question of the *values* of the work of art. This issue can also be connected to another regarding the peculiar essence of artistic images: what is it that indeed allows an image to be a work of art? Not all images are works of art of course, but what then makes an image a work of art? Here Husserl calls into question two points as simple as they are decisive. He tells us that in aesthetic experience we have 1) a 'value perception'[4] [*Wertnehmen*], we have intuition of an 'aesthetic-axiologic object', and 2) we experience particular feelings, specifically those feelings I referred to earlier, that is to say, all those feelings which do not have their reason in the existence of the object we are looking at.

Besides, Husserl makes it very clear that once the relevance of these two moments is stressed, we have to face the problem of their relationship. In this intricate question, one thing seems at least to be certain: value and feelings, in aesthetic experience, are not related to the existence of the object, not to the object-thing, to the thing in 'flesh and blood', as is the case in the perceptual attitude, but to the *image object*, to its 'way of appearance, in and for itself [*Erscheinungsweise, an und für sich*]'[5]. In 'the aesthetic attitude', Husserl specifies, 'I do not'

---

4   Cf. Ms A VI 1/2a.

5   Cf. Ms A VI 1/2a. And not only regarding the work of art, seeing that Husserl, still in 1906, writes that 'different appearances of the same object are not equivalent in this affective direction. The disposition of vases, ashtrays, and so forth, in the

even '*think* about the appearance and do not make it into a theoretical object' or a 'practical' one, 'tak[ing] delight in it as something actual'. The existence here is 'out of play'. Husserl himself, at this very point, significantly writes 'see […] *Kant's* theory'[6].

3. Perception and belief

Once made clear that the appearance of the image object is neither characterized as existent nor as non-existent, one could ask about the existence of values and feelings experienced when viewing it: do they exist? It is a very complicated question indeed. We can start off by saying that, for Husserl, value is not something external to the object, but something of which I have intuition while looking at the work of art. Of course, Husserl, in this manuscript, recognizes more meanings of value, among which the technical value of the work of art (the example singled out here as elsewhere is the *Madonna* by Raffaello) and the value of the work in the art market – that, at least on a first level, can be seen as something external to works of art. However, the most critical point is that concerning the relationship between the originary aesthetic pleasure I feel in enjoying the work and my intuition of values brought about in this experience.

Husserl himself poses the question in a way that is as simple as it is inextricable: 'Now, how are value perception [*Wertnehmen*], value intuition [*Werterschauen*], and aesthetic pleasure related? […] Are they the same thing, or do they coincide?'[7]. The difference seems to be qualitative: appreciating, Husserl affirms, might require 'no grade' and be independent from the joy: 'The joy can become 0 (zero), and yet I appreciate'[8]. From this point of view, value does not seem to be an intensive notion, there does not seem to be any gradualness in evaluating.

---

drawing-room. "Which arrangement is most beautiful?"' (Ms A VI 1/12b, in Husserl 2005: 168).
6   Cf. Ms A VI 1/12b, in Husserl 2005: 168, Husserl's note.
7   Cf. Ms A VI 1/2a.
8   Cf. Ms A VI 1/2a.

Nonetheless, this case regarding the 'zero grade' of evaluating could certainly refer to the evaluating moment regarding technical value. In other words, I could appreciate a work of art without feeling aesthetic pleasure. And yet we have to inquire about a more intimate relationship between the moment of the *nature* of pleasure and the *intuition* of value. Before moving on, however, it can prove very useful to focus on some other points that have a correlation with those we have pursued till now.

In a manuscript from 1912, about six years after the one we have just discussed, Husserl develops some decisive notions related to the issue of the peculiar 'non-existence' of the work of art. First of all, one point is made clear. While contemplating 'performances in a stage play' or a 'painting', I am experiencing *ficta* complying with *perceptio*, and that from the very beginning. We live from the very beginning in a consciousness that does not presuppose any statement about the existence of what we are experiencing. In this case, we are 'living in the iconic consciousness' and 'take the image neither as existing nor as nonexisting' (Husserl 2005: 457). It should be noted how here we have an interesting name for this peculiar consciousness without existence: iconic consciousness. The fact that Husserl chooses to refer also to theatre is of particular interest because, six years later, he will linger again on the art of the stage in a text that occupies a relevant role in the definition of his idea of image. If I am experiencing, as in these cases, artistic images, I am not considering the 'nullness' of the image object. Surely, 'I *can* turn toward the image object. I can also carry out an act of disbelief, the consciousness of nullity' (Husserl 2005: 457), but this is not the essential moment of my aesthetic experience[9]. The artificial dimension of the image object does not seem to concern my intuition of the *subject* of the image[10], 'of which I am conscious in the free iconic exhibiting from the beginning, without taking a position with respect to it' (Husserl

---

9   'Consciousness of nullity' does not come about when I turn toward the image carrier, the bearer [*Träger*] of the image object, as Husserl also called it, but rather 'toward the image object' (Husserl 2005: 457).

10 'In both cases, further position takings can be built up: Thus, on the one hand, I describe the subject of the oil painting. I carry out explications, comparisons, and so on, actual "acts of judgment" – actual acts, though modified, since I precisely have no belief. But I can also judge about the things, human beings, and so on, belonging to iconic phantasy' (Husserl 2005: 457).

2005: 457). Also according to what we discussed above then, the *subject* is not considered to be existing any more than it is considered not to be existing. Thus, echoing the terms of the 1907 letter to Hofmannsthal, one could also suggest that we are not *interested* in its existence. This is certainly true if we hold to the fact that 'when we are living in an aesthetic consciousness', Husserl says, 'we ask no questions about the being and nonbeing of what … appears in an image' (Husserl 2005: 459).

Before we proceed in this direction, let me spend only a few words to clarify why, in the last paragraph, I chose to translate the Husserlian expression 'perzeptive Fikta' with the English translation '*ficta* complying with *perceptio*' instead of the more immediate 'perceptive *ficta*'. The reason why I preferred the first solution is that I sought to highlight a significant distinction Husserl makes between the two German terms that could stand for perception, that is to say, *Wahrnehmung* and *Perzeption*. It is important to introduce such a distinction because it is highly relevant for the characterization of the peculiar consciousness that accompanies aesthetic experience. As early as 1912, the notion of iconic consciousness entails a vision that is merely complying with *perceptio* [*perzeptive*] without being perceptive [*Wahrnehmung*-]. By posing this distinction, Husserl tries to distinguish between perception as *Perzeption* and perception as *Wahrnehmung*. To put it in simpler words, we could say that *Perzeption* is '*pure positionless*' perception (Husserl 2005: 556). As Husserl already pointed out in his *Dingvorlesung*, *perceptio*, unlike *Wahrnehmung*, 'does not hold [*nimmt*] anything as true [*Wahr*]'[11].

Furthermore, in the manuscripts from 1912, significant attention is paid to the connection between aesthetic consciousness – namely, this iconic consciousness complying with *perceptio* – and both the axiological and sentimental dimension. Husserl states this very clearly: 'In the

---

11   Cf. Husserl (1973: 16). This decisive point came to light in all its significance in a text from around 1918, which is to be found in the *Hua XXIII* as well, and which we cannot linger on here. In this text, Husserl pays attention once again – and this time even more considerably – to the theatrical experience. In order to give a phenomenological account of what we see on the stage as spectators, he underlines how even if it can be said that we see actors in flesh and blood, we do not actually see them as existing people. We are not at all interested in their existence.

case of iconic acts', 'aesthetic valuation is essentially connected with … the object's *manner of appearing*', and through these appearances, 'we not only have the feelings of aesthetic valuing but also the feelings (or *quasi*-feelings) awakened in us as "reactions" – fear and pity, and so on' (Husserl 2005: 460–461). What is particularly worth specifying here is that 'aesthetic feeling … *does not aim through the appearance but aims at it, and aims at the object only "for the sake of the appearance"*' (Husserl 2005: 464). Hence, this aesthetic attitude does not require 'that I posit the correlate'. I do not pose '*what* appears as such': I am not living it as existing, rather 'I live in the appearing' (Husserl 2005: 521). We cannot linger too much on this question, but suffice it to say that here it becomes clear how the possible interest in the existence of the subject 'shifts back', as it were, to its ways of appearance, namely to artificial presences that are not touched by such an existential issue.

This point deservers particular attention because it would be interesting to ponder whether there is a possibility of finding a moment of position taking precisely in this living without position taking, in this experience which merely fits *perceptio* [*perzeptive*] without posing any existence. Such an experience, as we have seen, entails a contemplation of the ways of appearance of image objects, which are more *impalpable and ungraspable* 'presences' than they are *facts*. A hint for inquiring about this opportunity can be found in Husserl's effort to attain the peculiar quality of 'as-if' feelings thought in relation to the notion of value. Husserl seems to point in this very direction when, in another manuscript from 1912, while reaffirming that in aesthetic experience 'I do not … carry out any position taking with respect to what appears', he envisages the *exception* of an 'aesthetic position taking that belongs to feeling' (Husserl 2005: 521), namely the possibility of 'carry[ing] out' an 'aesthetic valuing' (Husserl 2005: 522).

At this very point, we could easily come full circle regarding the references to the Kantian aesthetic judgment seeing that Husserl often refers to aesthetic value as a value entailing the idea of beauty[12], as he writes in a text probably from 1917, a 'beauty-value [*Schönheitswert*]'

---

12   'The taking of a position toward what appears when it is valued as beautiful and ugly' (Husserl 2005: 482). 'Now what appears does indeed stand before me as beautiful "by virtue of its mode of appearance"' (Husserl 2005: 522).

(Husserl 2005: 649). More precisely, this is in the particular sense of a beauty *felt* as a value insofar as in this case 'what appears stands before me in its value-characteristic' (Husserl 2005: 522). Even though in aesthetic experience we do not value the existence of the objectivity concerned, it is not for this reason that we cannot 'value its modes of appearance … or we can value it *as* appearing in such and such a way' (Husserl 2005: 647). If we can experience an 'object of the beauty-evaluation', 'we would then … have something beautiful that *exists*' (Husserl 2005: 649; my italics). Consequently, it seems that we have a sort of existence also in the case in which we are not interested in the existence of what we are viewing. At least from this point of view, it can be said that we have an existence in the non-existence of a 'mere figment', of 'an "image"', an existence 'which is precisely an ideal object and not a 'real' object' (Husserl 2005: 649). More specifically, in these cases we are dealing with a non-psychological ideality because the value 'does not lie' in my enjoying it because the value can 'exist without being enjoyed' (Husserl 2005: 649).

4. *Interested* spectators

This last remark leads us directly to a fundamental point regarding the possibilities of experience opened up for us in aesthetic attitude. I am going to conclude calling upon a course Husserl repeatedly intended to publish and in which 'the phenomenological claim of philosophy is presented and developed in all its extent'[13]. I am referring to the course Husserl gave in Freiburg in 1923–1924, and which is known under the title of *First Philosophy*. Here, many of the aspects we have dealt with up to this point find an interesting and renewed formulation in very intense pages about the nature of images. We have already noticed how, in Husserl's letter to Hofmannsthal, the figure of the spectator implicitly arose from the parallel between the artistic

---

13   Cf. Costa 2007: xi.

and phenomenological attitude. In *First Philosophy*, this very figure explicitly becomes a theme for the Husserlian reflection on image consciousness. It is important to recall that also in the pages we are referring to, as in the 1907 Letter, Husserl draws a parallel between artistic and phenomenological experience. Once again, echoing issues already raised in the previous work, in these pages the way in which the notion of value and feeling are concerned is highlighted in order to characterize the aesthetic experience. In the letter to Hofmannsthal, Husserl concluded by stating, although very quickly, that the artist 'follow[s], purely and solely, his *daimonion* … which drives him to an intuiting-blind production' (Husserl 2009: 2). Interestingly, it could be said that in *First Philosophy* this very 'intuiting-blind' dimension is brought about by the notion of value, by a *dynamis* that must have the validity of value (cf. Husserl 1959: 100). The artist works value through 'pleasure' and 'disgust' (Husserl 1959: 101). And again this validity is something I must feel because the aesthetic attitude is – once again bearing in mind Kant's *Third Critique* – that of sentiment. That is why the art critic cannot be told to operate in aesthetic attitude but rather in a theoretical attitude where, of course, he can deservedly speak of value, but only in this sense did the 1906 manuscript seem to refer, namely, to the sense in which a specific aesthetic pleasure can also not be implied.

Above all, it is important to stress here that the paradox regarding the relationship between interest and disinterest is finally formulated. One decisive point is reached at the moment when Husserl calls upon '*affective interest* [*Gemütsinteresse*]' (Husserl 1959: 194) and when value becomes the theme of an aesthetic intention. Thus, it can be said that a work of art is not judged but valued. Here again Husserl plays, as it were, with the German word 'wahrnehmen', which, as we have seen, as early as in his *Dingvorlesung* from 1907, he characterizes as 'holding as true', as a perceiving that entails in itself the character of *belief* (in opposition, as we have seen, to the *perceptio*, that is to say, a pure *percipere* without *belief*). On this very occasion, he plays with words by saying that 'the value itself, in its value-truth [*Wertwahrheit*], is not perceived [*wahrgenommen*] but, so to speak, held as value [*wertgenommen*]' (Husserl 1959: 104). Hence, even though the attitude of experiencing a work of art cannot be described in terms of *perception* (but

only in terms of *perceptio*), nonetheless it can at the same time be stated that in the aesthetic attitude we still have an act of position analogous to that of perception as *Wahr-nehmung* (the one in which, as we have said, we are interested in the existence of the thing experienced). In the aesthetic attitude we suspended the latter and we have no interest in the existence of the thing, but only in its ways of appearance. What kind of position are we allowed to experience then? Even if we can in reason affirm that the aesthetic spectator experiences these non-positional nothingnesses which image objects are – and that he is not at all interested in the existence of the things he contemplates – Husserl can now state that the 'aesthetic spectator lives in the interest of evaluation'. And this is a *positional* moment, a *Wert-nehmen*. (Of course, the art critic can at a later stage direct his interest to the object theoretically, but, indeed, only subsequently.) The value, 'as the *telos* of the sentiment', has to be, 'so to speak, already available as valued', 'valued in itself in the value perception': that is what finally made possible the perception of a 'value-object' (Husserl 1959: 104–105).

We can then 'not be' interested in the existence of the image object (only in its way of appearance), but we are interested in the *existence* of the value-object. Hence, not only in positional perception do we experience values but also, finally, do we discover values in the *nothingness* of art. These *nothingnesses* that we have called image objects, these *simulacra*, reveal themselves to be endowed with a possibility of opening values to us and, according to what Husserl affirms in *First Philosophy*, opening a particular *telos* to us which we *feel* and which possesses the productive power of the *dynamis*. As Vincenzo Costa put it so well, 'a value can appear in itself as something that attracts' us insofar as we 'feel it as something whose existence concerns our very existence' (Costa 2014: 140). In this very sense, before works of art and before our 'possibilities' of responding to the *values we perceive*, we can finally be said to be *interested* spectators.

References

Costa, Vincenzo (2007). 'Introduction', in Husserl, Edmund, *Filosofia prima. Teoria della riduzione fenomenologica*. Soveria Mannelli: Rubbettino: XI–XLVIII.

____ (2014). 'Verso una fenomenologia delle tonalità emotive', in G. Matteucci, and M. Portera (eds.), *La natura delle emozioni*. Milano: Mimesis: 129–143.

Husserl, Edmund (1959). *Erste Philosophie (1923/1924). Zweiter Teil: Theorie der phänomenologischen Reduktion*, Den Haag: Martinus Nijhoff.

____ (1973). *Ding und Raum. Vorlesungen 1907*. Den Haag: Martinus Nijhoff.

____ (2005). *Collected Works, Volume XI: Phantasy, Image Consciousness and Memory (1898–1925)*. Dordrecht: Springer.

____ (2009). 'Letter to Hofmannsthal (Göttingen, January 12th,1907)'. Transl. S.-O. Wallenstein. *Site* 26–27: 2.

Lotz, Christian (2010). 'Im Bilde sein. Husserls **Phänomenologie des Bildbewusstseins**', in Neuber, S., and R. Veressov (eds.), *Das Bild als Denkfigur. Funktionen des Bildbegriffs in der Philosophiegeschichte von Platon bis Nancy*. München: Fink: 152–167.

Wallenstein, Sven-Olov (2009). 'Phenomenology and the Possibility of a Pure Art: Husserl's Letter to Hofmannsthal'. *Site* 26–27: 3–4.

Wiesing, Lambert (2010). *Artificial Presence: Philosophical Studies in Image Theory*. Transl. Nils F. Schott. Redwood City: Stanford University Press.

The Value(s) of Cinema: *Mise-en-scène,* Point of View and Ethical Problems

Maria Irene Aparício

> 'One need not be a chamber to be haunted,
> One need not be a house;
> The brain has corridors surpassing
> Material place.'
>
> ***
>
> 'Delight becomes pictorial
> When viewed through pain,
> More fair, because impossible
> That any gain.'
>
> *Emily Dickinson*

On April 3rd, 1986, Stanley Cavell delivered a *Tanner Lecture*[1] at Stanford University on the subject *The Uncanniness of the Ordinary*. In that lecture the author returned to some subjects he had already dealt with in *Must We Mean What We Say?* (1969), including *the philosophy of the ordinary* e.g. the subjects of *language use*, *experience,* and *world*, to J. L. Austin's similar questions and also to Wittgenstein's *Philosophical Investigations*, particularly the issue concerning the relation between thought and practice. Cavell underlines the idea of 'its relative neglect in contemporary intellectual life' (Cavell 1986: 83), and he explains the meaning of the 'Uncanniness of the ordinary' as '… the possibility or threat of what philosophy has called scepticism, understood … as the

---

1   *The Tanner Lectures on Human Values* were founded on July 1st, 1978, by the American scholar Obert Clark Tanner (1904–1993) at Clare Hall, Cambridge University. Presented annually, the Lectures reflect upon the scientific and academic learning related to human values.

capacity, even desire, of ordinary language to repudiate itself, specifically to repudiate its power to *word the world*[2], to apply to the things we have in common, or to pass them by'[3]. In a certain way, he considers that one has to trust the value and truth of an idea or belief, or one may be missing important opportunities and experiences. Previously, writing about the relation between 'artwork' classification and evaluation Cavell also emphasizes that 'works of art are valuable', but their 'value is inescapable in human experience, and conduct is one of the facts of life, and of art, which modern art lays bare' (Cavell 1976 [1969]: 216).

This time, I will not take these questions any further, but Cavell's concern relates to my point since the purpose of this essay is an attempt to understand whether the specific 'language' of cinema, as both an art and a popular medium, can embody the 'uncanny of the ordinary', with it being a chance to learn something about the most trivial but fearful or (in)significant experiences of life which otherwise would be difficult to apprehend. That means to assume that film can describe and shape common feelings, and deal with philosophical questions like ethics and moral standards. Sometimes cinema's description or representation are even able to change audiences' opinions and political thinking about sensitive issues like otherness or human rights, for instance. Thus, the cross-question is: *what is the power of film to depict and influence people's decisions throughout the course of their ordinary lives, simply by highlighting their dilemmas, gestures and the possible future consequences of every action?* For Cavell 'the magic of Hollywood is that it offers us not an escape from the burden of, say, Kant but, precisely, engagement with it' (Melville 1993: 172–192, 173), which means to assume *film as philosophy* through the connection between art, experience and thought. The author distinguishes between '*language*, *experience* and *world*', whereas our main objective is to show the relation between *film*, *experience*, *world* and *moral life*. The issue leads us to the cinema's

---

2   Our emphasis.
3   Cavell (1986: 84). Carroll would express a similar idea about the relation between art and ethics: 'There has been … a gap between theory and practice with respect to the ethical criticism of the arts throughout the twentieth century – a gap intensified by philosophy's silence about the relation between ethics and art'. Cf. Carroll (2000: 350, 350–387).

connection with intellectual life – since films are related to human perception and thought. Besides, we are interested in understanding the relevance of contemporary art's practices to the discussion of humanities, including their influence on a future 'way of life' – a culture. The *cinematic aesthetic form* is an artistic and cultural representation, but also a popular one, with a strong effect on behaviours and peoples' lives.

In short, I will attempt to find evidence of cinema's link with ethical and moral value(s) by discussing the following subjects: a) what kind of values and judgments are represented in films? b) Can we learn about *human values* like freedom, courage, loyalty and honesty by simply seeing movies? c) How does the *mise-en-scène* and the uses of *points of view* shape moral thoughts and bring about ethical decisions, leading to the sense that good should prevail over evil? Finally, what 'practice' of value is this about? If some films can raise the vexed questions of human values, the contrary is also true; ethics and moral issues provide the cinema with the substance for its stories and mark the high point in contemporary debates on values in the arts and visual culture. I am quite sure that *Sophie's Choice* (Alan J. Pakula, USA, 1982, 152') is one of those films, and if analysed with reference to the narrative historical context it can be very revealing of what humans are capable of, and what one can do and think under difficult or extreme conditions.

In order to identify the relevant problems which arise from the cinematic *mise-en-scène* and the film editing, or montage*, of Sophie's Choice,* I will look at the characters' performances and their involvement in moral judgment processes. From a philosophical perspective there are some important concepts, such as *free will* and *reasoning and decision,* that we can apply to both the characters' performances and the spectator's judgment. These concepts will help us to understand the problem of showing *ethical and moral judgments* in cinema and to think about the differences between *describing* a value system and *representing* it in cinematic narratives, according to the actions of everyday life.

In this context, I will consider the *mise-en-scène* as a multi-layered process – form and narrative – that includes the general elements placed in front of the camera, like settings, props, costumes, the actors' performances, gestures or facial expressions, and camera movements and angles, but also cinema's possibility of breaking the sense of trivial 'pictures' and moments of ordinary life by changing points of view.

This way, the *mise-en-scène* can mirror human experience and life by spotting its movements and details that, beyond the camera, have little value or no visibility at all. Besides, this is a common assertion that had already been transmitted by some philosophers and film scholars; films allow us to *see* life as it has never been seen before. Trivial moments of life, once projected, may appear odd, unusual and unexpected experiences. By changing a single point of view, what is normal suddenly becomes uncanny, i.e. too close or too far, even difficult or impossible to explain.

In *Sophie's Choice*, the characters' stories appear like real sequences of events which they cannot control or interfere with. By simply using a character's point of view – which is, in fact, the director's voice – 'life' becomes something *to look at* and *to think about*: from both an aesthetic and an ethical viewpoint. The film's sense depends on the director's particular vision and system, though decoding its political meaning is not difficult if one has particular knowledge about the subject, or recognizes something related to both history and everyday life. With this in mind, it is not difficult to identify the current philosophical focus of public debate on *film and values*. *Sophie's Choice* exposes, indeed, a common knowledge of principles (i.e. moral rules or standards of good behaviour) as well as the consequences of any personal or political 'transgression'. In order to perceive the common/moral level of the film, I will address to both the *description* of values and their *filmic representation* in the context of history. Actually, the social and political condition surrounding Sophie's life – the war and its aftermath as one of the greatest tragedies the world has ever known – is an important reference for the *mise-en-scène*, tracing out the complicated lines of reasoning that lead spectators to form their own judgments and conclusions.

1. Mise-en-scène: pictures and values

The concept of *mise-en-scène* has been discussed by many authors, filmmakers and philosophers. On the one hand, the *mise-en-scène* is connected with the concept of *space* and the film frame; on the other,

it is an ambiguous dimension of cinema because it figures a subjective vision influenced by or based on personal beliefs or feelings rather than based on facts. Underlining its ambiguity, Adrian Martin refers to the concept as a controversial issue and even as an undefined term, but also as a kind of 'pure style': '*Mise-en-scène* can transform the elements of a given scene; it can transform a narrative's destination; it can transform our mood or our understanding as we experience the film. Style is not a supplement to content; it makes content – and remakes it, too, in flight.'[4] In other words, the *mise-en-scène* – i.e. the construction of an atmosphere, the feeling of a place or situation – is the first motto of the film in the sense that it expresses the mood of the characters, their principles of good and correct behaviour, etc. The second one is the *point of view* as we will see in our case study.

Sophie's Choice is Alan J. Pakula's cinematic adaptation of William Styron's homonymous novel (1979)[5]. The film tells the story of Sophie Zawistowska (Meryl Streep) a Polish Roman Catholic immigrant living in Brooklyn with her lover Nathan Landau (Kevin Kline), who is a Jew and a rather mad genius diagnosed as a paranoid schizophrenic, and who is obsessed with the Nazis narrow escape from justice. The film is set in 1947, two years after the end of the Second World War, and it is narrated by another character, Stingo (Peter MacNicol), an aspiring young writer looking for inspiration for his first novel. He soon becomes Sophie's and Nathan's closest friend. Stingo is newly-arrived from *The South*, a detail

---

4   Martin writes a brief history of the concept and refers to several different definitions used by other authors. He also says: '… the term seems to mean (a little mystically) everything, cinema as an expressive art form becoming synonymous with *mise-en-scène;* … [But] *mise-en-scène* is nothing very specific" and, finally, he says: 'For my part … I want to hold onto Ruiz's sense of *mise-en-scène* as always potentially transformative – but transformative in ways that refer to the entire materiality of cinema, not solely the inspiration of a director on set or the phenomenological subjectivity of enraptured viewers. Transformation is not transcendence.' (Martin 2014: 13, 19–20). However, this controversy is not our point here and this matter will not be discussed.

5   In 1980, the American writer William Styron (1925–2006) won the National Book Award with the novel *Sophie's Choice*. However, the book was very controversial and widely criticized. It was banned in Poland and South Africa, and censored in the Soviet Union. In the USA, it was forbidden in some high schools, probably due to the sexually explicit scenes.

which may be seen as an indirect reference to slavery, another ghostly reference and itself a tragic episode of history. In his 'Introduction' to the book *The Cunning of History* (1978), William Styron would state that: 'If slavery was the great historical nightmare of the eighteenth and nineteenth centuries in the Western world, it was slavery's continuation, in the horror we have come to call Auschwitz, which is the nightmare of our own century'[6]. Although we cannot be sure if this detail of the story means a fate of inherited guilt for the characters, it is quite clear that *human rights* are the general background of the film and that all the characters have to deal with a kind of guilt by descent.

Following the story we learn that Sophie is a survivor of the Auschwitz concentration camp. Her father was an anti-Semitic lawyer[7], yet he was shot in a camp, just like Sophie's husband. Sophie's life is a tragedy, full of secrets, lies and hidden details of her past life that she will tell only to her friend Stingo in that long flashback in the film, bringing back painful memories including the moment when she had to choose which one of her children would be murdered to avoid them both being killed. Her confession challenges the spectator's emotions and makes possible the audience's doubts and questioning about *the right* of Sophie – if any – to choose between her son and her little daughter in the extreme conditions imagined and depicted by the movie story.

In the same way as melodrama, the spectator's point of view is eventually based on his (absence of) experience but human common knowledge is that '*right must win over wrong*', which means that '*evil should have lost*'. However, this is not a drama but a tragic representation of a vivid painful event because, in this case, Sophie had to face an impossible moral/ethical dilemma; a situation in which a difficult choice had to be made between two possibilities, though both would have equally devastating consequences. In Aristotle's *Poetics* 'a tragedy … is a *mimesis* of an action – that is, it is [morally] serious and purposeful' (Baxter & Atherton 1997: 67); '… tragedy is an imitation,

---

6   Styron *apud* Rubenstein (1975: vii).
7   The film is not very explicit about this character, but Oscar D'Amore (2010: 70) identifies Sophie's father, Zbigniew Biégansky, as the ideologist for the 'final solution' and the mentor of the monstrous project '*Die vernichtung*' – the extermination of the Jews.

not of human beings, but of actions and of life. … Story, then, is the first principle and like the soul of tragedy, and characters are second'[8]. Thus my decision to include this film in my brief analysis has to do with cinema's ability to tell life's stories and their dilemmas. Besides, the film points out how difficult it might be to make a decision against one's own free will, and how the decision-making process depends on reason and emotions. What we know and what we see affect the ability to decide under certain circumstances, such as putting lives on the line which is exactly the case here.

Dealing with our first question – *what kind of values and judgments are represented in the film "Sophie's Choice"?* – we are confronted with pedagogical issues about *what to see* and *how to 'read'* that are also the basis for the moral approach of our case study. In point of fact, at a basic level of practical wisdom, we know that *values* are the principles people have that help them to live in a reasonable and safe way. Values are also beliefs, and sometimes they refer to goals to be achieved. But in any case it is obvious that human values are connected to both emotions and reason, which influence people's decisions, and in the case of movie pictures, strong feelings arise quickly and easily. Films produce, and frequently reflect, emotional pleas and political meaning, and those results are achieved with the *mise-en-scène* and editing techniques – the primary features of cinematic representation. This is probably the reason why so many movies are controversial regarding the possibility of being a good or bad influence on the audience's opinion, and even their decision-making. However, the process is neither simple nor natural. It always depends on the spectator's experience.

Analysing the issue of 'Spectator, audience and response', Patrick Phillips says the relationship between the spectator and the cinema is always changing from one film to another. He also states that every particular response to a film depends on crucial factors, namely, the assumption of different selves, including 'a cultural self who makes particular intertextual references (to other films, other kinds of images and sounds) based on the bank of material she possesses', and 'a private self who carries the memories of her own experiences and who may find

---

8   Aristotle (2002: 20, 22).

personal significance in a film in ways very different from others in her community of interest' (Phillips 1996: 119).

The previous question leads us to our second question: *Can one learn about human values like freedom, courage, loyalty and honesty by simply seeing movies?* As a matter of fact, it is by the *mise-en-scène* and the uses of *points of view* that filmmakers influence the audience. Light, scale and continuity are precision tools which can be controlled very accurately and which can produce very effective results. In *Sophie's Choice,* light and scale shape the characters' morals and thoughts and lead audiences to the sense that, in those circumstances, it would be completely impossible that good prevailed over evil. The flashback to Sophie's experiences in the concentration camp, namely her arriving there years before, includes several inserts of her face: *close-ups* showing her sorrow and pain. This intimate *point of view* – which Deleuze would call an *affection-image* – creates empathetic spectators and attentive listeners, and the audience really understands how Sophie feels. The spectator himself becomes aware of Sophie's dilemma. People understand that she could not decide in good conscience how to achieve the so-called *terminal value: freedom.* Likewise, one can perceive why Nathan would be unable to find *happiness* or *inner harmony*. The film arouses a kind of *social conscience*; somehow audiences worry about the Holocaust survivors and their injuries. Annette Insdorf points out the contribution of the montage of *Sophie's Choice* to the evidence that the Holocaust is still a dreadful monster; a ghost haunting both our present and the coming future. Sophie and Nathan will never be truly 'free'. Both of them bear the consequences of the war and live in pain.

One can see that, despite Stingo's true friendship due to his close companionship, Sophie would feel always guilty although the 'choice' she made in her dark past had not been guided by any independence or free will, but in subjection to a German agent who restricted her decisions. She was *forced to choose*. It was not really a *free choice*; it was not a choice at all, but an order she had to carry out, and one that could never lead her to any ethical decision.

There is no point hypothesizing about how one would react and respond to an external problem like that since we shall never know, but the film clearly shows how Sophie's particular response had an impact on her future (and, indirectly, on our present). Moreover, by seeing the film,

people become aware of how life is full of uncertainties, which can be a positive quality of experience as an interactive process that demands the highest standards of behaviour, but also have a negative effect with tragic results like those at the end of the film. In fact not only the dilemmas of the characters but also the clarity concerning their actions and moral life make it possible for the *harmatia* to be seen.

First used by Aristotle in his *Poetics*, the idea of *hamartia* is commonly understood to refer to the protagonist's fatal error or flaw that leads to the downfall of a tragic hero or heroine, and culminates in a reversal of his/her good fortune to bad. Analysing the Greek Tragedy, Aristotle stated:

> 'It is necessary, then, for the beautiful story to be simple rather than, as some say, double and change not to good fortune from bad fortune but the opposite, from good fortune to bad fortune, not because of wickedness but because of a great mistake either of one such as has been said or of one better rather than worse.' (Aristotle 2002: 34).

Ari Hiltunen considers it a 'powerful special effect' and notes that Aristotle 'places *value on hamartia*[9] probably because it is an excellent means of achieving the state of undeserved suffering that in turn creates an intense emotional impact. Disaster caused by *hamartia* is the consequence of actions performed with the best of intentions and so is effective in arousing our sympathy' (Hiltunen 2002: 17). In Sophie's life the turning point of *harmatia* is the moment when she arrives at Auschwitz. When the German agent is turning his back on Sophie after a brutal moment of interrogation, she possibly changes her children's destiny, calling again for his attention and claiming that she is not Jewish but a Christian; she appeals to him, insisting that her presence there is a mistake. Of course we can argue that, one way or another, the children's lives were in danger, but Sophie's desire to save them might have speeded up the agent's hideous decision. And that was the first moment of her tragic fall even if she was an "impossible" heroine in the very sense of tragedy. The second fateful moment was when Sophie decided to renounce a *comfortable life* with Stingo and go back to her lover, Na-

---

9   Our emphasis.

than. The consequences of her final decision could not have been worse and the circle of tragedy is finally complete.

According to Aristotle, *the tragic hero is never an evil person*. His downfall is not a consequence of any vice or wickedness, but brought on him by *hamartia*. In this way, the story best suited to tragedy is the one of a man or a woman who falls from a state of happiness to a state of misery from some great error and not crime. And that is basically the story of Sophie's life. Julian Young underlines how 'the meaning of this word [*hamartia*] has been … debated. It has been variously translated as tragic "fault", "flaw", "mistake", "fallibility", "frailty" and "error"' (Young 2013: 35). Young also became aware of the implications in the debate of understanding the value of moral judgments since it makes clear why the tragic catastrophe cannot occur on account of extreme culpability. In short, Sophie could not stop blaming herself for sending her child to her death but, as a result of the *mise-en-scène*, audiences can understand her dilemma and forgive her.

2. Point of View: Is *Sophie's Choice* Right or Wrong?

It seems quite clear that the plot line of *Sophie's Choice* is not the Shoah. However, from the very beginning the film directs our attention to the Holocaust with a simple but effective *close-up*: the tattooed number on Sophie's arm. As a result of television's massive broadcasting of Second World War newsreels and footage, almost everyone immediately recognizes the tragic moment of history behind this picture. By *representation,* and a subjective point of view of the camera identified with Sophie's eyes, audiences are able to call to mind the most hideous facts which were beyond all reason and remain very close to the inhuman experience of what Kant describes as *radical evil*. But, despite *Sophie's Choice* dealing with the most disturbing historical events of the 20th century, the film is not exactly about a transcendental evil; rather it relies on the practical question of *supporting values with reasons*. It discusses the difficulty of one's decisions under circumstances beyond one's control and the probabilities of being right or acting wrongly.

According to Rokeach (1973), 'family security', i.e. 'taking care of loved ones', is a *terminal value* and refers to a 'desirable end-state of existence'. Moreover, this concept's descriptors include 'present and future responsibilities', 'rights and wrongs' as well as ethical and moral standards. Of course the problem of human values is very complex, but it seems very clear that there is a difference between *describing* a value and *judging* the consequences of an action based on values. In my opinion that is what this film is about; what kind of principles belong to an ethical and/or moral system and how is it possible – if it is possible – to put the proposals into practice unconditionally? After all, here is a mother who should have taken good care of her children but instead she 'let her daughter die' in a gas chamber just to keep her son alive. *Was Sophie's choice right or wrong? Would her life have been different if she had not had to make that choice, and both children had died?* Could she live with her own guilty conscience? Are these the right questions? After all, she was unhappy because of something she felt she had done wrong.

At this point, I would like to return to the topic of the *point of view* in film in order to show its relevance to the question of values in cinema. The *point of view* may be physical or moral; as a specific and technical rule of cinema it is described as a *form*, meaning both a *quantitative* and *qualitative* representation. It refers to *what* reality is framed and *how* it is framed. Thus, a *point of view* is always a question of perception and perspective. That is, it is a matter of *scale* and *position* from which something is viewed or judged. An object's dimension and angle determines its visibility and relative importance in a specific system, a 'landscape', for instance, but also a story. In *Sophie's Choice*, the physical *point of view* of the camera shapes the moral points of view of the characters: Sophie feels guilty, Nathan can neither forget nor forgive, Stingo just wants to be happy and share his happiness with Sophie. The historical events are distant enough to give the story its strength. The film itself looks at *Sophie's* motives, calling into question the mother's reasons and her own judgment.

Using a *comparison method*, the film makes an explicit and vivid analogy between the value that Sophie wanted to achieve – *to keep both her children alive* – and the value she had to sacrifice to keep at least one of them: *to give up her daughter and 'send' her to her death*. She consented to one child's murder to save the other. But, at the end of the

film, there are still questions without answers because, in actual fact, there is no solution to this particular dilemma. All the possible arguments are aporia and become inalienable. In any case, Sophie would be dealing with the *ultimate sacrifice* – her children's death, both or just one. And it would make no difference what decision she came to because any decision would go beyond any *moral and ethical system.* That is to say, on the one hand, none of her possible choices or arguments could ever be the best, i.e. the true or the right one; on the other hand, the cultural assumptions and ideological premises that inform the common audience's encounter with this film could never support any consideration concerning everyday actions for the reason that, only for Sophie was this *a matter of life and death*. And *real death* is the only fact in life that will never be trivial. In fact, our access to the tragic moments of Sophie's life is only possible by her narrative, which is basically a *tacit description* in the sense of Seymour Chatman: 'Sentences that tacitly describe … direct our primary attention … to the story events' (Chatman 1990: 38).

> 'Films, obviously, are more visually specific than novels, and filmmakers traditionally prefer visual representations to verbal ones. In other words, the medium privileges tacit Description. The choice of certain actors, costumes, and sets and their rendition under certain conditions of lighting, framing, angling, and so on all constitute what Aristotle called *opsis* or spectacle' (Chatman 1990: 38).

The *mise-en-scène* is also a *tacit description.* By remembering past events, Sophie gives audiences the right to look at her tormented soul. She describes the fateful moments of her past and what the spectator sees is not the history, but the story… the film of one mother's tragedy or, quoting Ruth Wajnryb, 'a space between speech and silence'. Wajnryb suggests that when someone describes events, all judgments depend on 'the mechanics of interpretation [that] are only partially contained in the text itself; they are also partially contained in the being(s) who serve as audience for the text' (Wajnryb 2001: 176). Perhaps we might say that, with regard to this Holocaust narrative, a post-memory syndrome should be considered which affects both the characters and the spectator, and this is definitely another major question concerning values for cinema that we cannot discuss here.

My limited analysis of just one film does not explain all the problems of cinema and the representation of values. This is a simple

introduction to an issue to be discussed in the near future – a work in progress. But there is one obvious foregone conclusion: in *Sophie's Choice* the spectator stays at the level of *description* and faces the problem of the interpretation of her narrative and its philosophical inquiry: the (un)fairness of her decisions, the limits of her freedom and free will. Consequently, one could ask if, by expressing more than one possible meaning, intentionally or not, films can prepare audiences for the dilemmas of life's ups and downs. Writing about teaching the Holocaust, R. Clifton Spargo argues that: '"Teaching the controversy" has become a pedagogical catch-phrase in certain circles, and it can be an artificial exercise, especially if all one intends by it is to illustrate how conflicting hermeneutical procedures … necessarily arrive at different interpretations of the same text' (Spargo 2008). *Sophie's Choice* seems to match this assumption, and that is probably the reason why it was a very much criticized film. However, values like freedom and free will, judgments like right and wrong, are frequently abstract, vague, relative or intangible, and cinema can make them 'real' and 'vivid'. Filmic narratives as well as pictures of life can engage audiences in different values existing in a form that can be seen or felt, giving the audiences the necessary concreteness to understand their complexity. Thus, Sophie's story reminds audiences how important values are, not only on great occasions and in decisive actions but also in every moment of an ordinary life. And the difference between the characters' (and the director's) *points of view* and the spectator's conviction is a kind of consideration beyond common sense, which is a decisive factor in shaping a *philosophical* and ethical *perspective* as well as shared experiences that create a community.

References

Aristotle (2002). *On Poetics*. South Bend: St. Augustine's Press.

Baxter, John, & Patrick Atherton (eds.) (1997). *Aristotle's Poetics*. Montreal & Kingston: McGill-Queen's University Press.

Chatman, Seymour (1990). 'What is Description in the Cinema?' in *Coming to Terms: The Rhetoric of Narrative in Fiction and Film*. Ithaca: Cornell University Press: 38–55.

Cavell, Stanley (1976 [1969]). *Must We Mean What We Say?* New York: Cambridge University Press.

Cavell, Stanley (1986). 'The Uncanniness of the Ordinary', The Tanner Lectures on Human Values, delivered at Stanford University.

D'Amore, Oscar (2010). 'The Sophie's choice: el factor letal'. *aesthethika, International Journal on Subjectivity, Politics and the Arts / Revista Internacional sobre Subjetividad, Política y Arte*, Vol. 6, (1) (Octubre): 69–80.

Hiltunen, Ari (2002). *Aristotle in Hollywood: The Anatomy of Successful Storytelling*. Bristol: Intellect Books.

Marques, António (2015). *A Filosofia e o Mal: Banalidade e Radicalidade do Mal de Hannah Arendt a Kant*. Lisboa: Relógio D' Água Editores.

Martin, Adrian (2014). Mise-en-scène *and Film Style: From Classical Hollywood to New Media Art*. London: Palgrave.

Melville, Stephen (1993). 'Oblique and Ordinary: Stanley Cavell's Engagements of Emerson'. *American Literary History*, Vol. 5, No. 1 (Spring): 172–192.

Phillips, Patrick (1996). 'Spectator, audience and response', in Jill Nelmes (ed.), *Introduction to Film Studies*. London and New York: Routledge: 113–141.

Rokeach, Milton (1973). *The Nature of Human Values*. New York: Free Press.

Rubenstein, Richard L. (1975) *The Cunning of History, The Holocaust and the American Future* New York: Harper and Row.

Spargo, R. Clifton (2008). 'Sophie's Choice: On the Pedagogical Value of the "Problem Text"', in Robert Eaglestone & Barry Langford (eds.), *Holocaust Literature and Film*. London: Palgrave Macmillan: 139–155.

Wajnryb, Ruth (2001). 'Holocaust Narrative', in *The Silence: How tragedy shapes talk*. Crows Nest: Allen & Unwin, 2001: 170–247.

Young, Julian (2013). *The Philosophy of Tragedy From Plato to Žižek*. New York: Cambridge University Press.

Vivre sa Vie: the *décalage* between language and life

Susana Nascimento Duarte

From early on (even in his debut works where the editing and the *mise-en-scène* still convey a certain preoccupation with narrative), Jean-Luc Godard's films not only show a taste for mixing categories, genres, discourses, media and stories, but in them also emerges the problem of the relationship between the seeing and the saying as one of its main dispositions. The array of words, quotations, cuts, images and references from various sources that had been with him since the beginning gradually loses its subordination to narrative concerns, insinuating itself on its own in its fragmentary autonomy. This results in two inseparable lines of events that, following the framework of analysis proposed by Jacques Rancière[1], are justly captured by the idea of the 'sentence-image': those that belong to the sphere of the iconic evidence of images, where all sorts of fragments give themselves as spectacle, irradiating the glow of their atomic and a-significant condition; and those that belong to the sphere of the articulation of these fragments into a sentence-like or discursive plane, combining in a significant way the various fragments detached from their initial contexts, stories and narratives. This is why we can venture to claim that the *Histoire(s) du cinéma* are the culmination of a process in which Godard tried to translate into cinematic terms an obsessive inquiry into language.[2] Guided by the idea of cinema's

---

1 Mainly with his *Histoire(s) du cinéma*, Jean-Luc Godard is the filmmaker that allows Rancière to illustrate the notion of the 'sentence-image' within the aesthetic paradigm. Cf. Jacques Rancière, 'La phrase, l'image, l'histoire', in Rancière (2003: 43–78).

2 This identification/confusion between language and words, though not very rigorous, is consistent with the way Godard uses several terms to name the same thing, i.e. everything that he opposes to the image and with which he is at war: words, text, language, literature, etc. It is, after all, a question of going back to a belief

lost childhood (the medium having, in the meantime, become shackled by a subordination to discourse, i.e. stories and narratives), Godard uses the 'sentence-image' to give film back to what resists language, meaning the enigma of life and bodies, thus returning the discourse back to them. Only now it is a discourse produced after undoing the predictable ties between the seeing and the saying demanded by the regime of representation, which means the possibility, as Deleuze said, of 'reaching bodies before words, so that they can produce their own words' (Deleuze 1985: 172, 173).

Regarding the desire to exorcise the model of representation and its limiting narratives and stories, found in this Godardian phase of reflexive cinema and understood as a retrospective and prospective meditation on what film and image are, Rancière draws attention to the survival of a discursive dimension whereby several heterogeneous elements are articulated into sentences, that is, by disfiguring the elements of the previous regime of representation, another drama arises.[3]

Vivre sa vie (1962) is one of the films that comprise the first phase of Godard's work. Here we start seeing the first signs of this disfigurement of the dominant order of representation through the dialectical use of editing as a way to link together heterogeneous material, meaning that the various fragments summoned by Godard link through shock.[4] This type of editing is influenced by the Brechtian paradigm, which intends to free representation from identification and the resulting fascination and absorption it creates in the viewers. This is achieved through

---

in the bodies, which demands, among other things, a work of 'deconstruction' of language and certainly also includes the matter of words.

3 Conversely, we could point out in this association of images and words, this sliding between them, a similarity with the work of dreams as it is described by Freud, more precisely the way in which it serves as a model for the figural and the study of art images. In it is emphasized the irruption of the sensory inside discourse and words, that is to say, the rupture with any idea of linkage that has the representative logic of the Aristotelian poem for a model.

4 Rancière identifies two types of editing in Godard: the dialectical editing that characterizes the first phase of his work, and the symbolic one, which alongside the former, Rancière uses to describe the *Histoire(s) du cinéma*. For the former, the main concern when bringing contradictory images together is the shock of opposites. For the latter, what matters, besides this shock, is the community of images, their conciliation through the 'sentence-image' (Rancière 2003: 43–78).

pedagogical procedures that break with the continuities and progressions of the narrative model.

A major contribution here is the 'sentence-image' mentioned above – the deviant use of words themselves being an example of it – in the way in which it meets/departs from the image to formally move alongside the film's theme: the heroine's difficulty in apprehending the world through language. The film culminates in a kind of *tête à tête* between the protagonist and the philosopher Brice Parain – that is also a *tête à tête* between cinema and philosophy – concerning the relationship between cinema and discourse in their connection to life and the world, in their different ability to externalize thought and express the possibilities of thought and life. Therefore, the film represents an inquiry into language that is also an inquiry into cinema through the means of cinema.

In fact, one can already find in *Vivre sa vie* a certain 'anxiety about the power of speech,' 'about the possibility of expressing those existential experiences' – on the one hand, love, on the other, the experience of oppression (in this case embodied by a prostitute) – that will mark Godard's later works (for instance, in Godard's politicized phase it will manifest itself in the question of how to represent in film the experience of class struggle) (cf. Badiou 2013/2010: 172).

In the case of *Vivre sa vie* it is important to look at the film right from the beginning from the perspective of its structure, as in Godard what one wants to show is inseparable from how it is shown, that is, from an explicit problematization of the narrative form and its cinematic conventions. Godard refers to the twelve episodes that compose the film as the stylistic option that most adequately gives visibility to the problem at hand, or the film's investigation – the apprehension of the movement of thought.

This deliberate structure echoes multiple references: St. Francis of Assisi's eleven *fioretti* from Roberto Rosselini's film (which brings to life in a completely novel way the narrative form of sketch-based films and to which Godard clearly refers in the script when mentioning the use of intertitles), the twelve stations of the cross and Bertolt Brecht. In interviews from the time, Godard explicitly ties this division in twelve episodes to his desire to emphasize what he calls the theatrical side of

the film and of cinema in general: 'This accentuates the theatrical, or Brechtian side.'[5]

As a playwright and also a theorist of what he practised, Brecht embraced the project of an epic, revolutionary theatre in which the author took a critical approach, regarding both the forms and the content, towards the illusions of a psychologically and narratively oriented theatrical drama. To put it briefly, Brecht wanted the audience to critically distance themselves from what they were watching instead of emotionally identifying with the flow of the story and the characters' psychology. By making people conscious of the theatre medium, Brecht sought to make them politically conscious. Consciousness about the medium would not only allow but also demand that the audience participate in a continuous process whereby the images, sounds and other phenomena that they were confronted with were examined. This enabled the spectator to produce constructive political and social critique. According to Brecht, the epic form should combine an old tradition with the current techniques of film, radio and theatre editing: it was a question of 'treating the elements of the real in the sense of an experimental composition through which epic theater, instead of reproducing the state of things, discovers it. Its discovery is achieved through the interruption of continuities.' This interruption consists in creating discontinuities, 'in undoing the articulations,' so that the situations can 'critique themselves dialectically,' that is, 'shock against one another'. 'Its main goal is to interrupt action –

---

5   Jean Collet, 'Entretien avec Jean-Luc Godard', *Cahiers du Cinéma* 138 (déc. 1962), revisited in Narboni & Milne (1986: 171–196). Here are some other excerpts of what was said by Godard in the interview from when the film came out: 'I start by the documentary in order to give it the truth of fiction,' 'I was trying to discover the concrete. The closer I get to the concrete, the closer I get to the theater. *Vivre sa Vie* is very concrete, and at the same time very theatrical,' 'By being realistic one discovers theater. Behind theater there is life, and behind life, the theater. I started from the imaginary and discovered reality [a reference to his previous films, described by Godard as films for cinephiles, whose starting-point is cinema itself and not life]; but behind reality, there is again imagination,' 'Cinema is spectacle (Méliès) and research (Lumière). What I have always wanted, basically, was to do research in the form of a spectacle,' 'The documentary side is: a man in a specific situation. The spectacle comes when one makes this man a gangster or a secret agent. In *Une femme est une femme* the spectacle comes from the fact that the woman is an actress; in *Vivre sa Vie*, a prostitute.'

instead of illustrating it or making it go forward. It is the retarding quality of these interruptions and the episodic quality of this framing of action that give epic theater its power' (Benjamin 2003: 24–26, 22, 27 and 21). Cut/division, interruption, suspense – terms markedly used in cinema – are procedures that create an effect of distantiation in the way theatre is apprehended and experienced (Didi-Huberman 2009: 61).

In *Vivre sa vie*, Godard makes use of similar techniques, directly inspired by Brecht, to question the cinematic act and the classic categories of representation, thus experimenting with the narrative form at the same time as he creates a self-reflexive cinema. Like in epic theatre, it is the showing itself that is shown, which allows the spectator to become conscious of the cinematic form as an element that one has to think and feel, as a form that informs reality.

Through the division into episodes and the use of intertitles, attention is drawn from the dramatic unfolding of Nana's story to focus instead on her reaction to each of the events as they happen. This distantiation from her story also serves as a way to constantly remind us that we are watching a filmed reality that despite some similarities with current life is a made-up reality nonetheless.[6]

For Godard, the twelve episodes, in their distance and exteriority to Nana's adventures, are paradoxically what makes it possible to reach and touch the opposite of this, meaning the maximum interiority identified with thought in action: 'I wanted to show the 'Adventures of Nana So-and-So' side of it. This division into tableaux corresponds to the external view of things that would best allow me to convey the feeling of what was going on inside… How can one render the inside? Precisely by staying prudently outside' (Narboni & Milne 1986: 187).

This underscores an important point to understanding both this film and Godard's cinema in general: by recording only the signs and external manifestations of beings, cinema becomes especially adept at revealing these as well as the existential reality in which they are immersed. That is to say that the fact that the camera superficially captures images and sounds, instead of making it inadequate for the apprehension

---

6   What we have here is a critical realism that supposes a critique of realism. Cinema does not give us access to a reflection of reality but to a snippet of reality, one that is simultaneously partial and inseparable from it.

of the interior 'selves', is exactly what guarantees its ability to show them in a new light. Nevertheless, the access to this interiority of a non-psychological nature cannot be confused, in Godard's case, with Antonioni's project, which proposed non-communication as cinema's fundamental theme.

> In any case, something is revealed. This is why Antonioni's cinema of non-communication isn't mine. Rosselini told me that I almost fell into the Antonioni error, but just escaped. I believe that sincerity is sufficient when one has this kind of problem. I think it is wrong to say that the more you look at someone the less you understand. Obviously, though, if you look too much you inevitably end up by wondering what the point is. If you look at the wall for ten hours on end, you begin to ask questions about the wall, and yet it's just a wall. You create useless problems. This, too, is why the film is a series of sketches: one must let people live their lives, not look too long at them, otherwise one ends by no longer understanding anything (Narboni & Milne 1986: 187).

This exteriority to which cinema, like other visual arts, is to a certain extent condemned can be highly suggestive if one believes that interiority is inscribed and manifests itself in actions, gestures and external situations that implicate us in our relations to the world and others: 'A painter that tries to render a face only renders the outside of people; and yet something is revealed,' says Godard, in an indirect allusion to the oval portrait episode. Or: 'It's very mysterious. It's an adventure.' *Vivre sa vie* was, then, 'an intellectual adventure. I wanted to try to film a thought in action – but how do you do it? We still don't know' (Narboni & Milne 1986: 187).

By wanting to cut out a thought in action, Godard identifies the thought with the body, overcoming the separation of spirit and body of classical philosophy (that is, it is mainly the behaviourist activity, which can be apprehended, more than its psychological content that he has in mind). In this respect, Godard reminds us of Merleau-Ponty who in 1945, in a conference called 'Le cinéma et la nouvelle psychologie', defended, against classical psychology, that a person's character is not a mysterious essence that reveals itself in time, but is immediately present in a person's acts. The philosopher challenged the notion that emotions can only be understood through the way in which they are individually experienced. He posited instead that they can be studied through behaviour since the only place where they exist in reality is in an individual's

expressions, gestures and acts. Merleau-Ponty added that cinema is particularly capable of illustrating this new psychology where thought becomes gesture and expression. To bring his point home, he finished the presentation with the following words by Goethe: 'The inside is the outside/what is inside is also outside' (Merleau-Ponty 1996).

Therefore, for Godard, psychology (in the classical sense) prevents the probing into the depths of the human being. The opening scene enunciates just that. In the sequence showing a conversation between Nana and her ex-husband Paul, the two play pinball while they exchange the scene's last words in a more relaxed tone than what was used previously. He mentions the school essays written by his father's students and how surprising some of them were. As we hear him quoting from them, the camera reframes Nana, isolating her in the shot in a pensive posture: 'A chicken is an animal with an inside and an outside. Remove the outside, there's the inside. Remove the inside, you see the soul.'

Godard's philosophical and cinematic program becomes apparent in this sentence (Cf. Sterritt 1999: 60–88): if films are nothing but audiovisual material suited for the recording of exteriority and if, despite that, some films risk going beyond these exterior representations and attempt to indirectly suggest man's interiority and psychology, Godard, on the other hand, without resorting to psychology and remaining at the surface of things, wants to go beyond to expose 'something else,' something more profound and enigmatic. This is what he tries to achieve in his process of film composition with the meeting of theatre and life (spectacle and research) mentioned above. He refers to this when he says, about *Vivre sa vie*, 'The film was made by a kind of second presence' (Narboni & Milne 1986: 187).

> What I am trying to do is to express thoughts rather than tell stories.
> It will not be a question of spying on the girl (François Reichenbach), trapping her (Bresson), or taking her by surprise (Rouch), but simply of following her; thus nothing else but being good and true (Rosselini)… Nana, who is gracious, which is to say, full of grace knows how to retain her soul all the while giving her body… I would like to try to make palpable what modern philosophy calls existence as opposed to essence: but at the same time, thanks to cinema, I want to give the impression that there isn't a true opposition between the two, that existence supposes essence and vice versa, and that it is beautiful that it should be so (Young 1963: 22).

In order to reach this something more – the soul – Godard disarranges the usual relations between word and image, that is, the way in which they mirror the articulation between interior and exterior, thus freeing them from the rules that have traditionally codified its manifestation within the regime of representation. The word is no longer a means to access what is hidden inside the soul, or what tells us about and describes what is far from the eyes. The dialogues in *Vivre sa vie* often seem like intrusions or interruptions in the storyline, having emancipated themselves from any intention to clarify motives and moods. And the visible, the image, is no longer subordinated to the restitution of a dramatic order, of an assemblage of actions and their causal logic of progression, devoting itself instead to the restitution of the being's unconscious gestures, of the silent words written on the bodies, to borrow Rancièrian terms.

Let us go back to the first episode, this time to its beginning.

Nana and Paul are filmed from the back. We can make out Nana in the mirror, but had we not seen her in the opening credits, we would have to wait until the end of the scene to actually see her. We will only see Paul properly at the end of the scene. By not making the characters visible to us, this apparatus of capture makes us focus on what is being said instead – what we have here is an extension of the procedures of distantiation, now applied to camera work. We find out that they are breaking off a marital relationship, that they have a son and, more importantly, that Nana does not know how to express herself. The lines revolve around the denunciation of incomprehension, meaning the difficulty in making oneself understood through language. Later in that scene, and afterwards, in the last scene, when talking to the philosopher Brice Parain, she says, 'the more one talks, the less the words mean.' Dissatisfied with words, Nana says she prefers to keep silent. Once again the plane of fiction, of Nana's story, uses the character's words to echo a problem that is not just her own but will also be the film's. Just as this character aspires to silence, the film aspires to be a silent film.

From this first scene then we are in the presence of not only a character and a fiction but also of a style – Godard's – which announces itself in the image and enunciates itself in the dialogues.

If we cannot see Nana, let us hear what she says instead. If she does not trust what she says, we will get used to the silence. The film prepares us for what will be its work, foreshadowed by the opening credits with the three shots of Nana's face seen in profile, from the front and in profile again: a physiognomic study that registers the human body's sculptural surface and gestures while sacrificing psychological depth and character development. The film tells us that we should not expect in-depth explanations of moods or ambitions, i.e. dramatic motivations in the basic causal sense. After this first scene, where we realize that Nana's breaking up with Paul represents a new start for her, we will find out what she does with her freedom and what the obstacles to that freedom are. The film will then circumscribe the obstacles that hinder the blossoming of this newfound freedom, which could also be called happiness. Nana's interiority is shown to us by a gaze that rests upon moments chosen at random from an itinerary.

> NANA: So things are not very happy then?
> YVETTE: No, it's sad, but I'm not responsible.
> NANA: I think we're always responsible for what we do. And free. I raise my hand, I'm responsible. I turn my head to the right, I'm responsible. I'm unhappy, I'm responsible. I smoke a cigarette, I'm responsible. I shut my eyes, I'm responsible. I forget that I'm responsible, but I am. No – it is how I say. To want to avoid it is foolish. After all, everything is beautiful. You only have to interest yourself in things to find them beautiful. After all, things are as they are – nothing else… A face is a face, plates are plates. Men are men. And life, is life.'

She outlines here a kind of ethics, which is also the film's: 'I'm responsible,' 'A face is a face.' An ethics that speaks for the film's aesthetics and allows it to be read as an ethics. It is a praise of surface, of a before or beyond language, that also reveals the film's metaphysical dimension. The difficulties of being, the contradictions and worries that come with trying to strike a difficult balance between oneself and the world, will be solved by her and the film by eschewing drama and accepting life as it is. This means not looking for a meaning behind things. We are faced here with a meditation on existence: 'Nana knows herself to be free, but that freedom has no psychological interior. Freedom is not an inner, psychological something – but more like physical grace. It is being *what*, *who* one is. Being free means being responsible' (Sontag 1994:205).

The *je suis responsable* implies the negation of the notion of *fatum*, or fate. That does not mean, however, that fate is not omnipresent in every shot and every scene throughout the film as some kind of threat (and, when seen retrospectively, it is clear that the film had been gathering and collecting evidence that Nana was destined to die in the end). Godard wanted Nana to be responsible in the existential sense, i.e. by creating her own destiny – and the tragic dimension is not any less powerful because of that. Her friend makes destiny responsible for her luck; Nana says that we are always responsible.

In the next scene, when Nana accepts to work for Raoul, she is responsible.

A 'normal' film would try to explore her motives, making us, at this point, dramatically conscious of her circumstances (lack of money, the separation from Paul) so that we could feel the pressure she was under (which is what happens in the case of Yvette, who, in this sense, represents that cinema). But Godard, in this scene, suspends the melodrama's usual expectations, the alibis of failure and compromise; in a word, the excuses. Godard wants us to feel the film's epigraph: *se prêter aux autres et se donner à soi-même.* It is through the body, through its postures, that Nana and the film reach the soul, reach thought.

The face, and the gaze, emerges as a privileged place for the camera to do its probing since it is traditionally thought to serve as a window to the soul. Maria Falconetti's face in the film *La passion de Jeanne d'Arc* (1928) by Carl Theodor Dreyer is a perfect example of just that: a face that in its maximum exteriority, in its passivity, shows the interiority; a face that is the place of ultimate revelation, with the soul or God passing through it. Anna Karina's face is, therefore, invested with spiritualism through its approximation to Maria Falconetti's. The approximation achieved through the editing of the two faces in the cinema gives rise to the approximation and the creation of correspondences between two far-apart worlds – Nana's itinerary and Jeanne d'Arc's passion. Jeanne d'Arc's face is not only the face that, in its absolute nudity, opens itself up to transcendence and attests to the passage of God as an invisible trace revealed in man. It is also a face that bares itself before threat. It refers back to a primordial ground that exists before individuation, to a space where all the faces of suffering and martyrdom converge to meet once more, (to the presence of the Other and of all humanity, as Levinas

would put it). On the other hand, *La Passion de Jeanne d'Arc* is a film that portrays the possibility of the language of the face becoming words, as stated by Dreyer himself: 'Joan of Arc is also words! Whether the text predominates, or the image, it is all the same. Each subject implies a certain voice. And that must be paid attention. One must find a way to express as many voices as possible' (Delahaye 1965: 15–34). In *Vivre sa vie*, Nana's true voice is her face and her body.

The face draws the limit between exterior and interior, but at the same time it erases that limit in the act of exposing it. Hence the preference for close-ups, especially after the scene in the cinema which introduces Nana's suffering. Nana's interior transformation, the metamorphosis of her soul, is something that exceeds and escapes her and that the camera tries to apprehend in its imperceptible manifestations, in the way that a face and a body might reflect it (the face and the body as places both of exposure and protection of interiority, producing an effect of ambiguity that oscillates between intimacy and distance or detachment: 'in your gestures there are words… but as silence… and it is that silence that I love…' ('*tu vois dans tes gestes, il y a la parole… mais sous forme de silence… et ce silence, je l'aime bien…*').

Therefore, drama is destroyed and replaced by the arbitrary observation of Nana's body's attitudes, movements and postures which, more than a political and social *gestus*, evoke a metaphysical, vital and also pictorial *gestus*. In that sense, we do not see Nana as a prostitute in the beginning, and in the end we do not either although we know that she was one. What interests Godard in prostitution is not the figure of Nana 'as already having become a typified character, but the moment when a woman like all the others, is driven, under certain circumstances, to prostitute herself; the moment of passage, when she is still the one that she was before, while at the same time being already different' (Bergala 2006: 102).

In Godard's words, *Vivre sa vie* 'is the portrait of a woman during a few months and it just so happens that in those months she prostitutes herself.' Godard refuses to turn this circumstantial activity into the female character's essence.

He asks the following question: what changes in a woman's consciousness and life when that frontier is crossed? (Bergala 2006: 102) We are not invited to observe a character thrown into a classical tragedy,

but to witness a woman in a contemporary situation living her life. Thus, prostitution is not an element that particularly defines her as a character in this fiction. Like other prostitutes of Godard's, she resists or seems aloof and detached in the sex scenes where she participates, looking out the window as she lends her body to men.

In *Vivre sa vie*, 'when a door opens into a room where the woman is supposed to be prostituting herself, it is always a living, sculptural and immobile tableau that is shown to viewers (as if these prostitutes' clients came to these places to satisfy primarily their *mise-en-scène's* ghosts, more than their crude sexual needs)' (Bergala 2006: 103).

Godard rejects the presence of sexual images in *Vivre sa vie*, which allows the reality of Nana's prostitution to go beyond eroticism and its visible actions to reach her being more profoundly – prostitution is a stage of learning in life. Prostitution's meaning in the film can be found, as pointed out by Tom Conley, in the term's etymological roots themselves – *prostatuere*: expose to the eyes, the loss of anonymity. There is a theological dimension here: this revelation results in the character's fall and transforms her into a kind of cinematic Christ.

The self-reflexive relationship that is established between Nana's movements as she lives her life, the thought of the body and the cinematic form is inspired, as we have noted earlier, by (Merleau-Ponty's) existentialist phenomenology. As a consequence, in the film all the meanings pass through the body.

Nana's presence, which is indistinguishable from the actress Anna Karina's, Godard's wife at the time, (especially in the last scene where Godard himself reads Edgar Allan Poe's 'Oval Portrait'),[7] is the subject of ethological inquiry – Can we go beyond appearances? Can we touch

---

7   What we see is also that: the study of a face by Godard's camera. A study that looks for the truth in the actress Anna Karina's face. In the end, the film makes it clear that it was also about intersecting with fiction – sometimes emancipating from it – a portrait of a painter/filmmaker's wife, made by him. In that sense, three planes meet in the film. Firstly, the fiction one, which deals with Nana's character. Secondly, the documentary/study about the actress Anna Karina's face as she plays Nana, which distances itself from fiction and Nana's story. Finally, encompassing the other two, is the film about the film, where cinema thinks itself and the very idea of cinema is taken apart. These three planes infiltrate into each other and the borders between them are diffuse.

a soul (essence)? It is also the object of plastic celebration – through the confrontation with cinema and painting (appearance), and by revisiting the division between appearance and essence (cf. Brenez 1998), action and thought. At the same time, the film strives to bring these together by embodying the character's journey and destiny, by turning the body and its postures into categories of the spirit, as Deleuze would say.

All her interiority pulsates on this exteriority – her face, mainly, but also her gestures and her body – that the film explores, showing the ambiguity present in the title – 'what I call mine, what is most personal to me, is also ultimately unknown, and, therefore, detached and impersonal at an ontological level' (Mathews 2006: 45). The film happens there, in that paradoxical mix of intimacy and detachment. In fact, the creation of a feeling of intimacy is ambivalent as we are constantly confronted with the inability to penetrate beyond the opaque surface of the words that are exchanged, of the expressions and actions that are manifested.

It is worth mentioning here the expression *donnez-moi un corps*, which is for Deleuze the formula of philosophical reversal: 'the body is no longer an obstacle that separates thought from itself, that which it has to overcome to reach thinking. It is, on the contrary, that which he plunges into or must plunge into to reach the unthought, that is, life.' It is not the case that, for the philosopher, 'the body thinks, but obstinate and stubborn, it forces us to think, and forces us to think what is concealed from thought, life.'

In this sense, for Deleuze, it is not a question anymore of 'making life appear before the categories of thought, but throwing thought into the categories of life.' The categories of life are, therefore, the body's attitudes and postures. 'We do not even know what a body can do, in its efforts and resistances,' writes Deleuze, quoting Spinoza. 'To think is to learn what a non-thinking body is capable of, its attitudes, its postures. It is through the body that cinema forms its alliance with the spirit, with thought. *Donnez-moi donc un corps* is first to mount the camera on an everyday body' (Deleuze 1985).

Brice Parain agrees that it is necessary to introduce the body. The philosopher observes that since the introduction of contingency by Leibniz – contingent truth alongside necessary truth, that is, everyday truth – we have to think with the limitations, the errors of life.

It is through the contingencies of her life as a prostitute, marked by the experience of silence and of the objectified body, that thought can emerge, not in terms of its contents but in terms of its own visibility – that allows us to read Montaigne's opening quote (his art of living) as a sentence about the urgency of experience, that is, the obligation and absolute need to pull life out of death.

On the other hand, in *Vivre sa vie* the narrative's usual causal sequence is, as we have seen, broken by the arbitrary decomposition of the story into twelve episodes – these episodes relate to one another serially and not through causality. In each series the film shows us that it thinks itself through both its characters and objects, and a collection of texts and references that serve as reflexive categories.[8] These compose a web of quotations and allusions that define Godard's morals and aesthetics of appropriation: Montaigne's epigraph; the essay about the chicken; the excerpt from a gossip magazine, read by a co-worker in the record store; the passage from Dreyer's *La passion de Jeanne d'Arc*; the shadow of Louise Brooks; Nana's name, which evokes Émile Zola's novel and Jean Renoir's film; the song; the dance music record; Brice Parain's reflection; Poe's text. There is no unified point of view, but a series of documents and descriptions that make up an unquantifiable body consisting of words, sounds and images. The traditional, unified interior discourse is replaced by free indirect discourse. These elements are quoted in the film in a way that preserves their difference. They might be used with analogical goals but they still keep their specificity, which means, for example, that the sentence about the chicken may clarify many planes in the film, not only Nana's story – *poule* in French also means prostitute – but also the film's program itself as we have mentioned. The film allows for relations between various terms to be established since these are not made to depend on identity relations. What

---

8   In Deleuzian terms, which could apply to any film by Godard: 'Each series refers to a way of seeing or speaking, which may be that of current opinion, operating through slogans, but equally that of a class, a sort, a typical character operating through thesis, hypothesis, paradox. Each series will be the way in which the author expresses himself indirectly in a sequence of images attributable to another or, conversely, the way in which something or someone is expressed indirectly in the vision of the author considered as other' (Deleuze 1989: 183).

we have here is the pinnacle of Rancière's 'sentence-image' within the 'aesthetic paradigm'.[9] Through the interaction of the various elements and a sensitivity to the gap between words and images, the wealth of connections that defines the 'sentence-image' in its acme of expressive saturation can materialize. Here the singularity of fragments, and its rupturing power, is interwoven with a discursive and reflexive density, which arises from the connections that those fragments establish between themselves at multiple levels through editing.

Deleuze hints at a similarity between Godard and Aristotle since he finds affinities with the philosopher's table of categories in the film-maker's table of montage. The fact that Godard resorts to categories in editing does not make them final answers, but problems that introduce reflection into the image itself. They are problematic or propositional functions. The question for each film is: what performs the function of categories or reflexive genres? For instance, Deleuze mentions that Parain fits into a reflexive category (that of language) towards which Nana moves (Deleuze 1985: 186).

In the conversation scene between Nana and Brice Parain, Parain reflects on the nature of language. Nana asks him why human beings must always talk and cannot live without words. He answers that there can be no life without thought and that in order to think one must talk, since speaking is the same as thinking, and thinking is the same as speaking. The question is not one of speaking compared to not speaking – even if sometimes one has to go through silence to speak well – but one of how to speak or think well. Parain anchors his position in the story of Porthos, one of the three musketeers, who has never thought in his life and who dies when he begins to think about how it is possible to put one foot before the other. Because he had to stop, to subtract himself from the course of life, the first time he

---

9   Rancière dates the 'sentence-image' back to Flaubert: 'if Flaubert 'doesn't see' in his sentences, it's because he writes in the age of clairvoyance (*voyance*) and the age of clairvoyance is precisely that in which a certain 'view/vision (*vue*)' has been lost, and saying and seeing have entered a space of community without distance, or correspondence' (Rancière 2003: 58). Thus, Flaubert uses the *gueuloir* ('one has to listen') and editing. The 'sentence-image' is not a sentence that allows meaning to be seen; it is not an image that carries meaning. It is something close to, though not totally interchangeable with, film editing – words, sounds and images arranged in such a way that they create a certain impact and/or meaning.

thought it killed him. Parain then concludes that the discovery of truth can be painful and that truth can never be reached without erring.

Naturally, this scene illustrates how important the reflection about language is for Godard, in general, and in this film, in particular. Brice Parain is a philosopher of language, occupying a space between pre-analytical philosophy and continental philosophy. For him, sentences, or propositions, are not separate from life. 'According to Sartre,' his work leaves us with the 'vertiginous feeling of the inexactness of language' and is the product of a man who 'words hurt and wants to heal,' 'who suffers for feeling that he is in *décalage* with language' (Fieschi 1962: 23; cf. Salanskis n.d.). His attempt is that of an unreconciled man who is looking for reconciliation. The same is true of Godard, who uses cinema to better understand life.

Central to Brice Parain's philosophy is the identity between thought and language, that is, the way in which the latter moulds and gives shape to the former. In fact, for the author, language almost becomes a religious experience since the truth of life is indistinguishable from the use of language. According to Parain, an instant of thought can only be grasped through words, and that is why to talk always implies the risk of lying. In this sense, lies too are part of our quest – we must pass through error to arrive at the truth, for instance when searching for the right word that one cannot find. Therefore, one cannot 'live in truth' without having recognised that truth is in everything, even in error (which means that errors and lies, in the above sense, are very similar). This emphasis on the importance of language, taken to the extreme, is at the heart of his critique of Blaise Pascal. For Parain, Pascal fails in trying to understand thought as a dialectical process when he identifies it with 'living in truth.' Parain does not believe that one could live directly in the truth unless one would adopt the silence of the medieval mystics, which is not possible in practice: the process of thought is inseparable from research, from a dialectical to and fro between truths and lies, changing one into the other – truths become lies (untruths) as soon as they constitute the premises for new truths. From his point of view, language is not something exterior to our being, a meta-level from which truth as a correspondence between meanings and the real world could be expressed. We cannot escape being inside language, it is our house, 'the house of being'. For

Parain truth in language is closely related to truth in life because not only does thought occur in life, with the servitudes and errors of life (cf. Parain 1953, 1972), but also it is inseparable from the elaboration of assumptions about life. These assumptions are not founded in logic, nor do they stem from factual observation, but come from a moral point of view towards life, from taking a stance towards it, which requires commitment. Assumptions are not factual states or states of life; they are promises: 'When I speak, I propose something to life' (Deleuze, 1984–1985: *Cours du 22 janvier* 1985). At the same time, speaking is not living and in order to talk, i.e. to think, one must have 'broken up with life,' that is, been through death, or through a kind of death. Nana mentions her difficulty in expressing herself as a result of her need to think, to use words to escape the states of life that she finds herself in and often oppress her. To this Parain answers, precisely, that renouncing life for a bit is the price to pay for speaking and thinking. He says: 'to live in speech, one must pass through the death of life without speech.' In this sense, to speak is almost like resurrecting to life, which means that speaking is a kind of ascetic exercise that makes us look at everyday life, or at too elementary a life, with detachment. To speak is to demand something from life, to interrupt it and escape from what is terrible about it. This means oscillating between silence, i.e. the moments when one has to 'live life,' and the moments when one 'gives orders to life' and demands something from it (while at the same time becoming involved in this demanding from life) (Deleuze, 1984–1985: *Cours du 22 janvier* 1985).

Conversely, in Godard, there is not an absolute identification between (spoken) language and thought. Like Nana, Godard distrusts words since words can betray you. Like her, the film has to pass through silence, through the silent observation of Nana's life in order to touch her truth and essence. Therefore, one has to question the meaning of words instead of passively accepting them. If Godard uses the dialogue between Nana and Brice Parain to question himself about how adequate words (thought) are to describe perceptions (of the real) and sensations, it is because he is looking for an answer to this problem in life, that is, inscribed in a situation, in a project. To find the right words, one has to work at it, that is, to suspend the legitimacy of dictionary definitions and language conventions. This is true

for cinema as well as for the learning of cinema. If in all of this one cannot stop hearing the words spoken by Parain in his dialogue with Nana, Godard, however, distances himself from the philosopher. In fact, he seems to believe in a manifestation of thought and truth that dispenses with language and involves reading through the faces and bodies in silence (like Nana thrown into the world and responsible for her actions – here we find phenomenology again). And cinema would be the place of that manifestation. Also from Godard are the words uttered by Nana: 'But why does one have to always talk? I think that sometimes we should just shut up and live in silence. It would be so much nicer to live without talking.'

How can we reconcile thought and action, i.e. life, without one cancelling the other? The character's concerns are also the film's and Godard's. Therefore, the last two episodes retroact on the film's different levels which we have briefly referred to. In the Brice Parain episode it is cinema that thinks itself through language and its ability to resurrect life, which would otherwise be condemned to silence and death, without a form on which to reflect itself. At the same time, in Nana's story, the conversation with Parain (in agreement with the idea of a thought that manifests itself directly in life and that cinema makes visible) determines that the character having acted and lived, by reflecting and talking about life, by separating life from thought, would have to die afterwards like in Porthos's story (from Alexandre Dumas's *Vingt ans après*). Read from this perspective, the episode no longer clarifies what brings cinema and language together but shows what separates them, what makes them different. It in turn ties in nicely with the oval portrait episode, which explicitly introduces, through the analogy between cinema and other arts, a reflection on the relationship between cinema and its ability to portray life,[10] to literally pull it out of death – the character's, but also the actress Anna Karina's.

---

10   Nana's itinerary allows Godard to live as if he were putting into his creation his own obstacles, his own difficulty, albeit transfigured. The difficulty of being is solved by Godard through the very being of cinema. By learning about cinema, he learnt about life. In an interview about *Vivre sa vie*, he says, 'It was cinema that made me want to make films, I didn't know anything about life, except through films, and my first efforts were films for cinephiles, the work of a cinema enthusiast. What I mean is that I didn't see things in relation to the world, life, History,

It is possible to recognize here the issue of cinematic ontology, inherited from André Bazin, as a critical realism that feeds on the (also Bazian) idea of an impure cinema. Bazin's influence, expressed as follows, can be felt in all the Nouvelle Vague filmmakers and crystallizes its modernity: cinema is modern when it refuses the old art of storytelling and asserts itself through what constitutes its specificity, meaning its ability to apprehend the spectacle of life without needing to resort to Aristotelian drama. Nonetheless, to directly apprehend life means, following Rancière's arguments, to apprehend it before language but after undoing the ties between world and language, the visible and the sayable, inherited from other arts and their narrative-representative conventions, and to find new ways to tie them together, rescuing those disjointed elements into a new non-unified configuration. In Godard, the direct capture of life goes hand in hand with the problematization of cinema itself in its relation to life. Due to its automatism of reproduction of the movement of the world, cinema – as a chaos of signs of variable signification (namely the ones produced by other arts and by cinema itself) – is the art that can best take in reality. These signs are simultaneously characterized by the image's autonomy and disruptive singularity, and by the potential for connection and arrangement found in the discursive logic. Hence Godard's confrontation with painting and literature, not in the sense of adapting them or interpreting their works but of using them as raw matter that cinema then re-works and re-links in a process that shows the possibility of new relationships with them and between them, life and history. In this way, cinema proposes itself as a new instrument of thought, before and beyond language.

In Éric Rohmer's words, 'Ontologically, film says something that the other arts don't say' (Rohmer 2004: 25). However, this idea, which distances cinema from older models taken, for instance, from literature and theatre (refusing the imitation of other arts and the narrative power of classic Hollywood films to favour a direct acceptance of the world instead), is inseparable in Godard from an inquiry into what it means to say something else. What follows is an investigation into

---

but in relation to cinema. Now I'm walking away from all of that' Cf. Fieschi (1962).

the way in which cinema says something else, that is, a profound reflection on language and on what makes cinema a non-discursive art different from others. This is achieved without falling into an essentialism incompatible with the practice of contamination of cinema by other arts.

References

Badiou, Alain (2013). *Cinema*. Trans. Susan Spitzer. Cambridge: Polity Press.

Benjamin, Walter (2003). 'Qu'est-ce que le théâtre épique?', in *Essais sur Brecht*. Trans. Philippe Ivernel. Paris: La Fabrique Éditions.

Bergala, Alain (2006). *Godard au travail*. Paris: Cahiers du Cinéma.

Brenez, Nicole (1998). *De la figure en générale et du corps en particulier. L'invention figurative au cinéma.* Louvain: De Boeck Université.

Delayahe, Michel (1965). 'Entre Terre et Ciel. Entretien avec Dreyer'. *Cahiers du Cinéma*, nº 170 (septembre): 15–34.

Deleuze, Gilles (1989). *Cinema 2: The Time-Image.* Trans. Hugh Tomlinson and Robert Galeta. Minneapolis: University of Minnesota Press.

Deleuze, Gilles (1985). 'Cours Cinéma/Pensée – octobre 1984/ juin 1985', La voix de Gilles Deleuze en ligne. <http://www2.univ-paris8.fr/deleuze/rubrique.php3?id_rubrique=17>

Didi-Huberman, Georges (2009). *Quand les images prennent position. L'oeil de l'histoire 1*. Paris: Les Éditions de Minuit.

Fieschi, Jean-André (1962). 'La difficulté d'être de Jean-Luc Godard'. *Cahiers du Cinéma*, nº 137 (novembre): 14–25.

Mathews, Peter (2006). 'The mandatory proxy'. *Biography*, 29, 1 (Winter): 43–53.

Merleau-Ponty, Maurice (1996). 'Le cinéma et la nouvelle psychologie', in *Sens et Non-sens*. Paris: Gallimard.

Narboni, Jean, & Tom Milne (eds.) (1986). *Godard on Godard*. New York: Da Capo Press.

Parain, Brice (1953). *Sur la dialectique*. Paris: Gallimard.

____ (1972). *Recherches sur la nature et les fonctions du langage*. Paris: Gallimard.

Rancière, Jacques (2003). *Le destin des images*. Paris: Éditons La fabrique.

Rohmer, Éric (2004). *Le Goût de la beauté*. Paris: Cahiers du cinéma.

Salanskis, Jean-Michel (s/d). *La dialectique de Brice Parain*.

Sontag, Susan (1994). 'Godard's *Vivre sa Vie*', in *Against Interpretation*. London: Vintage: 196–208

Sterritt, David (1999). 'My life to live', in *The films of Jean-Luc Godard. Seeing the invisible*. Cambridge: Cambridge University Press: 60–88.

Young, Colin (1963). 'Conventional-Unconventional'. *Film Quartely*. Vol. 17, nº1 (Autumn): 14–30.

'I'd rather be lucky than good.' Ethical variations in *Pickpocket* and *Match Point*

Susana Viegas

1. Ethics in cinema: An introduction

'The man who said "I'd rather be lucky than good" saw deeply into life. People are afraid to face how great a part of life is dependent on luck. It's scary to think so much is out of one's control. There are moments in a match when the ball hits the top of the net, and for a split second, it can either go forward or fall back. With a little luck, it goes forward, and you win. Or maybe it doesn't, and you lose.' *Match Point* starts with this off-camera statement by Chris Wilton (Jonathan Rhys-Meyers). Chris talks to the viewers (and to the film director) as if he were the Voice of God commentator. The movement of a ball hitting the top of a net in slow motion is complemented by the Italian operatic singer Enrico Caruso, who sings 'Una furtiva lagrima' from the Italian Opera *L'elisir d'amore* by Donizetti. Caruso's dramatic voice and the metaphor of a tennis match dictate the tone of the film.

What are the individual and social consequences of a rational decision that states that to be lucky is more important than to be good? Is this a practical and reasonable opposition? Can (good and bad) luck be an ethical principle and what is its role in one's attitudes and decisions? These are nothing but some of the ethical and moral questions we pose to ourselves every day. In a broader ethical scenario, we may say that we must choose the best way of life and we must decide the best way to choose it properly. But to what extent can film be a useful tool to scrutinize these problems in a philosophical manner?

Today, film may play quite a relevant role in the philosophical study of artistic images and aesthetic thought, but ethics is still a slightly unimportant section of contemporary philosophy of film. Even if

the panorama changed during the nineteen seventies and eighties with Stanley Cavell and Gilles Deleuze, ethics remains a secondary issue when compared with the ontological and epistemological questions. What is the ethical value of the aesthetic cinematic experience? How do we emotionally, existentially and ethically connect to the cinematic plot? Recently Robert Sinnerbrink defended a joint consideration of these different sides towards multi-layered artistic practice:

> Ethics in cinema (and cinema as ethics) represent much more than a minor branch of the philosophy of film. This is arguably the most culturally significant way in which cinema can be understood philosophically. It not only makes a contribution to our philosophical understanding of the world but enhances our ability to engage with complex forms of moral experiences.[1]

Thus does Sinnerbrink present us with the result of his research on the possibility and limits of cinematic ethics. He divided the ethical analysis of any film into three parts that, ideally, should be a collective reading of the three aspects in any film: 1) the ethics *within* cinematic representation, 2) the ethics *of* cinematic representation (for instance, questioning the decision of whether or not to show explicit scenes of violence and the viewer's response to these), and 3) the ethics of cinema as a cultural medium (as a medium for moral values, beliefs and ideological ideas). For this essay, I will restrict my analysis to the first one, analysing ethics in cinema via an analysis of the dilemmas, the themes and the film's plot. The use of verbal rhetoric in a film intrigue can be a good starting point to analyse the current debate around moral values because together with the visual images and visual (non-verbal) rhetoric they organise the film as a whole.

I will not consider film as a privileged medium for philosophising nor to illustrate someone's philosophy, whether the filmmaker or the main character in the film. Instead, I will look into these two films and read them as different fictional thought experiments that present us with vivid moral judgements and ethical dilemmas. If not fictionalised, some of them would not even be considered. But as fictions we certainly envisage and debate them in a different way to how we would

---

1   Sinnerbrink (2015: 9).

with a documentary film (maybe because we think that 'documentaries address *the* world in which we live rather than *a* world imagined by the filmmaker', as Bill Nichols says[2]). Besides that, it is important to stress that the general idea guiding this essay is not that of knowing if a work of art can be limited or criticised by moral values (is Riefenstahl's *Triumph of the Will/Triumph des Willens* (1935) a good work of art even if it disseminates Nazi propaganda?), or if taste and aesthetic judgement should contain other values besides purely aesthetic ones (is a film considered good only when it is not the medium for the dissemination of racist and hateful feelings?).[3] Let us recall here Berys Gaut's definition of ethicism: 'Ethicism is the thesis that the ethical assessment of attitudes manifested by works of art is a legitimate aspect of the aesthetic evaluation of those works.'[4] However, as Gaut also notes, it is most important to clearly distinguish the effects of a work of art that might morally corrupt its viewers from the effects of a particular character who represents dishonesty and immorality.[5] In this sense, *Pickpocket*, written and directed by Robert Bresson (1959), and *Match Point*, written and directed by Woody Allen (2005), are good works of art that do not intend to morally corrupt their viewers with negative, dishonest or hateful feelings, even if we may state with certainty that their infamous anti-hero protagonists are not ethical role models.

2. Gilles Deleuze: Ethics as ethology

When watching a film we do not just see it and hear it – we respond to it emotionally, existentially and ethically. We are not indifferent to what we see and hear in a fictional film. We may 'feel for' the character's feelings and actions (because we understand them), or 'feel with' the character's feelings and actions (we respond to the character's feelings but

---

2   Nichols (2001: xi).
3   Levinson (1998).
4   Gaut (1998: 182).
5   Gaut (1998: 188).

we do not share them, we do not feel the same).[6] As viewers, art gives us an ethical-existential dimension within which we analyse the moral and ethical qualities of any work of art. How can a fictional work as a film help us with the problem of the intertwining of ethics with aesthetics?

Gilles Deleuze had an immanent conception of ethics. Although he did not write a specific book on ethics, it is possible to find some remarks in the books he wrote with Félix Guattari and also in his two books on cinema. His ethics is defined as immanent and ethological,[7] as a divergent path from deontological ethics with normative and universal values. In this sense, Deleuze begins by opposing ethics (or moralism) to morality: 'morality presents us with a set of constraining rules of a special sort, ones that judge actions and intentions by considering them in relation to transcendent values (this is good, that's bad…); ethics is a set of optional rules that assess what we do, what we say, in relation to the ways of existing involved.'[8] For Ronald Bogue and D. N. Rodowick, to talk of a Deleuzian immanent ethics means to talk about cinematic ethics, especially the cinematic ethics of the time-image.[9] But when considering film, Deleuze does not have a single ethical view of it. Instead he presents us with two different approaches when considering pre- or post-Second World War films. In Deleuze's reading of film, films solve the existential crisis and the separation of the world during the post-war period. In the movement-image regime of the action-image as in the films directed by Eisenstein or John Ford, there was an identity between the human and the world. In this action-thought the subject of any film is not the individual and lonely hero, but the 'dividual', the collective subject. However, we can find in the action-image some films that 'vacillate' between the certainty of this regime towards the effectiveness and value of the actions in the world and some affection-images as in Dreyer and Bresson.[10] If in the movement-image regime there is a wide range of possibilities through the character's dilemmas and choices, in the time-image regime there is only one possible choice to make: 'to believe in this

---

6   Smith (2003: 90 ff).
7   Uhlmann (2011: 154–170).
8   Deleuze (1995: 100). Also in Deleuze (2009: 119).
9   Rodowick (2010: 97–114); Bogue (2007); Bogue (2010: 115–132).
10   King (2014: 59).

world.'[11] The link between the movement-image and the time-image re-
gimes is precisely in the modes of existence of the one who chooses.

A close reading of the two perspectives, although without any evo-
lutionary intention between them, delineates Deleuze's cinematic ethics.
'Affect as immanent evaluation, instead of judgement as transcendent
value: 'I love or I hate it' instead of 'I judge'.'[12] This immanent eval-
uation is a reminder of Deleuze's reading of Spinoza. The focus is on
what one can do, on what one might do, and not on what one should
do. What is the relationship between one's life as an authentic existence
and one's decisions and attitudes? Spinoza's ethology implies that there
is a singular connection between ethics and ontology with the way that
affects (the power to affect someone/something and also be affected by
it) shape a certain mode of existence. In this sense, first and foremost,
there is no divorce between human and the world, but there is only one
substance. This Spinozism perspective was the inspiration for the two
principles of the Deleuzian immanent conception of ethics: 1) to choose
to 'believe in this world' (and not in another, transcendental, world) is the
only authentic ethical choice to make after the crises of the action-image
and the breakdown of the sensory-motor regime (to see and hear replaces
to act/react). This world in which we should believe is the world of the
permanent becoming and change that manifests and expresses the power
of time; 2) this choice is based on the purely intensive criteria of an ethics
of affect that again has the power of time as a mechanism for change.

The different modes of existence express different types of forces,
either positive, negative or indifferent. For Deleuze, Michel in *Pickpocket*
represents the grey character – someone who is indifferent to his choices,
who is dominated by uncertainty about the good and evil of a normative
system. This does not mean that he should become good or evil, but
that by being indifferent to his spiritual alternatives he has become a
slave of his first choice. This means that he only had the opportunity to
choose once and since that moment, he has become entrapped by that
first choice and has lost the opportunity to choose again. The spiritual
choice of the grey character is, according to Deleuze, a false choice (as
it is a false choice to only choose good or evil).[13] On the contrary, an

---

11   Deleuze (2008: 172).
12   Deleuze (2008: 136).
13   Deleuze (2009: 117–118).

authentic choice would be the one made by someone who knows that a choice can be made and who is able to choose differently each time. A true choice is to restart each time and have the opportunity to choose again and not be entrapped by a choice made before. An authentic choice creates affects, immanent evaluation, and is defined by its power and not by its content. Deleuze opposes cinematographic expressionism to lyrical abstraction: if in Murnau there is a clear 'struggle of the spirit with darkness'[14], in Dreyer and in Bresson there is a spiritual choice, an alternative (either… or…). These alternatives are what Kierkegaard identifies as the three stages of the existential dialectic: the aesthetic, the ethical and the religious. According to this dialectical progression, the agent of the choice is in a permanent becoming. The question is how to choose what can be chosen. However, for Deleuze, without a belief in this world there are no reasons to choose properly. This will be the irrational point, as in Pascal's wager, and the illogical move of a leap of faith.[15]

Post-Second World War films do not register the world – they believe in this world as it is: 'The intolerable is no longer a serious injustice, but the permanent state of a daily banality.'[16] If before the historical fact of the Second World War, cinema were able to show us the different modes of existence of the one who chooses (through the choices that a character must make), in the post-war period this process was conveyed by the viewer's mode of existence. But how did we get to this point of not believing in this world? What disconnected humans from the world? To believe in this world decides both the original scepticism of Stanley Cavell's philosophy and the state of disbelief that Deleuze mentions.[17] For Cavell, cinema touches philosophy's main question: the epistemological scepticism about the existence of the world. According to Cavell, the photographic origin of cinema allows us to see and to think a world that is independent and exterior to any subjectivity. Thus, film refutes any metaphysical and subjective isolation. But in order to have an ethical experience from this aesthetic experience it is mandatory that a change take place: how to re-establish contact with the world?

14   Deleuze (2009: 115).
15   Deleuze (2009: 117).
16   Deleuze (2009: 164).
17   Rodowick (2010: 99)

To Cavell this ethical encounter transforms the sceptical subject (moral perfectionism), not the world. To Deleuze, cinema also explores new modes of existence in a way that interferes with the world. Thus the ethical dimension of film is for Cavell an 'ethics of self-transformation' and for Deleuze it is a 'belief in this world' through a leap of faith, as in Pascal. But if Cavell does not have a political reading of the moving image, Deleuze stresses the political vision with the idea of a 'people to come'. One of the most political and social (collective) consequences of this immanent conception of ethics is that becoming (other) is always possible. As a consequence, the agent does not have to carry the weight of his actions (in the sense that becoming comes from the past into the future) but can restart as something new, as something other (the becoming comes from the future and falls into the past).

The link between the human and the world is broken and it will never be fixed; it is impossible to link them again. However, the link can only be the subject of faith. We may decide not to give this leap of faith and never restore the connection, making any ethical view of the human's actions impossible. But we may also decide to believe in this world again. Just as in Pascal's wager, everything changes with this move. That way, the human (the viewer) can be connected with what he sees and hears.

3. Character Dilemma

According to Robert Sinnerbrink, cinema provides 'new ways of evoking and expressing *ethical experience*: not only emotional *engagement* facilitating moral sympathy and empathy but also emotional *estrangement* through which conflicting, clashing, or incompatible ideas, commitments, or beliefs can be revealed.'[18] Both *Pickpocket* (Robert Bresson, 1959) and *Match Point* (Woody Allen, 2005) are films with an obviously ethical dimension providing emotional estrangement. Their protagonists can hardly be considered heroes in the classical sense of the term, as role

---

18   Sinnerbrink (2015: 8).

models that do the right thing, that fight against injustice and that put the collective above the individual interest. They are closer to being negative heroes, not positive ones. In the classical films of the action-image driven by moral judgements, such as those by Griffith, Ford and Eisenstein, the true hero is the collective, the community that fights against evil believing that a better society is possible.[19] On the other hand, we should not mix up the film's plot, which might magnify the negative hero's actions, with the filmmaker's intention and, regarding this issue in particular, it is important to bring together the aesthetic qualities and the ethical values of films. How do we respond to the character's fate, such as in *Pickpocket* and in *Match Point*? What are the implications of Deleuzian immanent ethics, the necessity to have faith in this world, in these two movies? Both films raise important questions concerning decisions, values and consequences within different ethical backgrounds.

In *Match Point* Jonathan Rhys-Meyers plays Chris Wilton. Once a professional tennis player, he is now a tennis instructor at an exclusive club in London. During a night at the opera, *La Traviata*, he meets the sister of his new friend Tom Hewitt, Chloe (Emily Mortimer). He admits to her that tennis was an easy way to escape poverty in Ireland when he was a kid. Their different status is very evident in the film although his genuine interest in the arts is a way to overcome the monetary gap between them. Hard work is important, but for Chris luck plays an important role in life. But then, Chris falls in love with Nola Rice (Scarlett Johansson), Tom's fiancée. Eventually Tom breaks up with Nola but by chance Chris, who in the meantime has married Chloe, bumps into her at the Tate Modern. They restart their affair. Everything changes for him with a moment of bad luck when Nola gets pregnant when Chloe has been trying for years without success. However, although he feels guilty about the irony of the situation, Chris does not want to do the right thing and so plans to murder his lover. That way he would be able to continue with his own life project.

Chris dismisses an ethical life (to be good) in favour of what good luck can bring him. He commits two murders, which he disguises as a robbery (just like Raskolnikov in Dostoyevsky's *Crime and Punishment*), but he does not seem to be bothered by any feelings of guilt. He delivers his life and actions to the uncontrollable forces of luck (thus presenting

---

19   Deleuze (2009: 115).

us with a different version of the literary character of Raskolnikov, who had an obvious influence on Woody Allen's film). At the end Chris is the opposite of a man consumed by guilt, a recurrent murderer character in Allen's films such as Peter Lyman (Hugh Jackman) in *Scoop* (2006), Ian (Ewan McGregor) in *Cassandra's Dream* (2007) and Abe Lucas (Joaquin Phoenix) in *Irrational Man* (2015). In one of the first scenes in the film, Chris is reading *Crime and Punishment* but he switches his reading to *The Cambridge Companion to Dostoyevsky*. Shahrzad Siassi, who compares *Match Point* with a previous movie by Allen on the same topic, *Crime and Misdemeanours* (1989), argues that this switch is proof of Allen's existential atheism and nihilism[20]: unlike Raskolnikov, the protagonist does not believe in the (Christian) redemptive power of guilt, but instead prefers a more rational, expository, philosophical and theoretical analysis of these moral dilemmas (Fig. 1 and Fig. 2). The aleatory effects of luck replace the role of God in this godless and meaningless world.

Image 1 and 2: Chris prefers the Companion.

---

20   Siassi (2013: 133–146).

Chris decides that he would prefer to be lucky than good. Thus he commits a crime and expects that, with this decision, he can fulfil his life goal (to be married to a rich wife) without being caught. The film shows us the various circumstances in which luck prevails over bad luck thus guiding the consequences of his decision and somehow putting them beyond his control. He disguises the murder as a failed robbery and on several occasions he is not caught just by a fluke. When he throws the stolen jewellery into the River Thames, one ring hits the top of the railing (a similar visual composition to the ball in the film's beginning) and we imagine that the ring can either fall back and he will be caught, or fall forward into the river (Fig. 3 and Fig. 4). The ring falls back. Luckily for Wilton, someone finds the ring and picks it up. Later, the same person is involved in another crime related to drugs. He is caught with the ring and is formally accused of a crime of robbery and murder that he has not committed.

Image 3 and 4: If the ring falls back, Chris must lose.

In *Crime and Punishment*, Raskolnikov robs and murders a woman and her daughter but, consumed by his guilt, struggling with his most rational thoughts to justify his actions, he ends by confessing his crimes to the police and being exiled to Siberia. To be capable of killing someone (as a principle that justifies the crime) does not make that agent either superior or extraordinary. With that punishment, justice is established and Raskolnikov can finally restore his ethical (and religious) integrity. When the spectre of Mrs Eastby, Nola's next-door neighbour, appears in a dream scene to Chris, he says to her that 'the innocent are sometimes slain to make way for a grander scheme. You were collateral damage.' In order to justify killing his unborn child, he adds that Sophocles said: 'To never have been born may be the greatest boon of all.' (The dictum

slightly adapts the original one from the Chorus of the Greek trage-dy *Oedipus at Colonus.*) The next scene, with Detective Mike Banner awaking from his dream and figuring out how Chris Wilton committed the two crimes, is one of the few from which Chris is absent. At that point, we fear for him; his luck may be over. Although we recognize him as the infamous anti-hero of the story, we do not dislike him.

As for *Pickpocket* Martin La Salle plays Michel, a different ethical version of the main character in Dostoyevsky's *Crime and Punishment*. In *Pickpocket* Michel, unemployed, also commits a crime – pickpock-eting. The film also uses first person speech as a tool of persuasion. However, Michel's thoughts (including his fears and joys) are addressed to the viewers not as the Voice of God commentator, but in an intimate, confessional way. He does not see himself as an ordinary thief but as a pickpocket, a superior and extraordinary man, but he is willing to change his life and redeem himself. Michel prefers to have good luck than to decide his own fate. This decision is not a judgement. He steals because doing so he feels superior – Bresson never explains the char-acter's true intention in picking the pockets of others, why he never visits his dying mother, what his life project or 'grander scheme' is. His hands, his artful skills and his good luck make him feel superior to others: 'I was in control of the world,' declares Michel in voiceover at the beginning of the film. He knows it is wrong ('People can know that an act is ugly and still commit it,' he declares to Jeanne (Marika Green), his mother's neighbour, after she becomes aware that he is a pickpock-et), but he does it because he can do it well. In the final scene, Jeanne visits Michel in prison and the two become closer, which seems to be the precise opposite of the ending of *Match Point*. However, Jeanne and Michel are apart, divided by the jail bars. The final scene reveals that the world is fully restored at the end, even by the strange and par-adoxical lines when Michel finally understands that he is in love with Jeanne: 'Oh, Jeanne, to reach you at last, what a path I had to take.' But it is restored with irony: it may not be the expected Christian re-demption of the criminal who accepts his guilt and finds love. In this case, Michel's face remains blank as in another scene (a particularity of Bresson's style). It is a face without affects but where all affections can converge. According to Brian Price, the irony relies on the intertextual interpretation of the soundtrack of Jean-Baptiste Lully's *Atys*, an opera

possibly composed as a response to Louis XIV's attempt to clean Paris of all deviant behaviours (vagabonds, homosexuals, those who did not work, etc.): 'As Foucault has shown, the rise of the prison system in the time of Louis XIV was predicated on the latter's antipathy for beggars, sexual deviants, idlers, and those who challenged the authority of the state more generally.'[21] Thus, cinema is the art that makes visible the invisible everyday life of the common people, possibly of the infamous who are socially condemned and publicly judged: it may be a portrait of the life of chance of a murderer who leaves no clues behind, or of the daily training of the subtle movements of a pickpocket's hands.

Final remarks

Is Chris Wilton willing the eternal return of his actions (being a liar, being a killer)? Is he strong enough to deal with it? We certainly would share with Chris his own hopeful words: 'It would be fitting if I were apprehended… and punished. At least there would be some small sign of justice – some small measure of hope for the possibility of meaning.' At the end of the film, with Chris moving away from the rest of the family welcoming his newborn son, thinking that he is safe from the crimes he has committed but that at the same time he is trapped in a mere pointless existence, the viewer remains with a lost sense of meaning, of a world without any sign of justice and without any measure of hope. His grander scheme is accomplished but still there is no answer to his nihilism. He has no faith in this meaningless, intolerable and unjust world. From the classical Aristotelian conception of a good tragedy, we do not find a complex plot as there is no reversal moment (*peripateia*) that could free the character from his suffering and the viewers from their expectation of catharsis.

Is it possible to make an ethical judgement of Chris Wilton's initial statement ('I'd rather be lucky than good')? It seems to us that the film

---

21   Price (2011: 32).

is a way for Woody Allen to call into question this first premise. It is not without irony that Chris and his wife are portrayed. Allen is never sympathetic either to the couple in love or to the deceived wife, who is represented as being naïve. In the end, to be lucky is not a good, universal, ethical principle because it neutralizes all actions. Wilton's decisions about his actions will not become good just because he becomes unlucky and is arrested, and they do not become bad because he got lucky and escapes punishment. It does not matter if one's actions are good or bad: as long as one is lucky the actions are not qualified as good or bad. In this sense, good or bad luck is incompatible with ethics because it puts consequences beyond a person's control. Putting all one's faith in good and bad luck is a way of not being responsible for one's actions because it does not matter if what you decided to do is right or wrong as long as you were lucky – and in a deontological system to have luck is a totally arbitrary principle over any (im)moral choice. Chance just substitutes the role of God in this ethical understanding of actions. To be able to succeed with his ambitions (not be caught by the police for his crime and to marry his rich fiancée) cannot justify Chris' decision (to kill his lover, Nola).

An immanent perspective on ethics defends that someone's actions are driven by their power and qualities and not by a deontological orientation. That way, someone's decisions are not judged for fitting or not fitting the same deontological system that is at its origin. This is a fictional experiment that makes us rethink how we normally judge the other's decisions and the importance we give to them or the disappointment we feel. Someone's decisions express the qualities by which those actions are inscribed in the world and in the agent. Thus, not to believe in the existence of a link between the human and the world becomes a real problem and, as Deleuze states, to choose to choose is a way of restoring that gap. As we saw, Michel and Chris are cinematic examples of the grey character, the protagonist of uncertainty and indifference. They have made the first choice but then got trapped in the consequences that followed. But they certainly hope for some justice in the world. Chris never achieves it and is not capable of confessing his crimes to the police, accepting the universal injustice of the world and that 'so much is out of one's control'. He does not believe in inscribing his intensive actions in the world. Only Michel can accept the world as it is and he

acknowledges his crime by being arrested. That way, Michel knows that by choosing to believe some choices are no longer possible; he knows that becoming is always possible and a new mode of existence is possible. To be lucky makes Chris and Michel feel superior, a feeling they would not have had if they had decided to do the right thing, to be good, because in this case, eventually, they might very well have had to deal with the unlucky. They anticipate the risks they are taking by not being good and they calculate the best way of not being caught and punished for their decisions, but even so they decide to go that way. Ethics does not apply to a pure luck-driven decision but, however, it can apply to Chris' initial statement ('I'd rather be lucky than good'). That statement makes us rethink which values we take for granted in our everyday life by exploring fictional different and changeling modes of existence.

References

Bogue, Ronald (2007). *Deleuze's Way: Essays in Transverse Ethics and Aesthetics*. Abingdon: Ashgate.

Bogue, Ronald (2010). 'To Choose to Choose – to Believe in This World', in D. N. Rodowick (ed.), *Afterimages of Gilles Deleuze's Film Philosophy*. Minneapolis and London: University of Minnesota Press: 115–132.

Deleuze, Gilles (1995). *Negotiations 1972–1990*. Trans. by Martin Joughin. New York: Columbia University Press.

Deleuze, Gilles (2008). *Cinema 2*. Trans. Hugh Tomlinson and Robert Galeta. London: Continuum.

Deleuze, Gilles (2009). *Cinema 1*. Trans. Hugh Tomlinson and Barbara Habberjam. London: Continuum.

Gaut, Berys (1998). 'The ethical criticism of art', in Levinson (1998): 182–203.

King, Alasdair (2014). 'Fault Lines: Deleuze, Cinema, and the Ethical Landscape', in Jinhee Choi and Mattias Frey (eds), *Cine-Ethics: Ethical Dimensions of Film Theory, Practice, and Spectatorship*. Abingdon and New York: Routledge, 2014: 57–75.

Levinson, Jerrold (ed.) (1998). *Aesthetics and Ethics: Essays at the Intersection*. Cambridge: Cambridge University Press.

Nichols, Bill (2001). *Introduction to Documentary*. Bloomington: Indiana University Press.

Price, Brian (2011). *Neither God Nor Master: Robert Bresson and Radical Politics* Minneapolis: University of Minnesota Press.

D. N. Rodowick (2010). 'The World, Time', in D. N. Rodowick (ed.), *Afterimages of Gilles Deleuze's Film Philosophy*. Minneapolis and London: University of Minnesota Press: 97–114.

Siassi, Shahrzad (2013). *Forgiveness in Intimate Relationships: A Psychoanalytic Perspective*. London: Karnac Books.

Sinnerbrink, Robert (2015). *Cinematic Ethics: Exploring Ethical Experience through Film*. London: Routledge.

Smith, Daniel W. (2007). 'Deleuze and the Question of Desire: Toward an Immanent Theory of Ethics', *Parrhesia* 2: 66–78.

Smith, Daniel (2012). 'The Place of Ethics in Deleuze's Philosophy: Three Questions of Immanence,' in *Essays on Deleuze*. Edinburgh: Edinburgh University Press: 146–159.

Uhlmann, Anthony (2011). 'Deleuze, Ethics, Ethology and Art', in Daniel Smith and Nathan Jun (eds.), *Deleuze and Ethics*. Edinburgh: Edinburgh University Press: 154–170.

Smith, Greg M. (2003). *Film Structure and the Emotion System*. Cambridge: Cambridge University Press.

List of Contributors/Affiliations

António Marques
IFILNOVA, NOVA School of Social Sciences and Humanities

Erich Rast
IFILNOVA, NOVA School of Social Sciences and Humanities

Bartholomew Ryan
IFILNOVA, NOVA School of Social Sciences and Humanities

Claudio Rozzoni
IFILNOVA, NOVA School of Social Sciences and Humanities

Diana Soeiro
IFILNOVA, NOVA School of Social Sciences and Humanities

Dina Mendonça
IFILNOVA, NOVA School of Social Sciences and Humanities

Francesco Pallante
Università di Torino

Giovanni Damele
IFILNOVA, NOVA School of Social Sciences and Humanities

Gabriele De Angelis
IFILNOVA, NOVA School of Social Sciences and Humanities

Jamie Buckland
IFILNOVA, NOVA School of Social Sciences and Humanities

João Sàágua
IFILNOVA, NOVA School of Social Sciences and Humanities

Luís Sousa
IFILNOVA, NOVA School of Social Sciences and Humanities

Maria Irene Aparício
IFILNOVA, NOVA School of Social Sciences and Humanities

Marta Faustino
IFILNOVA, NOVA School of Social Sciences and Humanities

Michael D. Baumtrog
IFILNOVA, NOVA School of Social Sciences and Humanities

Nélio Conceição
IFILNOVA, NOVA School of Social Sciences and Humanities

Nuno Fonseca
IFILNOVA, NOVA School of Social Sciences and Humanities

Paolo Stellino
IFILNOVA, NOVA School of Social Sciences and Humanities

Pietro Gori
IFILNOVA, NOVA School of Social Sciences and Humanities

Regina Queiroz
IFILNOVA, NOVA School of Social Sciences and Humanities

Susana Cadilha
IFILNOVA, NOVA School of Social Sciences and Humanities

Susana Nascimento Duarte
IFILNOVA, NOVA School of Social Sciences and Humanities

Susana Viegas
IFILNOVA, NOVA School of Social Sciences and Humanities

LISBON PHILOSOPHICAL STUDIES
uses of language in interdisciplinary fields

A Publication from the Institute of Philosophy of Language at the New University of Lisbon

Lisbon Philosophical Studies – Uses of Language in Interdisciplinary Fields is the book series of the Institute of Philosophy of Language at the New University of Lisbon. Its aim is the publication of high-quality monographs, edited collections and conference proceedings in areas related to the philosophy of language, such as aesthetics, argumentation theory, epistemology, ethics, logic, philosophy of mind and political philosophy. The purpose of the series is to reflect the activities of the Institute as well as contemporary research in these areas, encouraging the interchange of arguments and ideas between philosophy and other disciplines.

Vol.   5 João Fonseca & Jorge Gonçalves (eds)
            Philosophical Perspectives on the Self.
            2015. ISBN 978-3-0343-1402-2.

Vol.   6 Paolo Stellino
            Nietzsche and Dostoevsky. On the Verge of Nihilism.
            2015. ISBN 978-3-0343-1670-5.

Vol.   7 António Marques & João Sáàgua (eds)
            Essays on Values and Practical Rationality –Ethical and Aesthetical Dimensions.
            2018. ISBN 978-3-0343-3058-9.