Normativity, moral realism, and unmasking explanations[1]

Josep CORBÍ

BIBLID [0495-4548 (2004) 19: 50; pp. 155-172]

ABSTRACT. Moral Projectivism must be able to specify under what conditions a certain inner response counts as a *moral* response. I argue, however, that moral projectivists cannot coherently do so because they *must* assume that there are moral properties in the world in order to fix the content of our moral judgements. To show this, I develop a number of arguments against moral dispositionalism, which is, nowadays, the most promising version of moral projectivism. In this context, I call into question both David Lewis' dispositionalist account of colour and Chistine Korsgaard's procedural realism.

Keywords: normativity, moral subjectivism, projectivism, dispositionalism, moral realism, explanation, morality.

I. Normativity and unmasking explanations

1. The individuation of a psychological state as a belief involves the satisfaction of some demands that are usually regarded as *normative*. The set of psychological states that constitute an agent's beliefs are to be individuated in such a way that they tend to form a coherent set and track variations in the world. If, for instance, an agent believes that his watch is on the table because he can see it, then if someone removed the watch from the table while the agent is looking at it, then he would stop believing that his watch is on the table. If, in those circumstances, the agent insisted that he still believes that the watch is on the table, we would not take him seriously, we would just think that he is joking or, perhaps, trying to make a desperate philosophical point.[2] And something similar goes for the coherence requirement.

So, there is a *first* sense in which beliefs involve *normativity*, namely: an agent's beliefs must be individuated in such a way that their coherence and truth is maximized. But there is also a *second* sense: when a particular belief is presented as true and coherent with other beliefs, we are enhancing the value of that particular belief. We can thus say that the individuation of a belief involves normativity at least in this double sense. There are, however, some instances of

(1a) *S* believes that *p*

where a *further* normative element is also present, namely, those in which the content of *p* involves some normative or evaluative concept, like in the following cases:

(1b) *S* believes that to *A* is morally wrong

---

[1] This paper has benefited from comments by participants in the *Workshop on Mind and Language* (Bologna, October 15-18, 2003), in the *Weekly Seminar of the Phronesis Group* and in the *Seminar on Normativity* (Granada, February 12-13, 2004). I must also thank Dan López de Sa and Marta Moreno for careful discussion of several aspects of this paper.

[2] Cf. Peacocke 1993; and Corbí & Prades 2000, ch. 6.

(1c) *S* believes that to *A* is acting cowardly (or generously, or honestly, ...)[3]

We see then that the individuation of moral judgements or beliefs involves not only the standard double sense of normativity, but an additional normative element that is present in the content of *p*, and this I will refer as *content-normativity* or *content-evaluativity*.[4]

2. Normativity (or evaluativity), in these three senses, *seems* to conflict with what natural sciences tell us about how the world is independent of us. From this viewpoint, the world consists of a number of causal processes which are not in themselves either coherent or incoherent, true or false, correct or incorrect, good or bad, generous or honest, beautiful or ugly. The identity of such causal processes is assumed to be fixed independently of our research practices, and the goal of such practices is precisely to track such independently individuated causal processes. The world as it is independently of us contains no evaluative feature and, therefore, gives us no norm for our action and no criteria to assess our psychological states (why should, for instance, truth be more valuable than falsehood?)[5] Values and norms should rather be interpreted as part of our response to certain non-evaluative, non-normative features of the world. For there is no norm and no value in the world as it is independently of us.

This line of reasoning presupposes what we may call, according to Barry Stroud,[6] the *bipartite image* of our conception of the world. In that image, our conception of the world is the result of the combination of two independent elements: on the one hand, the properties of the world as it is in itself, independently of how we are and the views that we have about it; and, on the other hand, our psychological properties ('independent' in the following sense: the identity conditions of these two elements are independently fixed),

$$CW = W + H$$

where '*CW* ', stands for 'our conception of the world', '*W* ' for 'the world as it is in itself' and '*H* ' for 'our psychological properties'.

---

[3] I do not think that moral judgements are beliefs, at least in the Humean sense. Yet, for the sake of argument, I will grant that moral judgements are, *prima facie*, beliefs in the sense that they aim at tracking the world or, at least, at being correct. For my stance about this issue, cf. Corbí (ms1).

[4] Emotivists contend that moral judgements do not *really* express a belief with a certain normative content. I do not think, however, that emotivist can yield a satisfactory analysis of moral judgements. For, in my view, the same line of objection that I will develop against the moral dispositionalist, also applies to the emotivist.

[5] Within this framework, we could surely provide an *instrumental* justification for the value of truth, but not a justification for the value with regard to which pursuing truth is a good instrument. Besides, why should we value what is instrumentally valuable?

[6] Cf. Stroud 2000, ch. 1.

In any case, if natural sciences have to provide a *full* conception of the world they must not only describe how the world is, independently of us, but also *explain* how it is that there are creatures in that world that have values and follow norms.[7]

The latter issue arises as soon as we acknowledge that agents have beliefs since the individuation of such psychological facts already involves some normative elements. And it may sound weird that there could be entities in the natural world that could only be individuated that way. There have been several attempts to bridge this particular gap between the normative and the non-normative: typically, by providing non-normative *sufficient* conditions for the tokening of beliefs. This is not the problem I want to address in this paper. My purpose is, instead, to focus on content-normativity. A full conception of the world must explain how it is that there are creatures with content-evaluative beliefs given that there are no moral *W*-facts (i.e., moral facts in *W*) and, indeed, such an explanation must be carried out without assuming that there are such facts. According to Stroud,[8] we may refer to this sort of explanation as an *unmasking explanation* insofar as they aim at explaining why such beliefs are always false and also why, despite such a fact, agents tend to acquire them.

3. *Moral Projectivism* takes the bipartite image for granted. There are different versions of that projectivist view, but all of them contend that moral properties are not among the properties of *W* and aim to unmaskingly explain moral beliefs. The overall projectivist idea is that the external world does not include any evaluative or normative property and, whenever an agent is in a psychological state like (1b) and (1c), he is just *projecting* upon that world his *inner response* to some *non-evaluative, non-normative features*.[9] If the unmasking explanation sketched by the Moral Projectivist is to succeed, he must be able to individuate the agent's *inner response* (which allegedly constitutes the *content* projected upon the world), as well as the specific relation called *'projection'*,[10] without relying on the existence of moral *W*-facts.

---

[7] For a discussion of the notion of an absolute conception of the world that lies behind this approach, cf. Nagel 1986, Putnam 1992, Stroud 2000 and Williams 1979, 1985, 2000.

[8] Cf. Stroud 2000, ch. 4. What Stroud's line of reasoning (and mine), show is *not* that I must assume that *p* in order to conclude that not-*p* (this is not incoherent in any relevant sense), but *instead* that I must assume that *p* in order to fix the content of *p*.

[9] In principle, the inner response at issue could also be a response to some *alleged* evaluative or normative inner feature. There is, however, a crucial difference between the fact that I value *p* (this is just a psychological fact) and the fact that *p* (even if *p* is a psychological fact) is valuable. For the Projectivist there are not facts of the latter kind except in the sense that facts of the former kind satisfy some *independent* principles or constraints.

[10] Apparently, Dispositionalist Theories do not need to talk of projection. Once we understand that moral properties are response-dependent properties, we needn't claim that these are properties of actions or situations in a response-independent way. But, still, Dispositionalist Theorist must accept that, before being supposedly convinced by their account, people used to believe that moral properties were response-independent and, therefore, such a theorist must account for the content of such beliefs and, at this stage, what else but projection could do the job for them?

In the coming sections, I will argue that moral projectivism cannot be coherently thought, that moral projectivists *must* assume that there are moral properties in the world (in the sense of the bipartite image, which is the only one available to them) in order to fix the content of the moral judgements that figure within our conception of the world.

It does not follow that, in order to defend a realist stance, *I must assume* that moral properties exist in the world in the sense defined by the bipartite image. For I regard my line of argument as a *reductio* of that image. For someone who is trapped in that image, the only alternative to moral projectivism is moral realism construed as claiming that moral properties are properties of the world in the sense fixed by the bipartite image, that is, as properties of *W*. But, once that image is dropped, the claims of the moral realist must be interpreted differently.

The structure of this paper goes as follows. As we have seen, the Moral Projectivist must be able to specify under what conditions a certain inner response counts as a *moral* response. I intend to argue that a proper elucidation of such conditions will call into question the possibility of coherently reaching a moral projectivist stance. To this purpose, I will initially focus on what I judge to be the most promising version of moral projectivism, namely, moral dispositionalism. In this respect, I will, firstly, distinguish between the individuating conditions of two features, namely, 'being nauseating' and 'being humiliating'. Secondly, I will challenge David Lewis' dispositionalist account of colour and argue that moral dispositionalism faces the same difficulties as Lewis' approach.[11] In the two last sections, I will present Korsgaard's procedural realism as a failed attempt to overcome those worries.

II. Troubles with Moral Dispositionalism

5. The individuation of moral features must meet some constraints which, features like 'being nauseating' or 'being frightening' need not. And such constraints have to do with the sort of limit that Isaiah Berlin detects regarding what may count as a *different* value:

> "Forms of life differ. Ends, moral principles, are many. But not indefinitely many: they must be within the human horizon. If they are not, they are outside the human sphere. If I find men who

---

This is quite clear in the case of colour terms (and, something similar, goes for moral terms). For the fact that colour properties are just secondary is just known *a posteriori*, it depends on the truth of a certain conception of the world. Hence such a fact cannot form a part of the content of the belief before that conception of the world arose. It might form a part of the content of such beliefs that colours *could* be secondary properties, but this holds for shape beliefs as well.

How does the Dispositionalist individuation of the content of our colour or axiological beliefs relate to the content of such beliefs before the disenchanted conception of the world? It does not sound like an analysis of such content for the reasons I have mentioned. Perhaps, it should be expressed like this: the most we can retain of the old beliefs. How is this continuum individuated? Some reasons can be mentioned, but, in the end, what matters is the perception of the continuum itself.

[11] I have elaborated this point in discussion with D. López de Sa (*cf.* López de Sa 2003) and Marta Moreno.

> worship trees, not because they are symbols of fertility or because they are divine, with a mysterious life and powers of their own, or because this grove is sacred to Athena —but only because they are made of wood; and if when I ask them why they worship wood they say 'Because it is wood' and give no other answer; then I do not know what they mean. If they are human, they are not beings with whom I can communicate —there is a real barrier. They are not human for me. I cannot even call their values subjective if I cannot conceive what it would be like to pursue such a life."[12]

I will use Berlin's quotation to compare the conditions under which an object or action can be identified as red, humiliating, or nauseating.

It is clear that the property 'being nauseating' is quite flexible regarding variations both in the *object* that may be identified as nauseating and in the *subject* that may have the suitable experience. For not only could *any object* be nauseating to some particular sensitive being, but also it is perfectly intelligible (even if rare) that a given object may be nauseating to that being on Mondays, Wednesdays, and Fridays, but not the rest of the week. In other words, the identification of an object as nauseating for someone does not require a certain future or past pattern of response on their side. The deep grammar of nauseating is 'being nauseating for someone at some moment'.

But this is not so for the property 'being humiliating'. It is true that, for any particular action that we now judge humiliating, we may imagine *situations* in which we would not deem it such or *people* who would not regard it as humiliating even in the present conditions. The problem is that to understand these *variations* we need some *sort* of explanation, a narrative. For instance, we cannot simply say that carrying the Star of David is humiliating whereas carrying the swastika is not. We need further details, we need to know that, even if the Star of David is a Jewish symbol, the fact that the Nazis imposed it as a stigma made the act of carrying it humiliating. But, of course, we can easily work out a context in which someone is proud of carrying the Star of David and ashamed of wearing the swastika.[13]

This suggests a significant flexibility as to what particular actions may be humiliating. Still, to render that flexibility intelligible, we need to tell a story that connects such actions to some morally relevant facts. This constraint is, nevertheless, absent in the case 'being nauseating': variations are not constrained by the availability of any such story.[14]

Constraints also apply to the sort of flexibility that can be tolerated regarding the nature of a moral subject. It is true that Germans Jews might have found it humiliating to carry the Star of David every day except the Sabbath, but to understand that, we need to tell a story: suppose, for instance, that it were part of the Jewish tradition to

---

[12] Berlin 1958, p. 11-12.

[13] "... a long time ago, while the first intifada was still taking place, it dawned on me that Israel was run by people who carried the David's Star is if it was a swastika" (a Palestinian journalist).

[14] This doesn't mean that there is no constraint involved. Presumably, some *patterns of regularity* are required, but not the kind of story pointed out in the case of humiliating: variations in the latter case must be justified in terms of variations of some morally relevant features, and nothing similar is involved in the individuation of nauseating.

carry the Star of David on the Sabbath but the rest of the days it was forbidden. So, an action cannot coherently be individuated as humiliating if the alleged individuating response on the side of the subjects were allowed to vary arbitrarily, like, for instance, the nauseating response it allowed. Variations on the side of the subject, like on the side of the object, must be justified *in terms of variations of some morally relevant features.*

6. The previous considerations allow us to distinguish between *moral pluralism* and *moral relativism*. The latter would affirm that moral properties are as flexible as being nauseating is, while the former would deny that, and would allow moral properties the sort of constrained flexibility that I have just ascribed to being humiliating. It is in the light of this distinction, that I interpret Berlin's claim that the conceptual inevitability of the conflict of values (about which we will talk later on) entail a healthy pluralism but not wild relativism.

I must also say that the constraints upon the flexibility of our moral responses are quite *naturally* expressed in terms of what the morally relevant features of the situation are. To put it another way, the individuation of a response as the humiliating response, quite naturally appeals to the attribution to the action (and not to the response) of some other moral features. And this fact *seems* to favour a *realist* interpretation of such features and, in the end, of 'being humiliating'.

The Moral Projectivist might certainly retort that the distinction I have just drawn between 'being nauseating' and 'being humiliating' is insufficient to make of the latter a *real, objective* feature of an action. For I have acknowledged that any particular action that is presently perceived as humiliating, may be regarded in some other context as not so or even as enhancing one's pride. Hence, even if our pre-theoretical conception of moral terms involves attributing them to actions, the individuation of any such features does not escape the logic of response-dependent features. In any event, the Moral Projectivist must be able to fix the content of our moral beliefs without assuming that there are moral features in *W*, but I do not think that such a demand can be met. To this end, I intend to challenge what I take to be the most promising attempt to meet that demand, namely: moral dispositionalism.

7. Moral dispositionalism claims that the moral features of actions are response-dependent properties, so that 'being humiliating' has two different, but closely interconnected, meanings. As a property of actions, 'being humiliating' alludes to the capacity of certain actions or situations to provoke, under certain circumstances, a specific inner response in some sort of agents. But 'being humiliating' may also refer to the specific inner response that figures in the previous characterization. The crucial question is whether these two interconnected meanings of 'being humiliating' can be individuated without assuming that actions and situations may be humiliating in a realist sense. My conclusion will be that they cannot. To develop the discussion, I will initially focus on colour properties and, in particular, on the dispositionalist analysis of

such properties that David Lewis proposes in 'Naming colors'.[15] Quite naturally, Lewis begins by characterizing both red and the experience of red as follows:

(7a) "*Red* is the surface property of things which typically causes experiences of red in people who have such things before the eyes.

(7b) *Experience of red* is the inner state of people which is the typical effect of having red things before the eyes."[16]

And he quite straightforwardly acknowledges that this "pair of definitions are almost totally useless, by reason of circularity."[17] He also accepts that Carnap's manoeuvre to avoid circularity in terms of Ramsey sentence, is insufficient because it does not allow us to distinguish between red and yellow, and between the experience of red and the experience of yellow. To solve this problem, we need, as Lewis points out, some further claims like the following ones:

(7c) 'Red is the colour of pillar box'

(7d) "... A living instrument: magenta is the colour such that I am disposed to say 'magenta' if you point to it and ask 'What colour is that?'"[18]

Lewis raises some worries concerning the parochialism of (7c) which I will not consider here. Suppose then that

> the combination of facts (7a), (7b), and some (7c)-like and (7d)-like facts fix the properties red and experience of red.

Yet, the need to introduce (7c)-like facts in order to distinguish between red and yellow, seem to have some serious implications. Prima facie, (7c)-like facts ascribe colour features to objects in the world without mentioning any response on the side of the perceiver. Of course, the dispositionalist could reply that such ascriptions should be interpreted in the light of (7a). But, recall, (7c)-like facts were introduced to *add* some further content to (7a) and (7b), whereby the content of an (7c)-like fact cannot then reduce to

(7a*) A pillar box has the property of being an object whose surface typically causes experiences of red in people who have such things before the eyes.

(7b*) *Experience of red* is the inner state of people which is the typical effect of having a pillar box before the eyes.

For, otherwise, there is no way in which the fact that a pillar box is red and not yellow could have been established, since we would be stuck with the same sort of emptiness

---

[15] Cf. Lewis 1997.

[16] Lewis 1997, p. 327. I am leaving aside other worries as to how to express these biconditionals in such a way that they are true of colours and not of shapes. Cf., in this respect, García-Carpintero 2002; Pettit 1991; Johnston 1989, 1992, 1993, 1998; Stroud 2000; Wedgwood 1998; and Wright 2001.

[17] Lewis 1997, p. 327.

[18] Lewis 1997, p. 336.

that led Lewis to add (7c)-like and (7d)-like facts in the first place. Hence, it seems that if Lewis' manoeuvre is to be successful, then (7c)-like statements must possess *some additional content*, they cannot *only* be interpreted in the light of (7a) and (7b). To put it another way, the sense in which the claim that pillar boxes are red cannot be reduced to the truth of the response-dependent biconditional implicit in (7a).

Alternatively, suppose (7c) is construed as an *example* of a red object whose role it is to *contribute to fixing* the property which 'red' refers to. Let us then reflect on the conditions under which a pillar box can play such a role. First of all, at least as it stands, (7c) involves the previous mastery of the concept of colour, that a pillar box can only help to fix the content of 'red' if the concept of 'colour' has already been fixed. To be consistent, the dispositionalist ought to provide a response-dependent account of the concept of 'colour' itself and it is uncertain how this could be done. Secondly, it is clear that if pillar boxes are to be used as examples that help us to grasp a certain concept (or to fix a certain property), then our capacity to grasp the colour of a pillar box cannot be reduced to our capacity to grasp the truth (7a) plus the claim that a pillar box satisfies the definition of 'red' in (7a), since, as we have seen, this definition does not tell us whether the object is red or yellow, whereby if the only fact that we grasp is that a pillar box satisfies that fact, we are not yet grasping that a pillar box is red. The colour of a pillar box can only be used as an example to fix the content of 'red' if, in grasping the colour of that object, we grasp something more than its capacity to satisfy the disposition specified in (7a). It seems that what we grasp is that pillar boxes are red and not yellow, and that this fact cannot be apprehended by the response-dependent biconditional, but, the other way round, that such biconditionals only characterize colour properties if they presuppose our capacity to grasp some independent facts like the ones I have just mentioned.

And this seems enough to challenge a response-dependent approach to colours, insofar as such an approach claims that colours are *merely* response-dependent properties, namely, that the sense in which an object has a colour is *exhausted* by (7a). I must, finally, stress that, even if my line of reasoning were right, this would not show that response-dependent theorists are wrong, but only that one cannot coherently reach that stance.[19]

8. It is easy to see how the previous line of reasoning applies to 'being humiliating'. Suppose we (provisionally and quite clumsily) reconstruct the biconditionals concerning humiliating and feeling humiliated as follows

---

[19] At this stage, I am adopting the sort stance that Barry Stroud defends with regard to subjectivism about colours and even about scepticism with regard to the external world.

By the way, I do not think (7d) may come to our help unless we assume that intentionality can be naturalized or that inner experiences can be individuated independently of (7b). Yet, if we could assume that, then there is no need to appeal to (7b) and therefore Lewis' would be irrelevant.

(8a) *Humiliating* is an action that typically causes feelings of humiliation in the person (or persons) to whom it is directed.[20]

(8b) *A feeling of humiliation* is the inner state of people which is the typical effect of humiliating actions.

It quite obvious that this pair of definitions and the corresponding Ramsey sentence, is insufficient to distinguish humiliating from shameful or coward.[21] To solve this problem, we need some further claims like the following ones:

(8c) Humiliating is being a Jew and being forced to carry the Star of David under the Nazi regime.

(8d) A living instrument: humiliating is the action such that I am disposed to say 'yes' if you point to it and ask 'Is this action humiliating?'

Like in the colour case, the sense in which (8c) claims that it was humiliating that the Jews had to carry the Star of David, cannot reduce the truth of the response-dependent biconditional implicit in (8a). For then (8a) we would be prey to the same accusation of emptiness that led Lewis to posit (8c), and then we could no longer distinguish between an action being humiliating, shameful or cowardly.

And this seems enough to challenge a response-dependent treatment of thick moral properties, insofar as such a treatment claims that such properties are *merely* response-dependent properties, namely, that the sense in which an object has a colour or a thick moral property is *exhausted* by (8a) and (8a), respectively.[22] Once again, even if my line of reasoning is right, this will not show that such theorists are wrong, but only that one cannot coherently reach that stance.

This worry does not arise regarding 'being nauseating' precisely because such a property is not constrained by the sort of demand that the individuation of an action as humiliating involves. Undoubtedly, some *patterns of regularity* must be present in order to individuate an object as nauseating. The crucial question is if such patterns necessarily involve the ascription to the object of features of the same kind that we are trying to individuate (or, in other word, if some (7c)-like facts are necessarily required). The sort of variability that is allowed for nauseating reveals that the patterns of regularity required for its individuation, do not involve such an ascription and this is the reason why they are uncontroversially regarded as subjective. My conclusion is that precisely because such an ascription is unavoidable in the case of red and humiliating,

---

[20] I am leaving aside here the sense in which an action may be humiliating not because the agent who performs it feels humiliated but, on the contrary, because it is performed in order to cause such feelings in other people.

[21] I think that Lewis' dispositional theory of value cannot help in this respect. Even if we assume that *"Something of the appropriate category is a value if and only if we would be disposed, under ideal conditions, to value it"*(Lewis 1989, p. 68); this does not tell us how to distinguish aesthetical from moral values, that an action is humiliating instead of shameful, and so on.

[22] If I am right, then, even though in the case of secondary properties, response-dependent biconditionals were true in virtue of the essence of such properties, they do not *exhaust* their essence.

we cannot coherently think of them as subjective, as *merely* response-dependent or, in other words, that, in the process of fixing those concepts, we must at some stage assume that there are colour and moral facts that are independent of our response. Let us examine Korsgaard's procedural realism and show why this projectivist approach does not escape the overall worries that I have raised against moral dispositionalism.

III. A challenge to Korsgaard's Procedural Realism

9. There are several texts where Korsgaard clearly endorses *moral projectivism*. In her view, the *content* that our moral judgements project upon consists of the content of those of our impulses that pass a certain normativity test:

> "Morality is grounded in human nature. Obligations and values are projections of our own moral sentiments and dispositions... Each impulse as it offers itself to the will must pass a kind of test for normativity before we can adopt it as a reason for action. But the test it must pass is not the test of knowledge or truth. For Kant, like for Hume and Williams, thinks that morality is grounded in human nature, and that moral properties are projections of human dispositions. So the test is one of reflective endorsement."[23]

As it stands, it seems that the content of our impulses, of our human dispositions, is fixed independently of the fact that they pass or fail to pass the normativity test. After all, the normativity test must be applied to a content that is previously fixed. In any case, Korsgaard takes it that the content of a human disposition is moral if and only if it passes a certain normativity test, even though no moral notion is required to fix the content itself. As we shall see, she proposes the moral law as the key normative test that an impulse must pass in order to be regarded as moral. And her distinction between private and public reasons, as well as her notion of unification, might be construed as attempts to elucidate the precise content of the moral law, even if they were initially designed to ground the need to be moral.

10. If Korsgaard's strategy succeeded, she would be able to retain both the projectivist idea that moral properties are not features of W and the normativity of moral judgements. This intuition can also be expressed, as Korsgaard does, by distinguishing two kinds of *moral realism*:

> "The procedural moral realist thinks that there are answers to moral questions *because* there are correct procedures for arriving at them. But the substantive moral realist thinks that there are correct procedures for answering moral questions *because* there are moral truths or facts which exist independently of those procedures, and which those procedures track."[24]

---

[23] Korsgaard 1996, p. 91, italics are mine. See also: "The reflective endorsement method has its natural home in theories that reject realism and ground morality in human nature... They [the sentimentalists] explicitly rejected the realism of the rationalists, and argued that the moral value of actions and objects is a *projection* of human sentiments." (Korsgaard 1996, p. 50.)

[24] Korsgaard 1996, pp. 36-7. Similarly, she claims that "Procedural moral realism is the view that there are answers to moral questions; that is, that there are right and wrong ways to answer them. Substantive moral realism is the view that there are answers to moral questions *because* there are moral facts or truths." (Korsgaard 1996, p. 35.)

As it stands, this distinction between procedural and substantive moral realism may be regarded as trivially circular. For, in this quotation, 'substantive moral facts' are characterized as facts that are independent of any such procedures, while 'a procedure' is tacitly individuated as a deliberation process that does not appeal to any substantive moral facts. Obviously, in order to track the intuitions that lie behind projectivism, we need a stronger notion of 'procedure', we must impose some constraints upon the sort of fact to which a procedure appeals, even if in the quotation itself, no such constraint has been explicitly imposed. For, in the quoted text, substantive facts are just those that are not mentioned within the procedure and no constraint is mentioned about what sort of fact may figure within a procedure. To avoid this circularity, a projectivist constraint comes quite naturally to our minds: substantive facts are facts about W and a procedure cannot rely on the attribution of moral facts to W if it has to help us to individuate the projected moral content within the boundaries of projectivism. Thus, I propose to read the previous quotation as tacitly appealing to this conception of a procedure and the corresponding notion substantive facts. So, we can say

> *Procedural moral realism* (hereafter, 'procedural realism') allows for the existence of right and wrong answers to moral issues, but claims that such answers are determined by some procedures that, *in their application*, do not involve the appeal to any moral facts to W and whose *legitimacy* does not derive from their ability to track any such facts.
>
> *Substantive moral realism* (hereafter, 'substantive realism') claims that any procedure that might fix the right answer to a moral issue must appeal, *at some stage*, to moral facts in the world.

Notice that, to avoid emptiness, procedural realism has been characterized by relying on projectivism and the associated bipartite image of our conception of the world; whereas substantive realism is not so committed. So, that someone (like myself) could endorse substantive realism without understanding the attribution of moral properties to the world as the attribution of such properties to *W*. But, indeed, a projectivist (like Korsgaard herself) must understand such substantive moral facts as *W*-facts. This is why we may call the substantive realism to which she opposes, *substantive W-realism*.

Quite reasonably, Korsgaard thinks that she has good reason to reject substantive *W*-realism and this is why she is interested in defending procedural realism. I will argue, however, that only if we *assume* substantive realism, can we make sense of moral discourse or, more precisely, can the content of moral judgements be fixed. Following up from my previous remarks on moral dispositionalism, I will conclude that a projectivist, insofar as she is committed to the bipartite image, must assume substantive *W*-realism in order to make sense of moral discourse. A first consequence of this will be that procedural realism falls short of grasping the content of our moral judgements. A second implication is that, insofar substantive *W*-realism is utterly implausible, this may be regarded as *reductio* against projectivism and the associated bipartite image. An alternative picture of our conception of the world will emerge out of Korsgaard's vindication of the space of public reasons as the bedrock upon which moral experiences

and judgements are elaborated. I am quite sympathetic with this alternative approach, but, in my view, it clashes with Korsgaard's projectivism and her defence of procedural realism. But let us begin by exploring Korsgaard's procedural proposal.

11. Korsgaard distinguishes between the *categorical imperative* and the *moral law*:

> "Now I'm going to make a distinction that Kant doesn't make. I am going to call the law of acting only on maxims *you* can will be laws 'the categorical imperative'. And I am going to distinguish it from what I will call 'the moral law'. The moral law, in the Kantian system, is the law of what Kant calls the Kingdom of Ends, the republic of all rational beings. The moral law tells us to act only on maxims that *all rational beings* could agree to act on together in a workable cooperative system."[25]

In this view, the only constraint that the categorical imperative imposes is that the agent must choose a law, that the choice of a free will must have "the form of a law".[26] By contrast, the moral law involves a more stringent demand on what may count as a universal law, namely, one must take into consideration not only one's own will but that of other people. Korsgaard seeks to ground morality (that is, provide a response to 'Why should I be moral?') by showing, first, that the categorical imperative is constitutive of action; second, that the moral law is constitutive of interaction and, thirdly, that the agent's actions involve the interaction of his parts. It follows then that the moral law is also constitutive of action. Leaving aside Korsgaard's foundational project, I want to focus on her attempt to fix the content of the moral law and argue that she cannot fix it without assuming that there are moral *W*-facts.

12. Korsgaard's distinction between public and private reason, as well as the notion of unification, play a crucial role in her foundational project, but also in her attempt to determine the content of the moral law, to fix what may count as a law that can (morally) be willed to be universal. In the latter respect, we might say that Korsgaard must provide an account of the modality of that 'can' which, despite avoiding any moral W-facts, tracks our intuitions about what counts as a moral maxim.

Let us begin with the distinction between private and public reasons. Consider, following up Korsgaard's own example, that a lecturer and a student are trying to arrange an appointment. Suppose that the lecturer proposes to meet in the afternoon but the student replies that, at that time, he has a class. The lecturer may treat the student's remark just as a reason *for him*, as a sheer obstacle to have an appointment that afternoon, or, alternatively, she could acknowledge the normative force of the student's reason and includes it in a cooperative search of a time that might be convenient for both of them. Korsgaard claims that, in the latter case, the lecture regards the student's reason as a *public reason*, while, in the former one, the student's reason is treated just as a *private* one.

---

[25] Korsgaard 1996, p. 99, italics are mine.

[26] Korsgaard 1996, p. 98.

But what are the implications of this distinction for the modality of 'can' in the moral law? The overall intuition is that the modality of 'can' must be constrained by public reasons in at least two senses:

(C1) one *can* will a law to become universal *only if* the normative force of other people's reasons have been acknowledged, that is, only if their reasons have been treated as public reasons. And

(C2) it is only the law that comes out of assessing all these public reasons that one *can* will to become a universal law.

The content of the moral law is thus fixed by applying constraints (C1) and (C2) to the modality of 'can'. Such constraints involve that the moral agent treat other people's reasons as public, as having a normative force in a process of cooperative deliberation. My question is: how is it fixed the fact that such an acknowledgement has taken placed? I do not see how this could be done but by examining how that moral agent deliberates on particular issues, the relative weight that he attaches to different reasons, the particular decisions that he makes, the way he acts. In other words, I think we would need to consider the content of the reasons at stake and see whether he is giving them the appropriate weight. For instance, we could not say that the lecturer is treating the student's reasons as public if a trifling inconvenience on the former's side would overweight a most serious problem for the student. In that case, the lecturer would fail to acknowledge the normativity of the student's reasons, but this verdict couldn't be reached independently of the assessement of the student's reason. The moral law must then rely on *some independent means* to assess the normative force of a reason and determine whether the moral agent have really taken it into consideration in his process of deliberation or, to put it another way, whether moral agents are *really* involved in a process of *cooperative* deliberation, in a real process of *interaction*.[27]

It is easy to see, however, that those independent means must, inevitably appeal to substantive moral facts. Suppose, for instance, that the student's reason includes some moral thick concept like 'being humiliating'. How could the content of that reason be specified? I do not think that such a content can be fixed within the boundaries of projectivism and, to this effect, we have already seen the most promising projectivist

[27] One may be tempted to appeal, at this stage, to the notion of *agreement* and claim that we can will a law to become universal if we all *agree* on accepting that law (cf. Korsgaard 1996, p. 99). This proposal has the advantage that agreement can still be regarded as a merely procedural criterion, but it also confronts a number of serious difficulties. Firstly, not every actual agreement will do, but only those agreements that satisfy some normative criteria and the procedural realist is committed to provide some merely procedural criteria to distinguish relevant from irrelevant agreements, but, as we have seen, Korsgaard's argument for public reasons can hardly help in this respect. Secondly, it is obvious that such an agreement will only be relevant if it is achieved after an appropriate process of deliberation, but how would the agents deliberate? how would they assess the different public reasons? and, more importantly, what will the *content* of such reasons be? Of course, the content of the claim 'Wearing the Star of David is humiliating' cannot be 'we agree that wearing the Star of David is humiliating' if there are no independent means to fix the content upon which we agree.

account (namely, moral dispositionalism) fails, that projectivists cannot fix the content of the property 'being humiliating' without assuming that it is a property of *W*.

13. Korsgaard's notion of a unified agent will not be of much avail either. For, if we construe that notion in merely formal terms, there are many ways in which an agent may be unified[28] which we (Korsgaard included) would like to discard as a criterion of what counts as a moral response. Consider for instance, the tyrannical person who is obsessed with a particular goal in detriment of other parts of himself. If, on the contrary, we adopt a more demanding notion of unification, like Korsgaard does, so that only those agents that are unified *in certain ways* give rise to moral responses, then Korsgaard needs some further procedural criteria to fix those particular ways. Her proposal is that, in the privileged sort of unification,

> "... reason must rule in the soul, if the soul is to be capable of action, and it must rule according to its own principle, if the action is to be a good one, and not defective."[29]

This is, indeed, the case of the aristocratic person. My worry is again whether Korsgaard supplies any procedure to fix the fact that reason rules in the soul and rules according to its own principle. The procedures that she actually mentions goes back to the distinction between public and private reasons, so that the aristocratic person takes into account the claims of all her parts and does not discard any such claims as merely private. Moreover, Korsgaard argues that the person who is aristocratic, who is inwardly just, will also be outwardly just, will supposedly take other people's reasons into account. No matter what you think of this last transition, it is clear that Korsgaard's approach is still affected by my remarks in the preceding sections.

IV. Moral Projectivism and the space of public reasons

14. So far, I have tried to show that Korsgaard's procedural realism sinks insofar as it is construed as part of a projectivist approach. The main reason is that there is no way in which the content of our moral judgements and reasons could be fixed without assuming the existence of some substantive moral facts, which a projectivist must interpret as *W*-facts. This becomes particularly clear when we realize that the morality of a maxim, its ability to pass the normativity test of the moral law, must rely on some independent means to fix the normative force of a reason and, in the case of reasons whose content involve thick moral concepts, such means inevitably comprise of the reference to some substantive moral facts. Of course, I am not thereby denying that the moral law apprehends the form of all moral maxims, but just that it could help to carry on the project of the procedural realist. And here we come to the second tension, within Korsgaard's approach: the tension between her projectivism and a crucial step in her attempt to ground morality, namely, the claim that our being naturally

---

[28] "The tyrannical soul, as I've just said, is consistently ruled and unified, though it is not self-governed." (Korsgaard 2002, lec. V, p. 25.)

[29] Korsgaard 2002, lec. VI, p. 1.

placed in the space of public reasons is the bedrock of moral experiences and judgements.

15. Traditional moral philosophers tend to think that the task of grounding morality consists in showing why an agent has private reasons to take other people's private reasons into account[30] and, therefore, to acknowledge their normative force and treat them as public reasons. I will not discuss the different attempts to go from private to public reasons in order to ground morality, since (I agree with Korsgaard) all such attempts fail,[31] but just challenge the plausibility of her alternative approach if it is to be interpreted within the boundaries of procedural realism. Her alternative view rests on the following intuition:

> "The solution to these problems must be to show that reasons are not private, but public in their very essence."[32]

To put it another way, we needn't have private reasons to place ourselves within the space of public reasons; since the latter is a space that we naturally inhabit:

> "You can no more take the reasons of another to be mere pressure than you can take the language of another to be mere noise." (Korsgaard 1996, p. 143.)

and, similarly,

> "Human beings are social animals in a deep way. It is not just that we go for friendship or prefer to live in swarms or packs. The space of linguistic consciousness —the space in which meanings and reasons exist— is a space that we occupy together." (Korsgaard 1996, p. 145.)

As Korsgaard points out, substantive realism has no trouble in acknowledging the existence of public reason, of reasons whose normative force must not derive from some independent private reasons. For this realist stance assumes that there are objective moral facts, that is, facts that are objective features of a public world. But, as we know, Korsgaard wants to avoid substantive realism (which, given her projectivism, must be identified as substantive *W*-realism) because she thinks that it is plagued with difficulties, whereby she proposes to construe "publicity as shareability".[33] Although, in contrast with the traditional view, she insists that we needn't have any independent reason to share other people's reasons, to acknowledge the normative force of such

---

30 "What such neo-Kantian arguments... and the Hobbesian arguments have in common is this: both assume that an individual agent has private reasons, that is, reasons that have normative force for her, and they try to argue that those private reasons give the individual some reason to take the (private) reasons of other people into account." (Korsgaard 1996, p. 133)

31 "All of these objections have something in common. They are all ways of saying that private reasons will remain forever private, that the gap from private reasons to public ones cannot be bridged by argument. In one sense, this is just what we should expect. We cannot know what an argument *does* until after we know whether the reasons it employs are private or public." (Korsgaard 1996, p. 134.)

32 Korsgaard 1996, p. 135.

33 Korsgaard 1996, p. 135.

reasons because that is our natural attitude towards them. Korsgaard appeals to Wittgenstein's private language argument to support the latter claim.[34]

I see, however, a deep tension between Korsgaard's projectivism and her claim that our position in the space of public reasons is the bedrock upon which moral judgements and experiences are elaborated. To unveil that tension, let us first ask what *sort of fact* is the fact *F*, i.e., the fact that an agent is placed in the space of public reasons.

According to the projectivist, *F* is either a fact about *W* or about *H*. Of course, the projectivist could not admit that it is a fact about *W* because, according to her, there are no normative facts in *W*. But, on the other hand, I will argue that Korsgaard cannot coherently claim that *F* is a fact about *H*.

16. Part of what Wittgenstein's analysis of rule following shows, is that no fact about *H* fixes the meaning of words, that there is no way in which facts about *H* may fix the distinction between following a rule correctly or incorrectly, that, in order to make sense of meaning, we must go beyond *H* and assume that we are already placed in the linguistic space. In other words, linguistic facts cannot be derived from any facts either about *H* or about *W*. And something similar goes from any other normative facts, like the fact that other people's reasons have normative force. This is, indeed, the reading of Wittgenstein's analysis on which Korsgaard relies in order to claim that we are naturally placed in the space of public reasons, that there is no need (and no possibility) of going from private reasons (which might belong to *H*) to public reasons.

To avoid some misunderstanding, let us point out that, when an agent stands in the space of public reasons, the *content* of his thought could not be rightly characterized as

(i) 'I take reason *R* of agent *A* as having normative force *N* ',[35]

which is a fact with no normative import, but as

(ii) 'Reason *R* of agent *A* has normative force *N* '.

Wittgenstein's and Korsgaard's point could then be expressed as follows:

A negative claim: (ii)-like facts do not reduce to (i)-like facts, and

A positive claim: (ii)-like facts are the bedrock upon which moral experiences and judgements grow.

And my challenge to moral dispositionalism (and, in the end, to procedural realism) should be construed as an argument that backs up the negative claim, and suggests the positive one.

---

[34] Cf. Korsgaard 1996, pp. 136-138.

[35] The same goes for 'We take other people's reasons as having normative force', insofar as 'we' refers to a set of individuals.

17. Yet, if this turns out to be so, if fact *F* is neither a *W*-fact nor an *H*-fact, then Korsgaard should either give up the bipartite image (and, thereby, projectivism) because there is no room within it for what she claims to be the bedrock of morality (i.e., fact *F*); or, on the contrary, give up fact *F*. Suppose we give up the bipartite image (which is the option that I favour for reasons that I have not mentioned here)[36] then what room can we make for *F* and any other facts that can be elaborated out of *F*? Well, once we give up the image of our conception of the world that led us to claim that, appearances to the contrary, the world has no moral features; we could just go back to our previous view and claim that both the meaning of words and moral facts are facts of the world (the world, which is to be distinguished from *W*, insofar as my description of the world is not committed to the bipartite image). In fact, we have no other option if my argument against moral dispositionalism is correct, if the content of thick moral concepts cannot be fixed except by assuming that the world has moral properties.

Of course, moral facts, like meanings, would not exist if creatures with a certain form of life would not exist, but this circumstance by itself does not render moral facts more subjective (where 'subjective' here no longer means 'pertaining to *H* ') than meanings themselves. Needless to say, fact *F* itself, like the overall fact that words have meanings, is not properly speaking a fact of the world, but a fact that sets one of the boundaries of the world. In any case, once more we must recall that this is no more than a transcendental argument, that the result is simply that if we want to make sense of language and morality, we must give up the bipartite image and assume that meanings and moral features belong to the world (which, of course, is not *W*).

18. To recapitulate, I must say that, in my case against Korsgaard's procedural realism, I have pointed out two problems: (1) There is no way in which her appeal to the moral law, the distinction between private and public reasons, and the notion of unification, may allow us to coherently fix the content that, according to projectivism, our moral judgements and experiences project upon *W*. (2) The fact *F*, which, according to Korsgaard, is the bedrock upon which moral judgements and experiences make sense, can find no room within the bipartite image that projectivism presupposes. For it can be interpreted neither as *W*-fact nor as an *H*-fact.

REFERENCES

Berlin, I. (1958). *The crooked timber of humanity*. Princeton: Princeton University Press.
Corbí, J.E. (2003). *Un lugar para la moral*. Madrid: Antonio Machado Editores.
—— (MS1). "Moral Motivation, *Ceteris Paribus* Clauses, and Directions of Fit", manuscript.
—— and Prades, J.L. (2000). *Minds, causes and mechanisms*. Oxford: Blackwell.
Johnston, M. (1989). "Dispositional Theories of Value", *Proceedings of the Aristotelian Society*, suppl. vol 63, pp. 139-174.
—— (1992). "How to speak of the colors", *Philosophical Studies* 68/3, pp. 221-263.

[36] Cf. Corbí and Prades 2000; and Corbí 2003.

—— (1993). "Objectivity Refigured: Pragmatism without Verificantionism", in J. Haldane and C. Wright (eds.), *Reality, Representation and Projection*. New York: Oxford University Press.
—— (1998). "Are manifest properties response-dependent properties", *The Monist* 81, pp. 3-43.
Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.
—— (2002), "Self-Constitution: Action, identity, and integrity", *The Locke Lectures*, http://www.people.fas.harvard.edu/~korsgaar/#Locke%20Lectures
Lewis, D. (1989). "Dispositional Theories of Value", *Proceedings of the Aristotelian Society*, suppl. vol. 63, 113-38. [Reprinted in D. Lewis (2000). *Papers in Ethics and Social Philosophy*. Cambridge (Mass.): Cambridge University Press, from which I quote.]
—— (1997). "Naming the Colors", *The Australasian Journal of Philosophy*, 75/3, pp. 325-42.
López de Sa, D. (2003). *Response-Dependencies: Colors and Values*. Barcelona: Ph.D. Dissertation.
Nagel, T. (1986). *The view from nowhere*. Nueva York: Oxford University Press.
Pettit, P. (1991). "Realism and Response-Dependence", *Mind* 100, pp. 587-626.
Putnam, H. (1992). *Renewing philosophy*. Cambridge: Cambridge University Press.
Stroud, B. (2000). *The Quest for Reality: Subjectivism and the Metaphysics of Colour*. Oxford: Oxford University Press
Wedgwood, R. (1998), "The Essence of Response-Dependence", *European Review of Philosophy 3: Response-Dependence*, R. Casati & Ch. Tappolet (eds.), CSLI Publications, Standford (Cal.), pp. 31-54.
Williams, B. (1979). *Descartes: The Project of Pure Enquiry*. London: Penguin.
—— (1985), *Ethics and the Limits of Philosophy*. Cambridge (Mass.): Harvard University Press.
—— (2000), "Philosophy as a Humanistic Discipline", *Philosophy* 75, pp. 477-496.
Wright, C. (1992). *Truth and Objectivity*. Cambridge (Mass.): Harvard University Press.
—— (2001). "On Being in a Quandary", *Mind* 110/437, pp. 45-98.

Josep E. CORBÍ is a lecturer at the University of Valencia (Spain). He is, together with Josep L. Prades, co-author of the book *Minds, Causes and Mechanism. A Case against Physicalism* (Blackwell Publishers, Oxford, 2000). He has also recently published *Un lugar para la moral* (Antonio Machado Editores, Madrid, 2003) and several papers on the philosophy of mind, epistemology and metaethics.

ADDRESS: Departamento de Metafísica y Teoría del Conocimiento. Facultad de Filosofía y Ciencias de la Educación. Univ. de Valencia. Avda. Blasco Ibañez, 30. 46010-Valencia. E-mail: Josep.Corbi@uv.es.