

## ***The Case against Evaluative Realism\****

Dan López de Sa

Universitat de Barcelona

[dlopezdesa@ub.edu](mailto:dlopezdesa@ub.edu)

FIRST DRAFT (20/10/03) – COMMENTS VERY WELCOME!

In this paper I want to present a case against evaluative realism. The considerations I will submit will not constitute a refutation of evaluative realism, given that in the central points they dwell on intuitions that if sound would support rejecting realism quite more directly, so that as arguments they might be accused of begging the question at stake. For the better or worse, I think that a stronger case against evaluative realism – nor for evaluative realism, for that matter – is forthcoming. But this does not make the considerations worthless, I hope, for they make explicit some of the unintuitive consequences of the realist approach, in a way that eventually may make one revise one's judgments about what one took to be one's own relevant intuitions.

The paper is in six sections. In the first section I focus on the target: dwelling on some earlier work of mine, I propose to characterize evaluative realism as rejecting what I call the *flexibility of values*, in contrast with other proposals that would make evaluative realism too easy or too hard. In the second section I present the particular flexible account of values I would favor, mainly due to David Lewis, and the scenarios whose intuitive description I take it strongly favor such flexibility, which are variants of the “Moral Twin Earth” submitted to related aims by Terence Horgan and Mark Timmons. People quite often claim, nonetheless, that they do not share the relevant flexibility supporting intuitions. The two main considerations I will offer aim to make them revise what they take to be their own intuitions. In section three I will claim that evaluative realism, including the dispositional variety of it, cannot account for internalism about *values*. In section four I will also consider the somehow trickier case of internalism about *value-judgments*. In section five I will present the Missing Explanation Argument, due to Mark Johnston, to the effect that flexible properties cannot be involved in causally explaining general dispositions of subjects to respond certain ways, and I will claim that evaluative properties intuitively do not appear in such explanations. In the final section six I will consider why, on the face of it, this is compatible with the views about so-called moral explanations of philosophers like Nicholas Sturgeon.

### ***1 Evaluative Realism vs. the Flexibility of Values***

The diversity of views intended under the label of “realism” is in my view particularly acute when realism about evaluative properties is concerned. Before presenting the one

---

\* Thanks to the LOGOS GRG, especially to Gemma Celestino, Sisco Gris, Pablo Rychter, Achim Spelten and Mike Wilms, and also to Agustín Arrieta, Josep Corbí, Esa Díaz-León, Manuel García-Carpintero, Kevin Mulligan, Ekai Txapartegi, y Agustín Vicente. Research has been partially funded by the research project BFF2003-08335-C03-03 (Spanish Government) and the research group 2001SGR00018 (Catalan Government).

I will use, I want to briefly mention some alternatives that, in my view, make evaluative realism too easy or too hard.

Consider for instance what is offered by Geoffrey Sayre-McCord:

Realism involves embracing just two theses: (1) that claims in question, when literally construed, are literally true or false (cognitivism), and (2) some are literally true. Nothing more. (1988b, 5)

I think it should be clear that realism *so conceived* will be a quite uncontroversial position. To illustrate, consider a caricature-like subjectivist account of values, having it that something like the following *defines* being good:

x is good iff we value x.

False as it might be for other reasons, the proposal *does* satisfy (1) and (2) and hence would be a realist proposal so conceived. But if this counts as realist, almost every possible view would as well.<sup>1</sup> Adding an epistemic element of the sort of

It is possible to find out about some moral sentences that they are true. (Thomson 1998a, 171)

does not seem to change the situation, as we can clearly find out sometimes what we value.

So it seems that one might have a non-realist approach to evaluative properties that respects that instantiations of them make simple predications of predicates signifying them straightforwardly, and sometimes knowably, true. On the other side, and, I take it, motivated by considerations like those just submitted, David Brink says the following:

A moral realist thinks that moral claims should be construed literally; there are moral facts and true moral propositions. Ethics is objective, then, insofar as it concerns matters of fact and insofar as moral claims can be true or false (and some of them are true). But moral realism claims that ethics is objective in another sense, which is not always distinguished, from this first kind of objectivity. Not only does ethics concern matter of fact, it concerns facts that hold independently of anyone's belief about what is right or wrong. This first kind of objectivity distinguishes moral realist and other cognitivist theories from nihilism and noncognitivism; the second kind of objectivity distinguishes moral realism from constructivist version of cognitivism. (1989, 20)

As it is specifically worded, though, it might seem that this also allows our former caricature-like subjectivist counting as a realist, as according to her the relevant responses on which goodness depends were not anyone's "belief about what is right or wrong" but rather a given desiderative attitude: *valuing*. But let us interpret Brink more liberally, as holding that evaluative realism require that evaluative properties have essences that are independent on relevant subjective mental responses, regardless whether they are doxastic or not. So understood, it would certainly exclude the considered subjectivist. But the problem now is that arguably it would exclude too much. There is a sense in which dispositions have natures that are not independent on their manifestations: disposition can be possessed when the manifestation does not occur, to be sure, but their relation to them is part of their essences, of what makes them the properties they are.<sup>2</sup> Take for instance dispositionalism about colors. According to

---

<sup>1</sup> I don't mean to suggest that Sayre-McCord is unaware of this, quite the contrary: he explicitly considers various possible "subjectivist" positions as varieties of realism, see Sayre-McCord 1988b, 16-9. It is only in this "cheap" sense, I take it, the Lewis himself describes his position as a realist one: values as he conceives them "do exist", see Lewis 1989, 93. For a similar view, consider Jackson: "*Realists* [are] cognitivists [who have it that the statements in question are semantically truth-apt] who take the extra step of holding that the ethical properties are instantiated" (Jackson 1998, 128).

<sup>2</sup> See Fine 1994 for an elaboration of the view on essences I am relying on here, and García-Carpintero ms. for the application to the distinction dispositional vs. categorical.

the view, colors are dispositions to produce in certain subjects, say, normal human perceivers as they actually are, certain mental responses, say, an experience as of a certain color being instantiated, under certain conditions, say, normal viewing conditions as they actually are. Hence colors have natures that involve mental responses. Being that so, one may say, makes a difference with respect to the alternative so-called *primary view* about colors: according to dispositionalism colors are less than fully objective properties but that is not so according to the primary view. But both views arguably are, and are certainly taken to be, varieties of *realism* about colors. *Mutatis mutandis*, one should expect, for the case of values: a view according to which values are fully objective properties, whose natures are independent of any mental subjective response, should certainly count as a form of evaluative realism. I will refer to such a view as (*evaluative*) *objectivism*. But realism should better not require objectivism by definition.

It is worth noticing that arguably for both primary, fully objective properties, and secondary, real but dispositional, properties, broadly conceived Fregean considerations require that there should be some descriptive material that fixes that they are signified by certain expressions and concepts. And in the case of colors, they arguably involve precisely the relevant chromatic subjective responses. For reasons that are familiar from Kripke (1980), this suffices for the following being not only true but also *a priori*

x is red iff x is disposed to produce in normal human perceivers an experience as of red in normal viewing conditions

even if only contingently true. According to these realists views about colors, you only get something that holds *necessarily* by rigidifying on the relevant expressions, as in

x is red iff x is disposed to produce in normal human perceivers *as they actually are* an experience as of red in normal viewing conditions *as they actually are*.

That is something that they both “primarists” and dispositionalists can, and do, hold. As suggested, the difference among them seems to lie in whether they hold that the former holds *in virtue of the nature of the color* or not<sup>3</sup> – and in my view this is something for settling which *a posteriori* considerations provided by the specialist are required.

I think that something like this is precisely characteristic of *realism* about the colors, and this is what I propose to generalize. Let me say then that if **F** is a property, an *rd biconditional* for (a predicate signifying) it is a substantial biconditional of the form:

x is f iff x has the disposition to produce in subjects S the mental response R under conditions C

or the form

x is f iff subjects S have the disposition to issue the x-directed mental response R under conditions C

where ‘is f’ signifies **F**, and ‘substantial’ is there to avoid “whatever-it-takes” specifications of either S, R or C.<sup>4</sup>

---

<sup>3</sup> See García-Carpintero 2002 and Wedgwood 1998.

<sup>4</sup> One such “whatever-it-takes” specification of, say, subjects S would be “those subjects, however they be, such that something is disposed to produce in them responses R under conditions C iff it is F.” *Mutatis mutandis* for the responses and the conditions.

Let me also say that a specification of the subjects in an rd biconditional is *rigid* iff the relevant predicate involved in the specification is rigid,<sup>5</sup> and *flexible* otherwise. So take for instance ‘normal human perceiver.’ This is not, as it stands, a rigid specification. For take the relevant predicate ‘is a normal human perceiver’ and suppose that in the actual world, it is true (even if knowable only *a posteriori*) that being such is being a human with a perceptual apparatus meeting condition ABC. Now consider a counterfactual situation in which, due to whatever reason you might think of, the human perceivers that are *normal there* are those with a perceptual apparatus meeting the different condition DEF. Now intuitively, it is this other property of being a human with a perceptual apparatus meeting condition DEF which would be relevant for evaluating sentences containing ‘is a normal human perceiver’ with respect to this other world. But then ‘is a normal human perceiver’ is not a rigid predicate but flexible. Its relevant rigidification, which can be put as something like ‘is a normal human perceiver *as they actually are*’ leads nonetheless to a rigid specification of the subjects, of the sort of ‘normal human perceivers *as they actually are*.’

An rd biconditional is *rigid* iff it involves a rigid specification of the subjects, and that it is *flexible* otherwise. Finally, a given property is *flexible* iff there is a flexible rd biconditional for (a predicate signifying) it which holds in virtue of its nature and hence necessarily.<sup>6</sup> With all these stipulations I can state my proposal about realism thus:

A property is *real* iff it is not a flexible property.

What considerations would be relevant for the issue as to whether a given predicate signifies a real vs. a flexible property?

Suppose that ‘is f’ signifies<sup>7</sup> property F, and suppose that S and C are the relevant flexible specifications of subjects and conditions, and S<sub>@</sub> and C<sub>@</sub> their relevant rigidifications, and that the only relevant rd biconditionals are

(R) x is f iff x is disposed to produce in S<sub>@</sub> the response R under conditions C<sub>@</sub>.

(F) x is f iff x is disposed to produce in S the response R under conditions C.

Both are, we may suppose, true with respect to the actual world, and *a priori* knowably so, we may also suppose. But the following asymmetry arises: –abstracting now from issues about essence vs. necessity– their metaphysical status covaries with the nature of F as stated in

F is real                      iff (R) is necessary      iff (F) is contingent

<sup>5</sup> I am assuming, with Kripke (1980), and a lot of people in discussions on philosophy of mind, philosophy of science or metaethics, that the notion of rigidity might be extended to be applicable to predicates, roughly along the lines of: a predicate is rigid iff it signifies the same property in all relevant worlds. Proposals like this have recently received criticisms, among which: that it would trivialize, making *all* predicates trivially rigid (see for instance Soames 2002), and that in any case it would overgeneralize, counting as rigid predicates some that do not signify natural properties/kinds (see for instance Schwartz 2002). I try to respond to these criticisms, respectively, in my unpublished ‘Rigidity for Predicates and the Trivialization Problem’ and ‘Predicates Rigidly Signifying the “Unnatural.”’ In the latter I also argue that the relevant simple predicates like those that will concern us here, ‘is red,’ ‘is funny,’ ‘is good’ and the like are, nonetheless, rigid. Given this I will speak of them *signifying properties*, without relativizing such talk to worlds.

<sup>6</sup> This is the notion I labeled *flexible response-dependence* in López de Sa 2003. I am abstracting here from issues related to response-dependence.

<sup>7</sup> See footnote 5 above.

F is flexible      iff (R) is contingent      iff (F) is necessary.

This provides a way of testing whether ‘is f’ signifies a real or a flexible property, and based just on *a priori* considerations. The recipe is, very abstractly put, this: consider what could be a counterexample of the necessity of the relevant statement on the assumption that the predicate signifies one particular kind of property. Those I will refer to as *target situations*. Then check how these should be intuitively described (with respect to the relevant predicate) and conclude accordingly.

I will instantiate this sort of relevant consideration in the next section. But let me end this one with the following remark about words. The question of the appropriateness of labels *per se*, when philosophical terms are at issue, does not appear to be particularly interesting philosophically, *once* the relevant distinctions are clear and attended to. There certainly seems to be a contrast between entirely objective properties *and* dispositional properties, *on the one side*, and flexible properties, *on the other*, as issued in the question of how the relevant target situations should be intuitively described. My aim here is to present a case against the view that evaluative are properties *of the former kind*, however they are called. I will call them, though, *real properties*.

## **2    *A Flexible Lewisian Theory of Values and the Intuitions about Evaluative Twin Earth***

That *some* evaluative properties are intuitively flexible is, I take it, quite uncontroversial. Consider the case of ‘is funny.’ Suppose that the following are the relevant flexible and rigidified biconditionals

x is funny iff x is disposed to amuse us under appropriately attentive conditions.

x is funny iff x is disposed to amuse us as we actually are under appropriately attentive conditions as they actually are.

Now take something funny, even something, as I am ready and willing to grant, *really really* funny, like *The Simpsons*. Gerald Lang suggests that we would not take very seriously the suggestion that it “would continue to be funny even if a comprehensive alteration in our comic sensibilities took place” (2001, 201). That is, in a very compressed form, an instance of the relevant consideration we have just considered, to the effect that being funny is flexible and not real. As there is no doubt that *The Simpsons* is actually funny, there is no doubt that it is disposed to amuse us as we actually are under appropriately attentive conditions as they actually are. Consider now a relevant counterfactual target situation, *w*, in which this alteration of our sensibilities takes place, but which resembles the actual world as much as possible compatibly with this. *The Simpsons* is *not* disposed to amuse us *as we would be in w* under appropriate attentive conditions *as they would be in w*.

So far we have the relevant target situation, appropriately neutrally described, as no hypothesis about the extension of ‘is funny’ *with respect to w* is introduced. The crucial question is now: how should it be intuitively described with respect to ‘is funny’? In particular, is it true or false, intuitively, ‘*The Simpsons* is funny’ *when evaluated with respect to w*? Lang says that we would not even take seriously the suggestion that it might be true. But now, if ‘*The Simpsons* is funny’ is false with respect to *w* and ‘*The Simpsons* is disposed to amuse us as we actually are under appropriately attentive conditions, as they actually are’ is true with respect to *w*, the rigidified biconditional is

only contingently true with respect to the actual world. Hence, ‘is funny’ signifies a flexible property, and not a real one.

One might say: “But we do say, at least sometimes, that *The Simpsons* is funny, in the objective mood, as it were, rather than *we find them funny*. Furthermore, we say those things even acknowledging that they could not amuse some people, for after all some days, although funny all the same, they don’t even amuse us. Why couldn’t we say then that *The Simpsons* are really funny even in the target situation, only that those unlucky people fail to be disposed to be amused by them?” The straight answer is that we *could* definitely say so: it’s only that intuitively we, or at least most of us, don’t want to. Remember that the crucial issue is how a given target situation should be *intuitively described*. In the submitted consideration, there is also another important element which is worth stressing to avoid possible misunderstandings. The fact that we have simple predicates like ‘is funny’ signifying the property of being funny does arguably entail that there should be a “real”/appearance distinction concerning what is funny, that being funny should be distinct from seeming funny or actually amusing. But that of course is also the case *even if funny is a flexible property*, and hence in particular does not entail anything about what the proper intuitive description of target situations should be. There are things which seem funny although they are not fun at all,<sup>8</sup> and conversely, as submitted, *The Simpsons* are funny even if they sometimes fail to seem so. But that is indeed entailed by the use of the *dispositional* idiom in the rd biconditionals. Dispositions can be possessed without issuing their characteristic manifestations. And conversely, their manifestation could occur without being the manifestation of a possessed disposition. Flexible properties are not dispositions, true enough. But with respect to each world, the things that have a given flexible property in this world are those that are disposed to produce the relevant response in the subjects as they are in that world under the conditions as they are in that world. Hence, in each world, having the property, being funny, is not the same as issuing the relevant response, seeming funny.

The same situation occurs, I claim, for a number of similar *soft* evaluative predicates: ‘is tasty’, ‘is disgusting’, ‘is comfortable’, let alone ‘is sexy,’ ‘is fashionable’, or ‘is cool.’ With respect to any of these, it seems, hardly anyone would claim to have the intuitions supporting their signifying real properties. Does it generalize with respect to *all* evaluative predicates, including the *hard* cases of moral and some aesthetic predicates? Consider the following general rd biconditional, adapted from the proposal by David Lewis in his ‘Dispositional Theories of Value’ (1989):

x is good iff we are disposed to value x in appropriate reflective conditions.

Some remarks are in order. First, *valuing* is the favorable attitude of desiring to desire. That valuing is an desiderative attitude rather than doxastic is arguably entailed by its being a *favorable* attitude. But “first order” desiring would certainly *not* do: we, unfortunately quite often, desire things we do not value at all. Weakness of will is, of course, a case at hand. Take “unwilling smokers,” as one might call them, like me myself. I do desire to smoke a *Ducados* quite often, I actually *love* smoking. But I find some uneasiness even in *reporting* it as I have just done. It is not, or at least not only, that I have a contrasting desire *not to* smoke: that will be a case of conflicting desires – which by the way could eventually issue in conflicting valuing or even in moral dilemmas. But phenomenologically, my case of smoking is not, or at least is not only,

---

<sup>8</sup> See Wright 1992, 101 for a dozen of examples of this.

constituted by what I experience when for instance I have contrasting desires about enjoying a good film this afternoon and remaining in my office finishing this paper. In this case I do not want to get rid of either desire: I would prefer the world to be so that they could both be satisfied in a way I know it unluckily isn't, so that finally I will act just upon one of these. My smoking is different: I *do* want to get rid of my desirings to smoke: even if I desire to smoke, I desire better not to desire to smoke at all. So that, when I light I cigarette, I'd say that my weakness is weak, given that I desire not to have the desire that makes me do so. So failing to desire as one values is failing to desire as one desires to desire. Hence, it seems, valuing is desiring to desire.

Second, *we* are, according to the proposal, those that are disposed to value, with respect to the relevant particular issue at stake, exactly like the speaker is. It is important to stress that, so understood, 'we' turns out to be a *flexible* characterization of a group of subjects. The relevant predicate signifies with respect to the actual world the property of being relevantly the way I am *actually*. But I could be otherwise, and in particular my disposition to value could be very different from what it actually is. But then, with respect to those worlds in which I am suitably different, it will signify the property of being relevantly the way I *would be* in those situations.

Third and finally, *appropriate reflective conditions* are a bit schematic. In Lewis' original paper, he submits that the relevant conditions are the conditions of fullest possible imaginative acquaintance with the thing in question, *possible* for the subjects in question and relatively to the thing in question (1989 77-9). This element has found some resistance in the literature.<sup>9</sup> Some would like to claim that further elements should be included, notably awareness of all (non-evaluative) relevant facts. I tend to agree with Lewis that this further element, crucial for the question of *balancing* different probably conflicting values, should not be included in the conditions determining the values to be balanced in the first place (see Lewis 79-82). But the issue is delicate, and I would rather not go into this. My proposal is then to characterize the relevant conditions as *the appropriately reflective conditions*, having in mind these Lewisian conditions of fullest possible imaginative acquaintance, but perhaps also some others like the ones considered if further thought renders them appropriate.

Let me come back to our issue. It would seem that if 'is good' signifies a flexible property then arguably all evaluative predicates do as well.<sup>10</sup> Does it? As we have seen, settling this depends on the status of the relevant flexible and rigid biconditionals

x is good iff we are disposed to value x in appropriate reflective conditions.

x is good iff we, as we actually are, are disposed to value x in appropriate reflective conditions, as they actually are.

which in turn depends on which turn out to be the intuitively proper description of suitably neutrally described counterfactual target situations. Such it seems to be, I claim, the generalized version of the Moral Twin Earth submitted in related contexts and for

---

<sup>9</sup> See for instance Johnston 1989 and Smith 1994.

<sup>10</sup> This is straightforward if one characterizes, as I am inclined to do, *evaluative* predicates as those that suffice for 'is good' (or 'is bad'). One should expect the claim also to hold, I guess, in some alternative formulation is adopted.

related aims by Terence Horgan and Mark Timmons,<sup>11</sup> that I will call *Evaluative Twin Earth* or *ETE* for short.

Take something such that I am –and hence *we* are– *actually* disposed to value under appropriate reflective conditions: (say) Santi’s lying to me in some particular occasion. We, as we actually are, are disposed to value Santi’s lying to me under appropriate reflective conditions, as they actually are, and hence it is –actually– good. But I could be different. In particular my dispositions to value under that relevant conditions this particular lie could be suitably more “deontologist,” as it were. I could be such I was not disposed to value it under appropriate reflective conditions. So let us consider a situation in which I *was* that way, but agrees with the actual situation in as much as it’s possible compatibly with this difference, and call it ETE. We, as we are in ETE, are not disposed to value Santi’s lying to me under appropriate reflective conditions.

So far, again, we have the relevant target situation, appropriately neutrally described, as no hypothesis about the extension of ‘is good’ *with respect to ETE* is introduced. The crucial question is again: how should it be intuitively described with respect to ‘is good’? In particular, is it true or false, intuitively, ‘Santi’s lying to me is good’ *when evaluated with respect to ETE*? My own intuitions, and as I understand him, also Lewis’, is that it should be *false* with respect to ETE. But then, if ‘Santi’s lying to me is good’ is false with respect to *ETE* and ‘We, as we actually are, are disposed to value Santi’s lying to me under appropriate reflective conditions, as they actually are’ is true with respect to *ETE*, the rigidified biconditional is only contingently true with respect to the actual world. Hence, ‘is good’ signifies a flexible property, and not a real one. Hence, arguably all evaluative predicates, and not only soft ones, do as well: the intuitive flexibility of values is then vindicated and thereby evaluative realism is rendered unintuitive.<sup>12</sup>

As I said from the beginning, people quite often claim, nonetheless, that they do not share the relevant flexibility supporting intuitions. As I tend to think that they *do* have them, the aim of this paper is to offer two considerations in the light of which some might revise what one took to be one’s own realist supporting intuitions when the proper description of the ETE is concerned. But as I have also announced, the considerations aim only at that, and they do not constitute a refutation of the alternative, realist, approach to values.

### **3 Internalism vs. Evaluative Realism**

John Mackie famously once developed an argument from queerness against there actually being objective goods, where:

An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact of this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. (1977, 112)

Here he is pointing to what it is sometimes called the *practicality* of the evaluative or *internalism*, roughly: values, whatever they are, have a to-be-pursuedness somehow built into them. That certainly seems something *constitutive* of values as we conceive them. Such it is according to Lewis:

---

<sup>11</sup> See Horgan & Timmons 1991, 1992a, 1992b, 1996 & 2000a. Their own positive views might differ from mine, though. For theirs see *inter alia* Horgan & Timmons 2000b.

<sup>12</sup> For further details and discussion see López de Sa 2003.

If something is a value, and if someone is of the appropriate 'we', and if he is in ideal conditions, then it follows that he will value it. And if he values it, and if he desires as he desire to desire, then he will desire it. And if he desires it, and his desire is not outweighed by other conflicting desires, and if he has instrumental rationality to do what serves his desires according to his beliefs, then he will pursue it. And if the relevant beliefs are near enough true, then he will pursue it as effectively as possible. A conceptual connection between value and motivation. But a multifariously iffy connection. Nothing less iffy would be credible. But still less it is credible that there is no connection at all. (1989, 72)

I propose to state this internalism claim about values thus:

- (I) It is necessary and *a priori* that: If something is good, we would desire it under appropriate reflective conditions (weakness of will and the like aside).

Evaluative realism cannot account for (I). The reason is straightforward: realism entails that the relevant flexible rd biconditional would be at most contingently true. But any counterexample to its necessity is such that the embedded conditional in (I) would be false with respect to it. Hence it would *not* be necessary with respect to the actual world, and hence (I) is false.

Evaluative objectivists, who hold that evaluative properties are fully objective ones, typically agree and even emphasize this, but then give (I) up and go externalists. They usually claim that that is indeed a virtue of their position, given that the externalist component is independently motivated. Some of them stress what is an undeniable fact: that sometimes people fail to be moved by what is good, even by what they know is good. That would challenge a strengthened version of (I) having it that values *directly* motivate the relevant subjects by directly issuing in them the relevant desire. That would be, I agree, as a matter of fact *not true*, let alone necessarily and *a priori* so: we have already considered cases of weakness of will in which we fail to desire as we value. These by itself would refute the strengthened version of (I). But (I) is suitably weaker, not only on that score, but importantly in requiring that one should value the good *only* under certain, appropriately reflective, conditions. So in order to refute (I) you will need a case of something which is good but such that the appropriate valuers don't desire it at all, *not even* under the appropriate reflective conditions and when their will is strong enough as to desire as they desire to desire. But this seems a quite hard thing to have. This case, one is inclined to say with Lewis, is simply not credible.

Someone like David Brink will agree with a lot of this, although he would put it the other way round, as it were:

[T]he internalist cannot rest content with the extensional claim that everyone is in fact motivated [by what is morally good]. Any externalist could claim that. The internalist about motives claims that it is true in virtue of the concept of morality that [moral goodness] necessarily motivate. According to the internalist, then, it must be conceptually impossible for someone to [know that something is morally good] and remain unmoved. This fact raises a problem for internalism: internalism makes the amoralist conceptually impossible. (1986, 29-30)<sup>13</sup>

The dialectical situation is weird enough, though, for the conceptual impossibility of such an amoralist, who is not disposed at all to desire something, which is good, even under appropriate reflective conditions and with a strong enough will, far from raising a problem for internalism is precisely what motivates it. The reason for (I) can be put by the thought that such an amoralist *is* conceptually impossible.

---

<sup>13</sup> In the original passage, instead of the inserted claims about moral *values* Brink has claims about moral *considerations* and judgments, but I take it that he would certainly, even readily, say the corresponding about the properties and facts as I've put it. I'll consider internalism concerning judgments in next section 14.

Do we have here an irremovable clash of intuitions? This could be, of course. But I take it is dialectically fruitful enough, for as I said some realists *do* indeed seem to appeal to (I) in rejecting objectivism and to claim instead that evaluative (and moral) properties are, though real, somehow more *subjective* by being essentially tied to (evaluative) responses, analogously as colors are according to the dispositionalist. But this move is unsuccessful.

In his response to Mackie, John McDowell took such an “analogist” line, by arguing that the model for real evaluative properties should not be looked for in the case of primary qualities, as Mackie did, but in the case of *secondary* qualities:

[I]t seems impossible – at least on reflection – to take seriously the idea of something that is like a primary quality in being simple *there*, independently of human sensibility, but is nevertheless (not conditionally on contingencies about human sensibility) such as to elicit some ‘attitude’ or state of will from someone who becomes aware of it. (1985, 111)

Shifting to a secondary-quality analogy renders irrelevant any worry about how something that is brutally *there* could nevertheless stand in an internal relation to some exercise of human sensibilities. Values are not brutally there – not there independently of our sensibility – any more than colours are: thought, as were colours, this does not stop us supposing that they are there independently of any particular apparent experience of them. (1985, 120)

I don’t want to go here into McDowell’s specific views concerning values – nor colors, for that matter. Rather, I want to claim that to the extent that one tries to accommodate (I) by claiming that values are real even if not fully objective properties, but rather dispositions to produce certain evaluative response in (rigidly) specified subjects under (rigidly specified) conditions; to that extent the try fails. For dispositionalists about value do not face (I) any better than objectivists did.<sup>14</sup> And this is so given that the previous remark about the incompatibility of realism and (I) did not appeal to any specific view about the nature of being good besides the assumption that it was a real property and, hence, applies in particular to the relevant, secondary, dispositions.

The dispositionalist about values can of course at this point simply deny that (I) should be true, and try rest comforted (say) with the *a priori* component of it, as we have seen evaluative objectivist do. The issue as to whether internalism about values is right or not depends on exactly the same intuitions that would support more directly the reality of flexibility of values. Hence this is not an independent consideration for settling the issue. But given that, as we have seen, some people think, mistakenly, that they can accommodate internalism about values within a realist framework, the consideration is worth making, as it is capable of making some revise what one took to be one’s own realist supporting intuitions, and as such is submitted.

#### **4 *Evaluative Judgment and Motivation***

Internalism in meta-ethics is sometimes intended as a related, though distinct, claim asserting an *a priori* and necessary connection between evaluative (moral) *judgment* and motivation. According to the flexible account of section 2, there is such a connection. Lewis says of it that

it is even iffier that the connection between value itself and motivation; and again I say that if it were less iffier, it would be less credible. If someone believes that something is a value, and if he has come to this belief by the canonical method [of putting himself in ideal conditions and finding whether he values it], and if he has remained in ideal conditions afterward or else retained the desire to desire that

---

<sup>14</sup> This has been also explicitly emphasized with respect to the original Moral Twin Earth by Holland 2001.

he had when in ideal conditions, then it follows that he values that thing. And if he desires as he desires to desire, then he desires that thing; and so on as before. (1989, 73)

One could here wonder whether it is really true that were it less iffy, it would be less credible. For the belief one reaches by the canonical method, if it includes *succeeding* in achieving the relevant conditions, would indeed constitute evaluative (moral) *knowledge*. But as it is sometimes stressed, the conceptual connection between evaluative judgment and motivation seems to be independent of whether the judgment is *in fact* true: false beliefs about what is good could motivate all the same than true ones.<sup>15</sup> But the canonical method might not be interpreted as necessarily successful: it is sufficient that one reaches what one *takes to be* the relevant conditions. So we have the following:

If someone believes that something is a value, and if he has come to this belief by the canonical method of putting himself in *what he takes to be* ideal conditions and finding whether he values it, and if he has remained in *what he takes to be* ideal conditions afterward or else retained the desire to desire that he had when in *what he takes to be* ideal conditions, then it follows that he values that thing. And if he desires as he desires to desire, then he desires that thing.

That is so even if the one is not right in what he takes to be the relevant conditions, and hence, even in one's belief is in fact not true. Now, to the extent one *typically* form one's evaluative by trying to approximate the canonical method, one's judgment typically entail that one's disposed to desire it, under appropriate reflective conditions (weakness of will aside).<sup>16</sup> But even if one typically does it, one needs not:

If someone reached the same judgement in some non-canonical way – as he might – that would imply nothing about his valuing or desiring or pursuing. (Lewis 1989, 73)

But this, it seems to me, accords with common-sense's view pretty well.

Can the realist account at least for this internalism about evaluative judgement? Brink thinks it cannot. I tend to think he is right, although arguing for such a further incompatibility would involve some complications.<sup>17</sup> In any case, the consideration that I wanted to offer was the previous one.

## **5 The Intuitively Properly Missing Evaluative Explanations**

Mark Johnston has recently argued against the view that colors and other manifest properties are *response-dependent*, when a property is response-dependent in his terms iff it is a flexible property in mine.<sup>18</sup> Abstracting from the details, it runs more or less

---

<sup>15</sup> See, for instance, Dreier 1990.

<sup>16</sup> If I understand them right, this is close to what is argued in Jackson & Pettit 1995, see also Jackson 1998.

<sup>17</sup> One in my view plausible sufficient condition would be what some philosophers have argued was right in verificationism: for a family of properties like evaluative ones it should be possible to determine sometimes that some of them are instantiated.

<sup>18</sup> His characterization of response-dependence is:

[A] property, Being F, is response-dependent if there is some predicate 'is f' which expresses the property (i.e., whose extension across possible worlds is just the things which have the property) such that some substantial way of filling out 'R', 'S' and 'C' makes

x is f if and only if x is disposed to produce x-directed response R in all actual and possible subjects S under conditions C

thus: the idea that some properties are perceptible require “receptivity”, that there should be causal explanations of the general dispositions of the subjects to elicit the responses under the conditions *in terms of those properties*. But those explanations would go missing if the properties were flexible: satisfying receptivity entails that the relevant flexible rd biconditionals are *contingent*. Hence the label *Missing Explanation Argument* or *MEA* for short.

This provides in my view a further consideration that could make one revise what one took to be one’s own realist intuitions in the evaluative case. According to the MEA, if a property is flexible, there will certainly be “deep” causal explanations of the general dispositions of subjects, explanation that will appeal to certain real properties that unify the relevant instances in the actual world, but those will not appeal to the flexible properties themselves. But precisely this is what intuitively seems to occur with respect to evaluative properties. Take a soft case. Our general disposition, as we actually are, to be amused by some things in appropriately attentive conditions, as they actually are, will certainly have casual explanations in terms of real properties: perhaps we are actually disposed to be amused by some things *because* they make us expect a connection among ideas that we know are not so connected. But intuitively we would not offer *as a causal explanation* of our dispositions to be amused that the things *are funny*. And *mutatis mutandis* for the general case: one should expect that there be a complicated causal explanation of why is it that we are actually disposed to value certain things and not others in the conditions, but intuitively it would not do as a causal explanation *that they are good*.

As before, this consideration again falls short of constituting a full-blooded *argument* against evaluative realism. As an argument it could be seen as presupposing that the relevant causal explanatoriness of the property in question vis-à-vis the relevant responses is a *necessary* condition for its reality, in the sense I am using the notion. Now, the evaluative realist could complain, one could grant that it would be a sufficient condition, and one could even grant, as occurred in the premises of the MEAs, that concerning colors, or any other kind of *perceptible* property, it *is* a necessary condition. But what should that be in general? In particular, why should it be the case for evaluative properties to be real that they were causal explanatory vis-à-vis the relevant responses in the envisaged way in which they intuitively aren’t? That is, as I understand it, the content of Nagel’s complaint (quoted in Sturgeon 1985, 235):

---

*a priori* and necessary; (Johnston 1998, 9)

once it is required

that the canonical biconditionals are not merely superficial necessities produced by “rigidifying” on a relation that is itself contingent. The equivalence “ $x = \text{Neptune}$  if and only if  $x = \text{the planet in the actual world which causes perturbations in the orbit of Uranus}$ ” is superficially necessary in this way. (1998, 10)

That the proper target of the argument are response-dependent properties *so understood* and hence not the views that most people submit under the label of response-dependent accounts of colors – dispositionalist theories of colors – is something I stress in my unpublished ‘The Explanations that are Missed according to the Missing Explanation Argument.’ This, acknowledged by Johnston himself (1998, 37), is rightly emphasized by Haukioja (2000, 109), but apparently had escaped other critics, like Miller (2001) and myself (2000).

it begs the question to assume that *explanatory* necessity is the test of reality in this area. (Nagel 1980, 114n)<sup>19</sup>

Fair enough, I'm inclined to acknowledge. But as with the previous issue concerning internalism, I take it that inasmuch as reflection upon the tension between evaluative realism and internalism about values could make one revise what one took to be realist supporting intuitions concerning the proper intuitive descriptions of the evaluative target situations, reflection upon the present issue about causal explanatoriness could oblige one to make a similar revision. This being so, and even if they fall short of constituting an argument against evaluative realism, I hope the consideration is not devoid of interest.

## 6 *Visiting Evaluative Explanations*

I have just suggested that some evaluative realists explicitly endorse the – counter-intuitive, as I take them to be – consequences of their views: externalism and causal explanatory impotence. The latter may come as a bit of a surprise, in that some other evaluative realists, notably Nicholas Sturgeon, are usually seen as precisely defending that there *are* moral explanations of the sort I've claimed evaluative realists and anti-realist alike acknowledge that intuitively there aren't. In this section I want to defend that this impression concerning Sturgeon does not stand up to a closer analysis, and that his arguments do not contradict what I have been claiming so far.

In his classical paper 'Moral Explanations' (1985), Sturgeon aims to rebut a claim he attributes to Harman,<sup>20</sup> according to which "*even if* we assumed the existence of moral facts they would still appear explanatorily irrelevant" (1985, 237), for discussing which, as he observes, he is free to, and does, "*assume*, for the sake of the argument, that there are moral facts" (1985, 237). One could think at this point that that is not a substantive assumption, amounting to something like "there are true simple modal statements." Not so: as he himself makes explicit, his assumption has a much richer content – and, as we are going to see, essentially so – that those moral facts involve moral properties that are, or supervene upon more basic, natural properties (1985, 247), so that for anything that has them, "it could not have differed in its moral quality without differing in those other [more basic features that makes it have it] as well. (1985, 249).

Let me say a few words on supervenience. There are good (in part a posteriori) reasons for holding that everything supervenes upon the way the world naturally is. How to characterize exactly the content of this rough claim is, of course, by no means easy. But it will be clear that evaluative properties such as the Lewisian, flexible, approach conceives them, *do* clearly supervene on the natural in this sense. (At least, they do so on the assumption that psychological entities, to which evaluative ones flexibly reduce, do.) Furthermore, in the literature there is also a claim sometimes intended as a supervenience claim such that anyone is committed (at any moment) to evaluate similarly things that she judges not to differ naturally. That also holds, again obviously, for flexible response-dependent values. What is *not* true according to the flexible proposal is that evaluative properties supervene *locally* to natural entities, and more in general, to entities which are independent of the relevant valuers. Indeed for

---

<sup>19</sup> As I understand her, something like this is also the view of Judith J. Thomson, see Harman & Thomson 1996, Thomson 1998a and 1998b.

<sup>20</sup> See footnote 21.

any target situation, if its proper description favors a flexible account, then it constitutes a counterexample of the relevant local supervenience claim. And conversely, the relevant realist alternatives could indeed be alternatively characterized by holding the relevant local supervenience claims.

It is then clear that the content of Sturgeon's assumption is, in my terms, that moral properties are *real* properties. This is OK for evaluating Sturgeon's target: that even if moral properties were real properties, they would be explanatorily irrelevant. And his argument is straightforward:

[C]onsider Harman's own example in which you see the children igniting a cat and react immediately with the thought that it is wrong. Is it true, as Harman claims, that the assumption that the children are really doing something wrong is "totally irrelevant" to any reasonable explanation of your making that judgment? Would you, for example, have reacted in just the same way, with the thought that the action is wrong, even if what they were doing hadn't been wrong, and could we explain your reaction equally well on that assumption? ... [I]f what they are actually doing is wrong, and *if moral properties are, as many writers have held, supervenient on natural ones*, then in order to imagine them not doing something wrong we are going to have to suppose their action different from the actual one in some of its natural properties as well. So our question becomes: Even if the children have been doing something else, something just different enough *not to be wrong*, would you have taken them even so to be doing something wrong? (1985, 247, my emphasis)

And the answer, I am ready to grant, could be 'no.' Suitably generalized, and in our terms: we have granted that there will be deep empirical explanations of our issuing the relevant evaluative responses when confronted with instances of evaluative properties, and our general capacity of so issuing them. If it is *assumed* that the relevant explanatory properties *are* the evaluative properties, then evaluative properties wouldn't be explanatorily irrelevant. That is something that a defender of the flexibility of values could, and I think should, accept.<sup>21</sup> But that is *not* an "argument to the best explanation" for evaluative realism, though it is sometimes seen in this way.<sup>22</sup>

### ***Conclusion***

Evaluative realism, I have claimed, should be characterized as denying the flexibility of values. But, I have also claimed, the intuitive description of the relevant counterfactual target situations, as the Evaluative Twin Earth, *does* support flexibility. As some people often claim they disagree with this, I have tried to make explicit some of the consequences of the relevant alternative realist descriptions, with the hope that some do acknowledge their counterintuitive character and revise thereby what they took to be their own realist supporting intuition. According to some evaluative realists, the flexibility supporting intuitions about the proper description of the relevant counterfactual target situations would have their own counterintuitive consequences. To the best of my knowledge, it is usually claimed that flexibility would have unacceptable relativistic consequences with respect to the evaluative domain, for instance being incapable of accounting for the fact that people *disagree* in normal

---

<sup>21</sup> And I don't know why Sturgeon thinks Harman wouldn't: Harman has always been explicit concerning "reductivist" straightforwardly solving the question of the explanatoriness of evaluative properties.

<sup>22</sup> If I understand her right, Orlando 2001 reconstructs Sturgeon's argument this way, and then she (rightly) accuses it of begging the question (see Orlando 2001, 339). As for her own "abductive argument" for evaluative realism, she also seems to grant that evaluative properties are explanatorily irrelevant *vis-à-vis the relevant responses*, but adds that moral facts could explain other moral facts. Provided that the explanation here is not empirical, that is fully compatible with the flexible account I favor.

conversations about evaluative issues. I do think that flexibility has relativistic consequences, but I would resist the claim that they are unacceptable. Rather, they are what intuitively seems predictable, in particular when a presuppositional element to the effect that participants in the conversations

### **References**

- Brink, David 1986: 'Externalist Moral Realism', *Southern Journal of Phil.* 24, 23-41
- Brink, David 1989: *Moral Realism and the Foundations of Ethics*, Cambridge, MA, Cambridge University Press
- Dreier, James 1990: 'Internalism and Speaker Relativism', *Ethics* 101, 6-26
- Fine, Kit 1994: 'Essence and Modality', *Philosophical Perspectives* 8, *Logic and Language*, 1-16
- García-Carpintero, Manuel 2002: 'A Non-modal Conception of Secondary Properties', plenary lecture at ECAP4, Lund (Sweden)
- García-Carpintero, Manuel ms: 'A Non-modal Conception of Dispositions'
- Harman, Gilbert and Judith J. Thomson 1996: *Moral Relativism and Moral Objectivity*, Oxford, Basil Blackwell
- Haukioja, Jussi 2000: *Rule-Following, Response-Dependence and Realism*, University of Turku, Turku.
- Holland, Sean 2001: 'Dispositional Theories of Value Meet Moral Twin Earth', *American Philosophical Quarterly* 38/2, 177-96
- Honderich, Ted (ed.) 1985: *Morality and Objectivity*, London, Routledge & Kegan Paul
- Hooker, Brad & Margaret Olivia Little (eds.) 2000: *Moral Particularism*, Oxford Clarendon Press
- Horgan, Terry & Mark Timmons 1991: 'New Wave moral Realisms Meets Moral Twin Earth', *Journal of Philosophical Research* 16, 447-65
- Horgan, Terry & Mark Timmons 1992a: 'Troubles on moral Twin Earth: Moral Queerness Revived', *Synthese* 92, 221-60
- Horgan, Terry & Mark Timmons 1992b: 'Troubles for New Wave Moral Semantics: The Open Question Argument Revived', *Philosophical Papers* 21, 153-75
- Horgan, Terry & Mark Timmons 1996: 'From Moral Realism to Moral Relativism in One Easy Step', *Crítica* 28, 3-39
- Horgan, Terry & Mark Timmons 2000a: 'Copping Out on Moral Twin Earth', *Synthese* 124, 139-52
- Horgan, Terry & Mark Timmons 2000b: 'Nondescriptivism cognitivism: Framework for a New Metaethic', *Philosophical Papers* 29/2, 121-53
- Jackson, Frank 1998: *From Metaphysics to Ethics*, Oxford, Clarendon Press
- Jackson, Frank and Phillip Pettit 1995: 'Moral Functionalism and Moral Motivation', *The Philosophical Quarterly* 45/178, 20-40
- Jackson, Frank, Phillip Pettit & Michael Smith 2000: 'Ethical Particularism and Patterns' in Hooker & Little 2000

- Johnston, Mark 1989: 'Dispositional Theories of Value,' *Proceedings of the Aristotelian Society*, suppl. vol. 63, 139-74
- Johnston, Mark 1998: 'Are Manifest Properties Response-Dependent Properties?', *The Monist* 81, 3-43
- Kripke, Saul, 1980: *Naming and Necessity*, Oxford, Basil Blackwell
- Lang, Gerald 2001: 'The Rule-Following Considerations and Metaethics: Some False Moves', *European Journal of Philosophy* 9/2, 190-209
- Lewis, David 1989: 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society*, suppl. vol. 63, 113-38, reprinted in Lewis 2000, from where I quote.
- Lewis, David 2000: *Papers in Ethics and Social Philosophy*, Cambridge, MA, Cambridge University Press.
- López de Sa, Dan 2000: 'Conceptos dependientes de respuesta: El problema de la explicación perdida', in Mary Sol de Mora et al. (eds.), *Actas del III Congreso de la SLMCE*, Dpto. de Filosofía de la UPV/EHU, Donostia, 165-73, English translation available at [http://www.ub.es/grc\\_logos/people/dlopezdesa/MEA.pdf](http://www.ub.es/grc_logos/people/dlopezdesa/MEA.pdf)
- López de Sa, Dan 2003: Response-Dependencies : Colors and Values, PhD thesis, U. Barcelona, available at [http://www.ub.es/grc\\_logos/people/dlopezdesa/RDs.pdf](http://www.ub.es/grc_logos/people/dlopezdesa/RDs.pdf)
- Mackie, J.L. 1977: *Ethics: Inventing Right and Wrong*, London, Penguin Books, chapter 1 reprinted in Sayre-McCord 1989a, from where I quote.
- McDowell, John 1985: 'Values and Secondary Qualities', in Honderich 1985
- Miller, Alexander 2001: 'The missing-explanation argument revisited', *Analysis* 61/1, 76-86
- Orlando, Eleonora 2001: 'Abduction, Realism and Ethics', *Theoria* 16/41, 331-52
- Sayre-McCord, Geoffrey 1988a: *Essays on Moral Realism*, Ithaca & London, Cornell University Press
- Sayre-McCord, Geoffrey 1988b: 'The Many Moral Realisms' in Sayre-McCord 1988a
- Smith, Michael 1994: *The Moral Problem*, Oxford, Blackwell
- Schwartz, Stephen P., 2002: 'Kinds, General Terms, and Rigidity' *Philosophical Studies* 109, 265-77
- Soames, Scott, 2002: *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*, Oxford University Press
- Sturgeon, Nicholas L., 1985: 'Moral Explanations', in D. Copp & D. Zimmerman (eds.), *Morality, Reason and Truth*, Totowa, N. J., Rowman & Allanheld, reprinted in Sayre-McCord 1988a, from where I quote.
- Thomson, Judith, J. 1998a: 'Précis of Part Two', *Philosophy and Phenomenological Research* 58/1, 171-3
- Thomson, Judith, J. 1998b: 'Reply to Critics', *Philosophy and Phenomenological Research* 58/1, 215-22.
- Wedgwood, Ralph 1998: 'The Essence of Response-Dependence', *European Review of Philosophy* 3: *Response-Dependence*, R. Casati & Ch. Tappolet (eds.), CSLI Publications, Stanford, Cal., 31-54

Wright, Crispin 1992: *Truth and Objectivity*, Cambridge, MA: Harvard University Press