# Circularity, Naturalism and Desire-Based Reasons

**Abstract.** This paper proposes a critique of the naturalist version of the Desire-Based Reasons Model. It first sets the scene by spelling out the connection between naturalism and the Model. After this, it introduces Christine Korsgaard's circularity argument against what she calls the instrumental principle. Since Korsgaard's target, officially, were non-naturalist advocates of the principle, the paper shows why and how the circularity charge can be extended to cover the naturalist Model. Once this is done, the paper goes on to investigate in some detail the different ways of responding to the circularity challenge. It argues that none of these responses succeed, at least not without serious costs to their advocates. The paper then ends with a brief summary and some concluding remarks.

**Keywords.** Christine Korsgaard, reasons, desires, circularity, ethical naturalism, normative question

## I. Naturalism and the Model

In the theory of normative reasons one popular approach is the Desire-Based Reasons Model (henceforth: Model). The Model typically comes with an ethical naturalist (henceforth: naturalist) background. Not everyone rushes to endorse naturalism, however. There are plenty of influential arguments against naturalism but my aim in this paper is to investigate a different and perhaps less discussed objection that does not originally target naturalists: Christine Korsgaard's circularity charge. To this end, we first need a suitable account of what naturalism is, and in this context we then have to locate the Model; after this we can turn to the objection mentioned.

My main focus will be on what is often called substantive naturalism. On this view, naturalism is understood as proposing an account of normative properties in terms of natural properties or relations.[1] The two main variations are the analytical and the non-analytical

[1] Here I set aside the difficulties surrounding the notion of natural property. Instead, I will act on the supposition that such an account can be given. For a good overview of different definitions and the difficulties they face see Copp (2003) and Ridge (2014).

versions.[2] Analytical naturalism holds that normative properties are natural properties and the two, normative and descriptive ways of capturing them are synonymous. In contrast with analytical naturalism, on non-analytical naturalism concepts and properties come apart: though to each normative term there is a corresponding descriptive term and these terms refer to identical properties, the two terms are not synonymous. Although normative properties are reducible to natural properties, the identity statements employed are synthetic, not analytical.

With these distinctions in mind, let us now see how naturalists interpret the Model. There are many versions but given the level of abstraction we will be operating at in the paper, it suffices to remark that they can be grouped roughly into two sets: those that use actual desires of the agent (with or without selection) and those that idealize (i.e., use hypothetical desires that may or may not be actual). For illustration, take Mark Schroeder's (2007, 193) Hypotheticalism that uses actual desires without selection:

> "For *R* to be a reason for *X* to do *A* is for there to be some *p* such that *X* has a desire whose object is *p*, and the truth of *R* is part of what explains why *X's* doing *A* promotes *p*."

This definition is easily readable along naturalist lines (Schroeder himself is a non-analytical naturalist). Take Schroeder's example (Ibid: 29). If Ronnie (*X*) likes dancing (*p*), then the fact that there will be dancing at the party (*R*) helps explain why Ronnie's going to the party (*A*) would promote Ronnie's desire to dance. "Explanation" here is meant in the metaphysical sense: explanations are facts about "what is true because of what" (Ibid, note 19). Hence this

[2] There is at least one other version, often called Cornell Realism. Like non-analytical naturalism, this view only makes claims about property identity, but unlike non-analytical naturalism, it does not claim that there is a descriptive way of capturing normative properties. Even though we may know that normative properties are natural properties, we may not be able to tell *which* properties they are. I set this version aside since it is unclear how it can give us the Model: it refuses to provide an explicit reduction, which the naturalist Model clearly is. For more on ethical naturalism including references, see Lenman (2014) and Tanyi (2009).

particular feature of Ronnie's situation becomes a reason for Ronnie. In short, the Model designates two properties; one normative (the relational property of being a reason for), the other natural (as it appears in the part of Schroeder's formula that contains reference to the promotion of the agent's desire), and claims that these properties are identical, and the terms used to capture the properties may or may not be synonymous.[3]

## II. Korsgaard's circularity argument

We now have a clear enough view of the Model and its connection to a naturalist meta-ethics. It is time to turn to criticism. To do so, we have to begin from somewhat afar, with an objection that does not, officially, target naturalists. In an influential argument, Korsgaard (1997, 240-1; 2003, 110; 2009, Chapter 4) has launched an attack on the non-naturalist or, to use her term, dogmatic rationalist account of practical reason. Dogmatic rationalists, she says, hold that there are "eternal normative verities", i.e., irreducible facts about what we have reason to do and it is our knowledge of these facts that we apply in action. But, Korsgaard goes on, the agent who is facing this claim can surely ask: "Why should I care about these facts?" "Why should I apply this knowledge in action?" And there is no answer, she says. This is a problem, she concludes, because it leaves practical principles without justification: even the most uncontroversial ones loose their grounds.

Her primary example is the instrumental principle (Korsgaard 1997, 241-2). Very roughly (here the details don't matter), the principle says that one has at least *prima facie* reason to take the means to one's ends. But if the fact that this is what we have reason to do is just another "eternal normative verity" that we apply in action then, by the above logic, one can ask

[3] For completeness' sake, here is an example from Dancy (2000, 28) for a version of the Model that uses idealization (but retains the focus on actual desires): "If its being the case that $p$ is a good reason for $A$ to $\varphi$, *this is because*, there is some $e$ such that $A$ actually desires $e$ and, given that $p$, $\varphi$*-ing* subserves the prospect of $e$'s being realized (or continuing to be realized)." And then Dancy adds the further condition: "and in condition $C$ $A$ would desire $e$" where 'condition $C$' is a placeholder for whatever idealization process the advocate of the Model would like to insert (there are many candidates).

why one should care about that fact. And what can the rationalist say to this? He can try the following route. One, perhaps necessarily, has the end of taking the means to one's ends; one has a reason to do what promotes one's end of taking the means to one's ends; taking the means to *this* end does just that. Obviously, for this argument to work we have to show that we indeed necessarily have the end of taking the means to our desires. But this is only the smaller problem and, perhaps, it is not impossible to prove this to be the case. For, the opponent can point out that even if we necessarily have the end of taking the means to our ends, the fact that we have a reason to take the means to *that* end is just the same old fact, which we have no more reason to care about than before. Consequently, instead of a solution, we get *circularity*.

Nor is the rationalist better off if he tries to widen the scope of the first premise of his reasoning. For what can he say? There are two options (Korsgaard 2003, 110-2; 2009, 65-6). Either he says that the end referred to in the first premise is the end of doing good action; or he says that it is the end of doing what is supported by reasons. But this produces the same difficulty as before. Imagine how the argument would work. It is our, perhaps necessary, end to do good action; taking the means to our ends is itself a means to our end to do what is good; hence we have a reason to take the means to our ends. We thus fall into the same trap: in arguing for it, we presuppose the principle itself. This should not be surprising. After all, our problem arises because we bump into the question of application. But the instrumental principle is the principle of application itself, thus it is no coincidence that we try to justify the principle with itself. Circularity is thus an unavoidable consequence of the rationalist account of the instrumental principle, and this puts the rationalist in deep trouble.

## III. Extending the argument

The first issue we have to deal with when considering Korsgaard's argument is whether it is relevant for us at all. That Korsgaard discusses the instrumental principle is of course good for us since the Model relies on a particular interpretation of the principle (on which ends are given by our desires, actual or hypothetical). However, the meta-ethical background is not fitting. Korsgaard explicitly speaks of dogmatic rationalists as her target and, as I noted, with this term she refers to contemporary non-naturalists (she mentions Parfit) and their predecessors such as the intuitionist Samuel Clarke or Richard Price. Although non-naturalism can claim to provide a version of the Model, advocates of desire-based reasons are typically not from this camp (and this, as has been shown elsewhere (Dancy 2000, 27; Tanyi 2007, Chapter 1), is for good reasons). In any case, in this paper my focus is on the naturalist version of the Model. Thus, before we proceed to any kind of substantial analysis of Korsgaard's argument, we have to tackle two questions. First, we have to ask what sort of justification Korsgaard has in mind and, second, whether the demand for justification understood in this way can be extended to the naturalist account of the instrumental principle (and hence to the Model) as well.

Start with the first question. The issue here is what drives Korsgaard's question of application. I think it is fairly clear that she has in mind a *practical* challenge: any rationalist account of practical reason must show that the property it identifies as an ethical property has a bearing on our deliberation and conduct. Korsgaard herself puts the problem in the following way. She says that when dealing with practical issues what we are dealing with is motivation, but not any kind of motivation. Take the case at hand: the idea that we should take the means to our ends. We are ordinarily motivated to take the means to our ends: the bare co-presence of an end and a suitable instrumental belief is enough to "effect a motive". But, she points out, this motive may be the result of mere *causation*: one can simply be so *conditioned* that he always takes the means to his ends (Korsgaard 1997, 221; 2009, 62-3). What we need, therefore, is that the motivation be the result of the agent's own *recognition* of the appropriate conceptual connection between the belief and the end: that he is moved to act because he thinks he has *reason* to act (Ibid., 1997, 243). In this way, we will have the agent put into the picture: the act will be the result of the agent's own mental activity "and not merely the result of the operation of beliefs and desires *in* her" (Ibid., 221). Korsgaard calls this "rational motivation" and thinks that this is what every theory of practical reason must account for.

A theory can fail to meet this demand in two ways (Korsgaard 1996, 12-6, 42, 46-7, 81; 2003, 112). In the first case, we are concerned with *motivational* issues, accordingly, if we fail here, we will fail to account for the motivation to act on one's normative judgement. What we investigate here is whether normative judgments necessarily motivate or only occasionally, and whether normative beliefs or only desires are able to move us to act and how the two issues are connected. This is what Korsgaard calls the criterion of explanatory adequacy. In the second case, we are concerned with the*justification* of that motivation, accordingly, if we fail here, we may have the motivation but the judgement involved won't be a *normative* judgement: it will lack justification in the sense that the agent won't see reason to act. Consequently, once the agent sees what is behind practical claims according to the given theory, he will refuse to endorse his own motivation. And if he does that, his action, though it might still be explained on the given theory, will be the result of mere causation and not rational motivation. This is what Korsgaard calls the criterion of normative adequacy. A proper theory, Korsgaard claims, must meet both criteria; but she also makes it clear that her primary concern is the second criterion: she calls it the *normative question*.

This puts our original problem in context. Korsgaard's question of application makes a practical demand, but now we know that this demand can take different forms. Korsgaard herself appears to have changed her mind as to which form the circularity argument uses. In her earlier writings, when she deals with this problem, she normally describes it as a motivational issue: the agent who faces the truths or facts rationalists propose is just not motivated to act on them (Korsgaard 1996, 37-8; 1997, 240-1). The criticism inspired by the other form of the demand she preserves for "empiricists", i.e. naturalists whose primary representative she takes to be Williams. Here she argues that the empiricist account of the instrumental principle cannot guide action because there is no way one can violate it (Korsgaard 1997, 223-33; 2009, Chapter 4). In the absence of error, however, we only have the agent acting on a purely causal basis: there is no rational motivation involved since there is no reason to act. I propose to set aside this argument since it is no concern for us here; it has been assessed, and I think refuted, elsewhere.[4] Our question should be whether the circularity argument and the question of application can only be given a motivational reading, or the alternative reading is also available.

I think it is. To begin with, in some places in her earlier writings and clearly in her 2009 book, Korsgaard herself puts aside the motivational reading. In the book she is crystal clear on this: the problem with rationalists is not motivational, but normative. In particular, she says that the dispute between Hume and contemporary rationalists concerning the motivational power of moral perception "may just be a standoff". But, she claims, "if we think about normativity, *rather than motivation*, then we will find that there is something in Hume's complaint" (Korsgaard 2009, 64; Italics are mine). And then she goes on to introduce the question of application and discusses it as a problem of normativity: why we would not be bound by (obligated to act on) the instrumental principle on the rationalist account (Ibid., 65-8). Also, in her first employment of the question of application, she uses the question to refute theories on the ground that they fail to answer the normative question (Korsgaard 1996, 28-48). There are, moreover, just too many ways of overcoming the motivational challenge, other than the one Korsgaard herself considers. There are internalist as well as externalist alternatives and burgeoning literature concerning their plausibility. Finally, the question of application appears to be neutral as to the two rival readings. One can just as well ask the question because, though they are moved to act on their knowledge, they see no reason why to: they are reluctant to endorse the motivation. That is, in accordance with the two points just made, we can grant the

[4] For discussion and references, see Tanyi (2007), Appendix III.

rationalist some account of motivation, but still think that agents would refuse to act on this motivation once they see what is behind the claims these theories make on them, i.e., once they see what is supposed to guide their conduct.

This also connects us to our second original inquiry: whether the question of application can be extended to naturalism as well. We just have to look at Korsgaard's more general philosophical enterprise. It is to argue against what she calls substantive realism by employing the question of application. She defines substantive realism by contrasting it with her own procedural realism (Korsgaard 1996, 36-7). The former holds that there are procedures for answering normative questions because there are ethical facts or truths that exist independently of these procedures. Whereas the latter claims that there are answers to normative questions because there are procedures for arriving at them: there is no need to suppose the existence of ethical facts or truths independently of these procedures. Now, it is clear that naturalism and non-naturalism belong to the same substantive realist camp. For both, ethics is a theoretical enterprise: to find out and then apply in practice the ethical knowledge we gain in the world (cf. Korsgaard 1996, 38, 40, 43-44). Hence they both invite the question of application, and if what I have argued for is accepted, this can take the form of the normative question. The naturalist Model is different from this general picture only in the way it fills in the details: it speaks of facts of desire-satisfaction. Consequently, we just have to substitute "desire" for "end" in Korsgaard's circularity argument and then follow the same reasoning as before.[5]

## IV. Responses to the argument

I suggest therefore that we read the circularity argument as posing a normative, not merely motivational challenge against any kind of substantive realist view, including the naturalist Model. We should first realize that the problem Korsgaard describes, though technically not an

[5] Cf. Schroeder (2005), but he appears to have a different understanding (the *Cudworthy* argument, as he calls it) of Korsgaard's challenge than what I have argued for here.

infinite regress, is nevertheless something very similar to it: it begins with a question and then, instead of getting an answer, we end up asking the question over and over again. Consequently, the types of responses also follow the strategies one approaches an infinite regress with. There are three: to show that questioning doesn't start; to accept that it starts but argue that it can be stopped; or to admit both that it starts and that it cannot be stopped but claim that this is not a problem. In what follows I assess all three types of responses. I start with the first, continue with the third and end with the second.

## First response

The best representative of the first response is an argument by Peter Railton. He begins with the familiar observation that "a substantial amount of our rational, intentional activity must be 'automatic', unmediated by reasoning and recognition" (Railton 2004, 185). On Railton's view what happens in such situations is that the agent relies upon his senses, memory or thoughts blindly. He *trusts*, that is to say, attributes default *authority* to his senses, memory or thinking. He expects that things are as he perceives them to be, or that they happened in the way he remembers and thereby also learns important information about his environment or about his own thinking (Ibid., 187). Although the trust involved is defeasible, it serves Railton's purpose well enough. For it allows that the agent's intentional activity may not involve an element of judgement, while taking his activity at the base justified. And this rules out a regress-type of questioning. At the ground level, default trust provides sufficient justification but being non-judgmental and passive, it doesn't invoke standards of reasoning, which would stand in need of justification. Circularity is avoided.

However, the claim that the need for justification elapses because we have default trust in our senses, memory and thinking is puzzling. Railton himself names the problem. "But how could so passive and non-judgmental an attitude as default trust be the foundation for a form of reason-disciplined agency?”, he asks (Ibid., 191). Of course, this is just a rhetorical question to which Railton readily provides the answer (Ibid., 191-4). His idea is that while both belief and desire are instances of default trust, they are also *normative* attitudes. They are normative because their nature is to realize a certain normative *role* in the individual’s mental architecture: truth-tracking and value-tracking respectively. And they are instances of default trust because they need not involve judgement: it suffices if they constitute an *expectation* that something is the case or that something is desirable. Through belief one learns how things are in one’s environment or how one’s thoughts fit together and through desire one learns about what is valuable and what is not.

The question is whether the account delivers the results Railton expects from it. I don’t think so. Let us begin with an idea from Scanlon (1998, 24). He claims that what sustains automatic intentional activity are certain *standing normative judgments* to the effect that if some putative evidence is not good ground for forming beliefs, or if certain reasons are not good grounds for action, then one does not even unreflectively form beliefs on the basis of such evidence or act on the basis of such reasons. Since these judgments are standing, they are not present in the agent’s consciousness but are only activated on certain occasions. Yet, since they *are* judgments they invoke standards of reasoning, which in turn invite justification. And though the need for justification rarely arises, perhaps it happens only in cases of great distress, this is still enough to open the ground for questioning and circularity.

This idea serves as an ideal supplement to Railton’s proposal. We can say that, in virtue of their normative role, desire and belief provide the material for our reasoning: what they mediate serve as ‘candidate reasons’ for us. We can also grant that these attitudes are instances of default trust: they function automatically without reasoning and judgement. This is not an unusual thought, others, such as Korsgaard or Scanlon, are also willing to grant this role to desires (though their grounds are different) (Korsgaard 1998, 51-4; Scanlon 1998, 65). But, and this is the important bit, what they are directed at is just an apparent reason, which becomes a real reason, i.e. a normative reason only after the agent has decided in its favour. And if Scanlon is right, this decision and the subsequent forming of the belief or action on its basis, is governed by the standing normative judgments he describes. Consequently, our intentional activity, even when it is spontaneous and automatic, is the result of a process that comprises both the aspect of default trust and that of norm-governed decision (cf. Korsgaard 1996, 243). We are back where we started: questioning gets off the ground.

**Third response**

Turn now to a radical alternative that accepts both that questioning starts as well as that it is circular. But then it adds that no problem arises from this: an infinite chain of questions is not vicious. The idea is well formulated by Copp (1995, 43-4). He distinguishes between the *process* of justification, i.e., the agent's performing the steps of justification and the *status* of being justified, i.e., the claim's having a place in an infinite chain of justifications. Like in geometry: perhaps there is one theorem of geometry only if there is an infinity of theorems, but this is not to say that there is one theorem only if an infinity of theorems has been proven to be such. Copp claims that in the matter at hand what we need is the first reading, and we need it only in a weak sense: the fact that a principle has a place in an infinite chain of justifications is not sufficient to show it *not* to be justified. If this is true, our problem evaporates. A principle can be justified just because it has a place in an infinite chain of justifications; it is not needed that the agent goes through the infinite steps of justification in order to confer justification on the principle. The same is true of the instrumental principle: it can *be* justified without us, the agents justifying it.

Of course, we would then still have to say what that justification consists in. But this is not difficult for we can simply revert to the original realist accounts of the principle. The crucial issue is not this. What carries the argumentative weight in Copp's argument is his claim that in the present case what we need is the first reading. This choice clearly builds on a particular understanding of justification, what we might call third-person justification. It claims that the agent need not have access to the full justification of a given principle; he need not be able to comprehend it. But this is only an *assumption*: Copp does nothing to support his choice; he only assumes that he is right. And it is not obvious that he is right. Korsgaard, for instance, makes it clear that the need for justification arises from a first-person standpoint: it is the agent who asks the question about how to regulate his conduct. Hence the answer to the question had better be accessible to him, had better be something he can grasp and appreciate the import of. The third-person standpoint, Korsgaard (1996, 16-7) admits, also appears in a theory of practical reason but it has a different role. We saw what this role is. It is the task of motivational explanation, when we show how and why practical claims have psychological effects on the agent.

The question, then, is which of the two readings we should opt for. I don't think there are decisive reasons in favour of either reading. But there are certainly good reasons that support the first-person reading, so all I will do here is to list those reasons. First, attributions of justification try to pick out agents as acting conscientiously, i.e. in a responsible, blameless manner with regard to what they should do. Therefore, agents should at least be given the opportunity to detect the justification of their actions. Perhaps they won't always seize this opportunity, but if they did they could find the missing justification (Radzik 1999, 39). Second, justification is a regulative notion. It signifies a property that people should be able to think usefully about in order to decide how to act (Ibid.). What we expect from a theory of justification are guides for our action and this is only possible if we are capable of detecting what the justification of a given act consists in. This idea is familiar from debates on indirect consequentialism; it is often called the transparency or publicity condition (Rawls 1971, 130; Williams 1973, 128; 1985, 101-2). The claim is that a theory that requires widespread ignorance of its account of justification is self-defeating. For its distinctive contribution to ethical theory is exactly this account, and the distinctive interest of such an account is to provide a basis for decision-making. Therefore, when a theory doesn't require people to be aware of its account of justification it abandons the very basis on which its own foundation, as a distinctive ethical theory, is laid.

**Second response**

Our best choice, then, is to stop questioning. Here we find several alternatives. Let us begin with the simplest one. Recall Korsgaard's argument. Her basic problem is that a realist construal of the instrumental principle always leads to the question of application. But this invites the obvious response. Suppose we don't try to justify the principle with itself but with reference to another, more basic principle. What then? Korsgaard thinks that this wouldn't work. Due to the problem of application, she says, such a move would only produce a chain of justification, i.e. an infinite regress of principles. This is why the instrumental principle is so crucial: in their efforts to avoid regress, realists must employ the instrumental principle (Korsgaard 1997, 242; 2003, 111-2; 2009, 64). We saw how this would work: say that conforming to the more basic principle is an end of the agent, then claim that following the instrumental principle is a means to satisfying this end. But then you realize that you implicitly re-employ the instrumental principle. The problem, to repeat, is that the instrumental principle is the principle in accordance with which we apply truth in practice, hence it cannot be used as its own support.

However, according to some philosophers, this response only begs the question. It squarely denies what they think is possible: that there are principles in the case of which the question of application does not arise (Parfit 1997, 121-9). And indeed, Korsgaard makes it clear that on her view no such appeal is acceptable: it is a refusal to answer the normative question (Korsgaard 1996, 34, 39-41; 2003, 112). I don't want to settle this debate here; instead, let me point out two things that show why the debate is not directly relevant for us. First, on the face of it, the present response is not available to advocates of the Model, since they take the instrumental principle to be fundamental. Moreover, second, those who make this response are not naturalists but non-naturalists, thus the meta-ethical background they use to stop questioning is again not available to naturalist advocates of the Model, which is what our concern here.[6]

Yet, naturalists may point out that the instrumental principle does not exhaust the Model. In fact, they could say that the focus on the principle is misleading because this is not what matters for a thesis about desire-based reasons. This is Hubin's (2001) central point in his response to Korsgaard: what he calls "pure instrumentalism" – "the thesis that reasons communicated across causal, criterial, and mereological relations" – is, he says, uncontroversial and is the not the defining part of the Model. Instead, what the Model really claims, is that "reasons…are grounded, ultimately, in the subjective, contingent, conative states of the agent" (Ibid., 459). If this is so, Hubin goes on, what we need to stop questioning is to point out that these reason-grounding desires are "brute facts" like the ultimate rule of recognition in Hart's legal theory. That is, just as we define legal validity as being in accordance with this ultimate rule and hence we cannot meaningfully ask the question whether this rule itself is valid, we can say that an action's being justified - "rationally advisable" is the term Hubin uses - is the very same thing as its sub-serving certain of the agent's desires, full stop.

I think this is an intriguing idea but there are several problems with it. First of all, the analogy with legal theory is not complete. As Hubin (Ibid., 464) also remarks, Hart did anchor the ultimate rule of recognition in something outside the legal system: in the complex social

[6] There are further developments along these lines that bring in what Scanlon (2014) dubs 'reasons fundamentalism' (in a nutshell: normativity is understood in non-naturalist terms and the fundamental normative property is taken to be that of a reason). See Parfit (2011), pp. 415-420 for responding to Korsgaard in this way; Dreier (2015) is an excellent critical discussion (he re-interprets the normative question as one about what he calls 'rational necessity', a form of motivational connection).

fact of acceptance of the rule in a population. Now, this does mean that, ultimately, it is people's attitudes ("desires") that ground the rule and I suppose this is what Hubin is driving at to support his point. However, the analogy with legal validity then breaks down: although we cannot ask questions about the legal validity of the ultimate rule, we *can* ask the question whether we should have follow the rule. So the proper analogy with practical reason would be asking whether we should follow our desires – which is of course what their being brute facts is supposed to deny. Setting this problem aside, there is the question whether Hubin's understanding of the Model is what other advocates of the Model would also like to have. In particular, and again as Hubin (Ibid., 466-7) appears to notice, normativity in his version of the Model is carried only by the *internal coherence* of one's system of desires with one's pursuits and actions: the desires themselves have no normative standing. I think many would want to deny this, leading to debates about the inherent normative nature of desires and about the question whether desires are based on reasons.[7] Besides, this picture of the Model – echoed in what Dancy (2000, 34) calls the Advice Point – leads to the unsettled debate about the normativity of rationality: are such coherence/consistency requirements normative? Why are they normative? Is it because we have reason to be rational? Hubin would have to settle this debate in his favour before his response to Korsgaard could really make its point.[8]

The next attempt starts with a distinction. According to Velleman (2000, 176), the object of any enterprise is either formal or substantial. A formal specification gives us the concept of the object of the enterprise: 'winning' in the case of a game, for example. A substantive specification, on the other hand, specifies what it is to achieve the formal aim: to run the fastest time, for example. Put in this framework, the formal object of practical reasoning is to do what

[7] For discussion and references, see Tanyi (2007; 2009; 2011).

[8] For the debate (including references), see Tanyi (2007) and more recently, Fink (forthcoming). Hubin would have to tackle challenges such as that rational requirements have only a wide-scope and thus no detachment of a normative directive is detachable, and he would also have to argue that the normativity of his coherence requirement does not itself stem from reasons (but, then, *where* would it come from?).

one should do, whereas the substantive aim could be anything including, as on the Model, doing what satisfies one's desires. Using this distinction, Wedgwood (2002, 142, 147; 2005, 468) has claimed that there is a way to meet Korsgaard's objection. For, it is hard to see what sense would be in asking why one should do what in the formal sense he should do: it would be like asking "Should I do what I should do?" Sure, one can just announce that he sees no point in acting for reasons. But then he opts out of practice altogether and can just as well commit suicide: his life is devoid of all value and is pointless. At the same time, for the others the question of what to do is settled. There is no question of application; their conclusion is regulative of their choice.

Wedgwood's proposal, however, does not necessarily provide us with a solution. First, the instrumental principle appears to articulate a substantive aim, Wedgwood certainly treats it so, so how do we get to it? Wedgwood himself is not concerned with this question. Yet, he is wrestling with another problem and his reasoning may give us a hint (Ibid. 147-8). In response to Velleman's objection that the notion of a formal object is empty, he points out that the formal reading is compatible with specifying what one in the formal sense should do. That is, it does not deny the existence of "non-trivial general truths" akin to Velleman's substantive aims. What it claims is that these truths are not given to us in advance of our deliberations about what to do; instead, we have to discover them there.[9] But once we found them, once we know what in the formal sense we should do, we have to act accordingly (Wedgwood 2005, 468). Yet, this still doesn't give us the instrumental principle. There is no assurance that the principle will be among the truths discovered and that it will be the only one. Nor is it clear what the meta-ethical

[9] The contrast here is with "basic principles of rational choice". It is interesting that Wedgwood doesn't say much about this, for him crucial category. His main idea is that certain principles are given to us "merely in virtue of our being rational beings" by which he seems to mean that it is constitutive of rational agents that they have a disposition to follow these principles. See Wedgwood (2002, 144, 147). This is a possible interpretation but it is puzzling in light of Wedgwood's (2005, 465) own endorsement of Railton's criticism of such constitutive arguments. Wedgwood's talk of principles is also strange. For he seems to suggest that these principles are substantial enough - his analogy with the principles of logic and mathematics seems to suggest this at least. But then it is mysterious why he thinks that Velleman's emptiness objection is a serious problem that should be handled through singling out these basic principles.

standing of these truths is. It can turn out that what we get in the end is an account that is not consistent with ethical naturalism.

Up to this point, our strategy has been to tackle Korsgaard's argument from the theoretical side by trying to show that certain normative truths can stop the circle of questioning. Let us approach the problem from the other, practical side now. Here is where Korsgaard's own proposal comes into view. It is built up of several steps. First, she suggests that we should approach the problem of justification as a problem of practical problem-solving. We should start with a real practical problem the agent has to solve and then show that the given principle does indeed solve that problem. As she puts it, "If you recognize the problem to be real, to be yours, to be one you have to solve, and the solution to be the only or the best one, then the solution is binding upon you." (Korsgaard 2003, 116) Now, the question is what that problem is. Korsgaard's answer is simple: action. Our plight as self-conscious beings is that "we find ourselves with the necessity of making choices and so in need of reason to act" (Korsgaard 1998, 62). "Human beings", she says, "are *condemned* to choice and action …[this] is the simple inexorable fact of the human condition" (Korsgaard 2009, 1-2).

The next step is to clarify what Korsgaard means by this claim and how the instrumental principle fits the picture. The two issues are connected, so I don't separate them either. Korsgaard's views allow for two interpretations (FitzPatrick 2005, 664-5). On the first reading, certain principles, including the instrumental principle are literally necessary to exercise agency, that is, we need them to be able to act at all. We find two variations here. The first starts from the widely accepted idea that in order to remain an agent one must have ends. Having an end is *constitutive* of being an agent: acting is a teleological enterprise (Korsgaard 1996, 122; 1998, 51, 60-2). Then a further thought comes: willing the means is *constitutive* of willing the end. That is, something the realization of which does not at all concern us in our deliberations, cannot qualify as an end for us. It is like walking and putting one foot in front of the other: one cannot walk unless one puts one foot in front of the other (Korsgaard 1996, 36; 1997, 249). If we add these two ideas together, we get this: taking the means to our ends is *constitutive* of agency. We cannot act unless we take the means to our ends. And since action, unlike walking, is "our plight", the instrumental principle is justified for us.

The second variation takes the claim about constitution for granted but combines it with a further psychological thesis: practical principles are necessary for the *unification* of agency. In particular, those who don't follow the instrumental principle will disintegrate as their agency degenerates into a passive arena for the operations of competing desires (Korsgaard 1997, 247, 254). In her most recent writings Korsgaard puts this idea at the core of her views about justification and normativity. "The necessity of confirming to the principles of practical reason", she says, "comes down to the necessity of being a unified agent…[which] comes down to the necessity of being an agent … [which in turn] comes down to the necessity of acting…[which] is our plight" (Korsgaard 2009, 25-6). Korsgaard thus gives us a second and perhaps even more pressing reason to conform to the instrumental principle. The principle is not only constitutive of our agency, but is also something we must in the majority of cases follow if we are to maintain our integrity as unified persons. Too many violations of the principle result in disintegration: our agency will fall apart making us incapable to act. Again, the principle is justified for us: questioning is stopped.

The same problems beset both proposals. There is good reason to think that they are not correct and even if they are, they don't help the defence of the Model. Let us proceed in reverse order. Korsgaard's driving thought is the idea that principles of practical reason serve as solutions to the practical problem of acting. But neither of her proposed solutions employ normative truths and facts; they are left out of the picture altogether. Nor are they needed. If either of the above variations is correct, conformity to the instrumental principle is literally practically necessary. And this is not surprising. Korsgaard intends her account as the constructivist *alternative* to realism: normative concepts name the problem and the principles propose the solution. There is no aim to track normative facts outside the will; instead, practical principles emanate from the will (Korsgaard 2003, 116). In addition, both variations encounter problems of their own. The appeal to the preservation of agency by avoiding disintegration works only in a general way, pointing to the problem that will plague us if we *regularly* fail to take the means to our ends. It therefore cannot explain why someone should obey the principle in a *particular* case, which is obviously the kind of requirement that the idea of practical justification is premised upon (FitzPatrick 2005, 674).

The claim that the instrumental principle is constitutive of agency, on the other hand, can be read in two ways. The first conforms to the literal practical necessity account and holds that in order to remain an agent one must *actually* employ the principle. But it is not impossible to imagine cases in which one does not act on the principle; in fact, this *has* to be possible since, as Korsgaard herself emphasizes when discussing the empiricist account of the principle, it must be possible to violate the principle if it is to count as normative. Hence the best solution is to give up the literal practical necessity reading and look for an alternative interpretation. Korsgaard's (1996, 36; 1997, 245) candidate is this. We can say that in willing an end the agent is *committed* to taking the means to that end. This is why there is a problem with failing to take the means to one's ends: it involves a failure to follow through on one's commitments. This solution also preserves the claim of constitution. Although actually acting on the instrumental principle is not constitutive of willing an end, thus of agency, it is nevertheless something the agent *implicitly endorses* while willing the end. (FitzPatrick 2013, 47) We cannot just shrug off the principle; it is justified for us.

I can accept this account of the principle. Yet, it is not obvious that it also solves the present problem. The issue here is not that advocates of the Model cannot appeal to this understanding of constitution; as a matter of fact, to the detriment of Korsgaard's constructivist enterprise, they can (FitzPatrick 2013). They can say, think of the first version of the Model, that the instrumental principle is grounded in a fact about the internal relations among the will's operations, namely, that of willing ends and willing means. We can construe this fact as a psychological fact of consistency between these two operations of the will.[10] But even when this is done, there are still two challenges to face. First, reference to this fact re-invites the question: why should I care about this fact? And now we have no literal practical necessity involved either: failing to act on the principle does not put an end to our agency. Second, even if this question receives an answer, the Model still needn't follow.[11] For, now we have based justification on a fact that opens the way for other practical principles that are also grounded in facts about consistency.[12] Again, the burden of proof is on the advocate of the Model: he must show that this is not the case.

We have one attempt left. It again comes from Railton. Just like Korsgaard, he argues that the instrumental principle is constitutive of agency. And for the same reasons he also thinks that we must understand constitution in a weaker sense: it does not require that one actually acts on the principle. What it requires, and this is the first difference between him and Korsgaard, is that the one has some *disposition* to follow the principle (Railton 2003, 307-13).[13] But the really important difference is that in Railton's view this kind of defence cannot provide justification. His problem is even more radical than mine above. It is not only that we can ask why we should not resist our disposition, but also that in certain situations we can genuinely wonder why we should not eradicate it by putting an end to our agency (Ibid., 313-5). For example, being a patient with an incurable, painful and costly disease, one can reasonably

[10] Alternatively, we can claim that when this fact is present, another non-natural, irreducible fact also occurs: that the act is rational. But, recall, this is not an option available to us in defending the Model in this paper.

[11] FitzPatrick (2013) provides an excellent discussion of possible answers to this challenge.

[12] Indeed, this is what Korsgaard claims within the constraints of constructivism: she thinks that categorical imperatives also follow due to the commitments taken up in the course of our exercise of agency. See Korsgaard (1996b, 120-3). For a critique of her argument see FitzPatrick (2005, 677-81).

[13] This is a restatement of Railton's view in which I follow Wedgwood (2005, 465).

question the point of staying alive.[14] Or, to take an example from Parfit, when one is attacked by a mob who seeks revenge on his family because he testified against them, he can be tempted by the idea to knock himself senseless temporarily or even permanently. Constitutive arguments, Railton says, have no resources to answer either of these challenges.

At the same time, however, he thinks that this problem and, presumably, mine problem above too can be remedied. For, he says, the agent will ask these questions in such a way that betrays deference to the instrumental principle again. What he will wonder about is whether crossing the line between agency and non-agency (or whether to act on his disposition or to follow through on his commitment) is the best or only way of getting what he most wants from life (Ibid. 315; cf. Rosati 2003, 522). But it is not obvious that the agent *must* make reference to the instrumental principle in order to raise his challenge. There are two options. There may be further principles that are also constitutive of agency and the agent invokes *these* principles in his challenge. We would then have a set of principles a member of which we have to invoke in order to raise a challenge about another member. But, crucially, we would not get the Model since the instrumental principle would no longer be fundamental. More tentatively, it is at least conceivable that the agent makes reference to none of these norms when posing a challenge (Wedgwood 2005, 466). He may just be genuinely puzzled about how to make up his mind about what to do, and unpersuaded by the proposals philosophers have offered so far.

A further problem looms if we accept Railton's proposal. If someone is puzzled about the instrumental principle, then repeated reference to the principle hardly helps him out. This is Korsgaard's problem again. To take Railton's example, the agent would ask: why should I do what gets me what I most want in life? But Railton thinks that it is exactly the possibility of

[14] Note that this kind of suicide is different from the suicide I referred to when dealing with Wedgwood's proposal. There we were asked to imagine an agent who, in a hands-up fashion, announces that he is not going follow any principle he is presented with. He is basically saying that he sees no point in acting for a reason; hence he cannot be offered a reason. In contrast with this, the present problem is exactly that the agent is looking for a reason to remain an agent but he sees none. I think Korsgaard is referring to the same sort of difference in Korsgaard (1996, 243), which makes it even more interesting why she doesn't consider Railton's problem.

circularity that shows why the agent cannot ask this question. To explain, he brings an analogy with Carroll's argument concerning *modus ponens* (Railton ibid., 316-7). Carroll has shown that if one doesn't reason in accordance with *modus ponens* when forming beliefs, then adding *modus ponens* as a premise in his reasoning doesn't help. For to effect a conclusion from the new premises, the agent would have to use *modus ponens,* and this is exactly what he doesn't do. As Railton rightly points out, there is a clear parallel between this argument and the present challenge. If one doesn't reason in accordance with the instrumental principle, then adding the principle as an end or a means in his reasoning doesn't help. For to effect a conclusion from the new premises, he would have to use the instrumental principle, and this is exactly what he doesn't do. Hence, Railton concludes, the instrumental principle cannot be just another premise in the agent's practical reasoning.

This is puzzling. For the point of Korsgaard's charge was exactly that on a realist construal the instrumental principle will *be* a premise in the agent's reasoning *because of* the push of justification. In fact, she explicitly uses the analogy with Carroll's paradox to illustrate her problem (Korsgaard 1997, 239-41; 2009, 66-7). Hence, from Korsgaard's point of view, Railton merely restates the problem without offering a solution to it. Of course, the analogy with Carroll's problem does suggest possible ways of solving it. Korsgaard, for instance, takes it to show that just as the agent's theoretical reasoning would be "a mere heap of premises" were she refuse to employ *modus ponens*, his practical reasoning would also fall apart without the use of the instrumental principle. This then leads directly to her second variation on the idea of literal practical necessity: the claim that we need principles of practical reason in order to unify our agency (Korsgaard 2009; cf. Blackburn 1995, 709-10) And, at least we cannot rule this out, there can be other lessons one might draw from a parallel with Carroll's paradox.[15] But

[15] The best such idea comes from Dreier (2001, 38-45). His problem is this. If one believes that a rule requires him to act but is not moved to act, then what is missing must be a desire. But this means that the instrumental principle cannot be just another rule that we need a desire to comply with. And the reason for this is just the parallel with Carroll's paradox: if we require such a desire, then we will never get to the end of the questions. This is significant.

Railton does no such thing and in the absence of such an attempt it is hard to see what difference his suggestion makes.

## V. Summary and concluding remarks

This paper has put forward a critique of the naturalist version of the Desire-based Reasons Model. It first set the scene by spelling out the connection between naturalism and the Model. After this, it introduced Christine Korsgaard's circularity argument against what she calls the instrumental principle. Since Korsgaard's target, officially, were non-naturalist advocates of the principle, the paper showed why and how the circularity charge can be extended to cover the naturalist Model. Once this was done, the paper went on to investigate in some detail the different ways of responding to the circularity challenge. It argued that none of these responses succeeded, at least not without serious costs to their advocates.[16]

## References

Blackburn, S. (1995), 'Practical Tortoise Raising', *Mind*, 104: 695-711

Copp, D. (2003), 'Why Naturalism?', *Ethical Theory and Moral Practice* 6 (2): 179-200

Copp, D. (1995), *Morality, Normativity, and Society*, New York: Oxford University Press

Dancy, J. (2000), *Practical Reality,* Oxford: Oxford University Press

Dreier, J. (2001), 'Humean Doubts about Categorical Imperatives' in. E. Millgram (ed.): *Varieties of Practical Reasoning*, Cambridge, Mass.: MIT Press, pp. 27-47

Dreier, J. (2015), 'Can Reasons Fundamentalism Answer the Normative Question?', in. G. Björnsson, C. Strandberg, R. F. Olinder, J. Eriksson & F. Björklund (eds.), *Motivational Internalism*. Oxford: Oxford University Press, pp. 167-181

Fink, J. (forthcoming), 'The Property of Rationality: A Guide to What Rationality Requires?', *Philosophical Studies*

For all other rules are applied in action as a result of two things: a desire that has them as its object plus the instrumental principle. Hence, were we to question the principle, thus get entangled in a Carroll-type of regress, our practical reasoning would collapse. Not in the sense as Korsgaard describes but in the sense that nothing would *count as a reason* for us. This is the component missing from Railton's argument. Once it is in place Dreier can conclude that the instrumental principle must have a special status among principles of practical reason. But Dreier's argument is premised on two things. First, he takes that a rule can only be applied in action via the instrumental principle. This is just the point made by Korsgaard: that the instrumental principle is the principle of application itself. And, as we saw, there are philosophers who deny this assumption. Second, as Dreier also admits, there may other principles that have the same status as the instrumental principle. They can function both as a supplement as well as an alternative to the instrumental principle.

[16] [Acknowledgments]

FitzPatrick, W. J. (2005), 'The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity', *Ethics* 115 (4): 651-691

FitzPatrick, W. J. (2013), 'How Not To Be an Ethical Constructivist: A Critique of Korsgaard's Neo-Kantian Constructivism', in. C. Bagnoli (ed.): *Constructivism in Ethics*, Cambridge: Cambridge University Press, pp. 41-62

Hubin, D. C. (2001), 'The Groundless Normativity of Instrumental Rationality', *Journal of Philosophy*, 98 (9): 445-468

Korsgaard, C. (1996), *The Sources of Normativity,* New York: Cambridge University Press

Korsgaard, C. (1997), 'The Normativity of Instrumental Reason' in. G. Cullity and B. Gaut (eds.): *Ethics and Practical Reason*, Oxford: Oxford University Press, pp. 215-254

Korsgaard, C. (1998), 'Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind', *Ethics* 109 (1): 49-66

Korsgaard, C. (2003), 'Realism and Constructivism in Twentieth Century Moral Philosophy', *Journal of Philosophical Research*, pp. 99-122

Korsgaard, C. (2009), *Self-Constitution: Agency, Identity, and Integrity*, New York: Oxford University Press

Lenman, J. (2014), 'Moral Naturalism', *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/spr2014/entries/naturalism-moral/ (Accessed: 28/10/2016)

Parfit, D. (1997), 'Reasons and Motivation', *Proceedings of the Aristotelian Society*, Supplementary Volume 71: 99-130

Parfit, D. (2011), *On What Matters*, Vol. 2., Oxford: Oxford University Press

Radzik, L. (1999), 'A Normative Regress Problem', *American Philosophical Quarterly* 36 (1): 35-47

Railton, P. (2003), 'On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action' in. *Facts, Values, and Norms: Essays Towards a Morality of Consequence*, New York: Cambridge University Press,, pp. 293-322

Railton, P. (2004), 'How to Engage Reason: The Problem of Regress', in. R. J. Wallace, M. Smith, S. Scheffler and P. Pettit (eds.): *Reasons and Value: Essays on the Moral Philosophy of Joseph Raz*, Oxford: Oxford University Press, pp. 176-201

Rawls, J. (1971), *A Theory of Justice*, New York: Oxford University Press

Ridge, M. (2014), 'Moral Non-Naturalism', *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2014/entries/moral-non-naturalism/ (Accessed: 28/10/2016)

Rosati, C. S. (2003), 'Agency and the Open Question Argument', *Ethics* 113 (3): 490-527

Scanlon, T. M. (1998), *What We Owe to Each Other*, Cambridge, Mass.: MIT Press

Scanlon, T.M. (2014), *Being Realistic about Reasons*, New York: Oxford University Press

Schroeder, M. (2005), 'Cudworth and Normative Explanations', *Journal of Ethics & Social Philosophy*, 1(3): 1-28

Schroeder, M. (2007), *Slaves of the Passions*, New York: Oxford University Press

Tanyi, A. (2007), *An Essay on Desire-Based Reasons Model*, Doctoral Dissertation, Central European University, <http://politicalscience.ceu.edu/sites/politicalscience.ceu.hu/files/basic_page/field_attachment/attilatanyi.pdf > (Accessed: 24/10/2016)

Tanyi, A. (2009), Desire-based Reasons, Naturalism, and the Possibility of Vindication: Lessons from Moore and Parfit.' *Polish Journal of Philosophy* 3 (2): 87-107

Tanyi, A. (2010), Reason and Desire: The Case of Affective Desires.' *European Journal of Analytic Philosophy*, 6(2): 67-89

Tanyi, A. (2011), Desires as Additional Reasons? The Case of Tie-Breaking.' *Philosophical Studies* 152 (2): 209-227

Velleman, J. David (2000), 'The Possibility of Practical Reason' in. *The Possibility of Practical Reason*, pp. 170-199

Wedgwood, R. (2002), 'Practical Reasoning as Figuring Out What is Best: Against Constructivism', *Topoi* 21 (1-2): 139-152

Wedgwood, R. (2005), 'Railton on Normativity', *Philosophical Studies* 126 (3): 463-479

Williams, B. (1973), 'A Critique of Utilitarianism', in. J. J. Smart. and B. Williams: *Utilitarianism, For and Against*, Cambridge: Cambridge University Press, pp. 77-151