Korsgaard and the problem of teleology

Alvaro Rodriguez-Gonzalez Barredo

1. Introduction

The fourth lecture in Christine Korsgaard's seminal work, *The Sources of Normativity* (1996),[1] ends by launching a direct challenge. J. L. Mackie, Korsgaard claims, "is wrong and realism is right." The founder of modern moral error theory failed, to her judgment, to take into account that there in fact are "queer" entities in the world that bring objective values with them: "They are people, and the other animals" (SN: 166). Throughout her later oeuvre, Korsgaard would double down with regards to the central role animals – human and non-human – have for her theory of normativity and ethics, further developing a teleological picture of organisms to back these claims. In this paper, I move to take Korsgaard's claim at face value, and test how her conception of living beings as teleologically ordered helps her at fending off the challenge from moral error theorists. I will argue that, although her account suffers from some major vulnerabilities against moral error theory, a more robust commitment to teleology can dispel some of these problems.

I begin by presenting an overview of Korsgaard's constitutivist theory of normativity. She famously argues that the normativity of actions cannot be properly explained either by appealing to objective mind-independent properties, nor by "empiricist" desire-driven models of our psychology. Only internal, constitutive principles of action can be truly normative. Afterwards, I contend that these principles cannot stand on their own without what I go on to call "bridge theses", which link teleological judgments about the world with modal claims about action. I then move on to present Korsgaard's arguments in favour of teleology and conclude that they fall short: moral theorists can easily dismiss them without acquiring new commitments. Finally, I present what sort of teleological conception of living beings could potentially withstand objections from moral error theorists.

[1] I will be using the following abbreviations for Korsgaard's works: *Creating the Kingdom of Ends* (1996a) (CKE); *The Sources of Normativity* (1996b) (SN); *The Constitution of Agency* (2008) (CA); *Self-Constitution* (2009) (SC); *Fellow Creatures* (2018) (FC).

2. Constitutive principles and normativity

Korsgaard's argument about the normativity of our actions is highly controversial and it has faced numerous challenges (Enoch, 2006) (Timmermann, 2006), the merits of which I will not address in this paper. By the end of my argument, I will hope, not to have proven her right or wrong, but to have uncovered a particularly faulty quirk in her picture of normativity, and one easier to address than more substantive disagreements about the constitutivist project.

Mirroring Kant's own exposition methods, Korsgaard finds the discussion around normativity to be best characterized by two opposing views, both of which fall short in what the other one manages to accomplish: an empiricist position and a rationalist opponent[2] (CA 1: 30-1). In order to get over this antinomy, a solution will be right to spot what both sides of the argument have got wrong about the normativity question, and what brings them, in that way, together. Korsgaard's main piece of insight about this quarrel is that the source of normativity cannot lie outside actions themselves, be it in objective values to be found "out there" or in a feature of human psychology; it must be a *constitutive* standard, part of what makes an action *be* an action. With a constitutive standard, "there is no further room for doubting that [it] has normative force" (SC 2.1.3), for a failure to, at least, attempt to follow such a standard will amount to the activity in question not being done. If you say you are building a house and, rather than laying down bricks and mortar, you get your fishing rod and go to the river, you are not merely working towards a very unorthodox house – you are fishing.

But why should it be this way? Is this the best way to respond to the normativity question? Following her more recent writings on the matter[3], I will begin by presenting her discussion about the normativity of the instrumental principle, and then I will assess how this argument applies to any putative principles of practical reason. The instrumental principle commands that, if one desires, or ought to, or wills to achieve certain end E, then one has reason, or ought to, undertake means M, where M either

[2] In *The Sources of Normativity,* Korsgaard distinguishes a third view that contrasts too with her own Kantian proposal: "voluntarist" accounts, like those of Hobbes and Pufendorf (SN 1.3.). However, they fare the same problem as empiricist theories, only in a more obvious way, so we need not go over them explicitly in this exposition.

[3] However, she does not seem to have retracted in any major way from her arguments in *The Sources of Normativity* (SC 1.4.8), so I will keep using references to that earlier work when they help elucidate some of her claims. If some discrepancy arises, the later books should, at any rate, take precedence.

constitutes the end E or is conducive to it (CA 1: 27; SC 3.1.2). Self-evident as this principle may be, earning the status of "analytic" according to Kant (SC 4.3.1), the question of its normativity can still arise; it is not obvious at all how we are supposed to explain that failing to carry out this principle is a specifically *irrational* deed, and not simply imprudent or odd behaviour.

Korsgaard starts off by discussing Hume's account of rationality, which serves as flagship for "empiricist" theories at large. The core commitment of this doctrine is that it is no more *rational* for one to choose any given end over the other, reason has no bearing over such matter, which is up to desire alone (CA 1: 38-9). This may seem to have no clear implications over the instrumental principle at first glance, but Korsgaard asks us to consider what we should say transpires when somebody violates the instrumental principle. Somebody, say, resolves to go to the dentist, and yet is overcome with terror the second before entering the clinic, fleeing home with their teeth unchecked and an embarrassing call ahead, failing to have taken the means to their ends.

Now, the desire to flee the dentist was no less real than their prior resolution to go in the first place. The instrumental principle, in this view, operates as a fact of human nature; whenever we act, some desire must be present so that it explains why we did so (SC 4.1.1).[4] So, two choices[5] remain: either nobody is at any point truly irrational, because at any minute they are acting on their desires, all of which are equally as rational as the next one; or there is a sense in which the resolve to go to the dentist is "more truly" what the person wanted to do than fleeing. But, if that is so, either this is a purely psychological matter, and – implausible as this is (CA 2: 76) - our cowardly fellow simply misjudged what they "actually wanted to do", or we must deny our initial premise and accept that some ends are more rational than others. In the first case, everyone is, strictly speaking, rational at all times once more, because they are just mistaken about what they want; in the second case, we have veered into rationalism.

Rationalism, or substantive realism (SN 1.4.4), "dig[s] in [its] heels" (SN 1.4.2) by accepting that there are some ends that simply are rational on their own (SC 4.2.1). The difficulty in trying to ground a practical principle like this, however, is not any the less

[4] "What looks like the principle of instrumental reason turns out simply to be a *description* of the inevitable effect that a certain kind of judgment has on the human will" (SC 4.1.1).
[5] There is the option of theoretical error too (CA 1: 40), but that would not be irrational either. If we drink a glass of bleach because we believed it was water, we are not so much irrational as very unlucky.

formidable[6]. Let us suppose that there are such objective facts about rationality, "out there" as it were. For you to conclude, case in point, that you ought to take the means to your ends just by incorporating such principle into your line of reasoning, you may try committing to the thesis "whoever wills the end, wills the means", but that binds you with logical necessity, taking us back where we were with the empiricist account, you would *always* be following the imperative, irrationality turns into impossibility (CA 1: 50). On the other hand, if you accept the true statement that "whoever wills the end, wills the means, insofar as he is rational", and if we assume that everyone ought to be rational, then we find that, for you to put two and two together, you need a prior commitment to the instrumental principle. The instrumental principle is a means to the end of being rational, so the objective end of rationality can only kick off if you accept that you ought to take the means to your ends – which was what needed proof (SC 4.2.2).

The takeaway from this argument is not exclusive to the instrumental principle, it is only all the more shocking that such an elementary principle would prove so hard to see its normativity grounded. But consider now a moral principle, a categorical imperative, if you will (SC 3.1.3). If we follow the empiricist route, and reason is an effete faculty of the intellect, then we are once more stranded between the trivial response of giving our rational checkmark to every action one might do, and the implausible thesis that what we want is not actually what we want, but some other thing[7]. Crucially, if we try the rationalist way instead, the question will always remain: on what principle do we apply our normative principle of choice? For us to incorporate it, it needs to somehow be already normative for us before we accept it as our principle, and that, Korsgaard claims, will not do (SC 4.2.2).

At this point, Korsgaard's solution should come as no surprise. The misstep empiricists and rationalist share is to approach practical principles as though they were

[6] In *The Sources of Normativity,* the main concern with substantive realism seems to be motivational: "If someone finds that the bare fact that something is his duty does not move him to action, and asks what possible motive he has for doing it, it does not help to tell him that the fact that it is his duty just is his motive. That fact isn't motivating him just now, and therein lies the problem." (1.4.5). Later on, however, Korsgaard appears to disavow this approach, opting for a problem with normativity in the rationalist's conception: "But if we think about normativity, rather than motivation, then we will find that there is something in Hume's complaint" (SC 4.2.1).

[7] Of course, nobody would dare disregard the possibility of an "empiricist" ethical project with just these meagre lines. The argument Korsgaard is making here (vid. especially CA 2 in its entirety) is that not even something like a principle of prudence can be derived from an empiricist stance without very costly psychological commitments, let alone anything resembling a *rational* principle of action that could found morality.

outside constraints on action (CA 1: 56); by doing this, their normativity becomes incomprehensible. However, "if we can identify something as an internal norm, the question why you should conform to the norm answers itself" (CA 1: 61). And so, Korsgaard takes on the enterprise of uncovering the constitutive principles of action. Glossing – inevitably – over the details, she comes to conclude that these are efficiency and autonomy (SC 5.1.3). An action is to be distinguished from mere movements, and it is so by virtue of its constituting an *agent*, which, being something more than the impulses and desires transpiring under her purview, has to constitute herself as the cause of her ends. In constituting herself as a cause (i.e., in being efficacious) she must follow the hypothetical imperative; and, in being *herself* the cause (i.e., in being autonomous), and not *something about her* causing a movement, she must follow the categorical imperative (SC 5.1.1).

3. The bridge theses

After this elementary assessment of Korsgaard's theory of normativity, we have not really arrived at any proposition that could evidently belong to the scope of metaethics, or at any programme that could guide us in doing so. Let us consider more carefully what we have managed to secure so far:

> (**P1**) If action is possible, then it is governed by the constitutive principles of efficacy and autonomy.

This first thesis follows directly from our previous discussion. As mentioned earlier, I will not discuss how plausible or defensible it is. Some doubt may be cast, though, over what "governed" means here. An exhaustive and completely clear explication of this term may be out of order, but we can say two things. First, it expresses a sort of necessity that is neither logical nor metaphysical; it is, indeed, a necessity that can fail to obtain in a sense: "The possibility of self-government essentially involves the possibility of its failure" (CA 1: 60). But the principles truly govern the activity[8] since, even when the activity has somewhat failed, for it to properly belong to the activity-type in question, it must have been directed by it: "Although it is not true that you are not performing an activity at all unless you do it precisely, it is true that you have to be *guided by* the precise

[8] This is a quirk about Korsgaard's terminology, where it seems that "action" is a type of "activity" - this may sound unintuitive, but not much seems to hang on this.

version of the activity to be doing the activity at all" (SC 2.1.5). This takes us straight into the second comment that may be made; only when something is "governed" by constitutive principles can it be said with any objectivity that it is a "defective" member of its class, rather than belonging to a different class altogether (SC 2.1.8; 8).

(**P2**) If we, as self-conscious beings, must do X, the constitutive principles that X is governed by are normative for us.

This statement does not follow straight from the former discussion, but it seems like the most plausible reading of the transition from the merely internal description of an activity to *our* being held to certain standards. This, I take, is the implicit thesis that motivates statements such as "[h]uman beings are *condemned* to choice and action", and "[i]t is our *plight*: the simple inexorable fact of the human condition." (SC 1.1.1). Now, if there were problems beforehand with the specific sense in which a principle "governs" an activity, the way in which we are to understand this 'must' is at least as problematic (Enoch, 2006: 188). It cannot be a 'must' of obligation, or the whole argument becomes irredeemably redundant: if we already know that we 'must' do some things, the normative question is thereby rendered moot. But then again, it cannot express "logical", "causal" or "rational necessity" (SC 1.1.1). We can already guess that it will be, essentially, the same sort of necessity we drew from (**P1**), but we may let it rest for the time being and come back to it later. Similarly, I will address the reasoning behind the qualification "as self-conscious beings" at a later point, though the consequences of leaving such a premise unqualified may already be apparent.

From these two theses, if the argument is to qualify as a proper and positive reply to the normative question, the goal will be to arrive at something like this conclusion:

(**C**) The constitutive principles of efficacy and autonomy are normative for us.

Even though the goal may seem at hand, I am going to show that the argument requires somewhat of a detour to reach its conclusion. Korsgaard still needs to show that (i) action is possible, and (ii) we must act. This appears close to trivial, but we must be careful not to confound the terms here. That action, in the relevant sense, as an activity that constitutes agents through principles of efficacy and autonomy, is possible is not obvious (SC 5.2.1), even if it undoubtable that *something* we call action is. Similarly, that we *must* act in the precise sense that is required here, we need to show that we have a

necessity to act that is not logical and is not causal, so it leaves out some space for our not doing it quite right. Correspondingly, in the remainder of this section, I will argue for two "bridge theses" that will make those assertions possible.

(**BT1**) If animals are teleologically constituted, action is possible.

(**BT2**) If animals are teleologically constituted, they must act.

These are "bridge theses" in the sense that they connect a teleological matter of fact with a correlative modal condition about action.[9] In what follows, I will contend that the two theses are to be found, at least, implicitly in Korsgaard's account, and that they account for the relevant senses of 'action' and 'must' warranted by the argument.

a. First bridge thesis: the possibility of action

The claim that the possibility of action should be linked somehow to teleological constitution will strike many as weird and uncalled for. Korsgaard, however, has a pressing issue with regards to her conception of action that she tries to remedy through ostensibly teleological means. If action is a self-conscious commitment to a means-and-end package, only adult humans can be said to properly act, and just sometimes at that (SC 5.3.3). Yet, we may want to say that other animals act too, without biting the bullet of imagining squirrels and hogs applying the categorical imperative to their movements and acting from a motive of duty. As a result, Korsgaard identifies three features of action through which we can extend it beyond rational humans whilst maintaining the constitutive standards that govern it.

First, as a corollary from the efficacy principle, an action can fail in the sense that it can be infelicitous and not achieve its end. "If the rock runs into an obstacle and stops rolling before it gets to the bottom of the hill, it has not failed. But if the cockroach does not make it under the toaster, he *has* failed" (SC 5.4.4). We, however, do not necessarily attribute a mental state to the cockroach by which it decides to go under a toaster. But if we judge the cockroach as a teleologically arranged being, that is, as having the function of keeping itself alive – maintaining its form (SC 2.2.1) - then we can assign intentional

[9] In Korsgaard's work, and in this paper, "teleology" can be broadly equated with "functionality", or "teleonomy" (Mayr, 1974); more precisely, it most closely resembles Kant's conception of "internal teleology" (Ak 5: 368-9). It does not refer to transcendent ends particular beings aim towards, but to a thing's having a certain form, and being of the kind it is by virtue of said form, such that a certain function can be said to follow from it (SC 2.1.1).

content to an action even without a mediating thought (SC 5.4.5). It will not be too difficult to figure out what a butterfly is doing in flying away after you maliciously nip the flower it had alighted on.

Secondly, an action can fail in the sense of being inadequate. If a moth flies into our lightbulb (and into its doom), it is not that it has failed to achieve what it sought so much as that it has followed a policy of action that, in the context, was inadequate for its function; the action "can lose [its] self-maintaining function" (SC 5.5.3). This, albeit less intuitively, follows from the principle of autonomy. In the case of animals, an action is autonomous, Korsgaard argues, if it follows from their *instinct*, understood as a principle of action they have *as an entire and whole living being,* according to which they translate incentives into actions (SC 5.6.3; FC 3.2.4). This instinct makes only sense as a category subject to evaluation if animals, as Korsgaard argues, have self-maintenance as their function.

Finally, action must be guided by valenced perception. A clock has a function, and its movements can be subjected to evaluative judgment, but it does not *act*; it is only those beings that react to a representation of the world (SC 5.4.2), and a representation by which they perceive things as being evaluatively loaded at that (FC 2.1.7), that can be said to act[10]. This evaluative load refers back to the function of the perceiving being; it is with that criterion in place that an incoming object may be regarded as good, bad or none of the above.

The three features of action depend directly on the teleological arrangement of the agent. Without a teleologically constituted being whose function is self-preservation – or preservation of its form – none of the three characteristics could hold water; there would be no criterion for its movements failing or succeeding, and there could be no valenced perception.

This much is clear for non-human animals, but one may resist this argument by noting that the teleological crutch does not appear necessary when speaking of humans.

[10] It is far from clear where this caveat is meant to fit amid the other two features of action. Whereas intentional content follows from efficacy and autonomy is expressed by means of principles through which the whole entity, and not one of its parts, effects a change into the world, perception seems eerily ad hoc, especially since the unwanted conclusion that plants may check all the boxes looms close (SC 5.4.6; FC 2.2.3). Though such a view, attributing agency to plants, is not unheard of (Alvarez and Hyman, 1998: 243ff.), that would have disastrous consequences for Korsgaard's tight link between agency and morality.

After all, we choose the principles of our actions, and we can – at least, in principle – abstract from whatever our biological impulses tell us about how good or bad an object is. With other animals, the need for these judgments arises since we cannot claim that they have the mental states we are certain we do have.

Although (**BT1**) is certainly more clearly motivated when attempting to address non-human animals – which, given Korsgaard's interest in proving their moral relevance, is not a minor issue[11] - there are, I countenance, at least three reasons why we cannot dismiss it out of hand when assessing human action. First, Korsgaard adheres to a broadly Aristotelian classification of lifeforms in her philosophy, where humans as rational animals are distinguished from the rest of animals by the way in which we maintain our forms throughout time, that is, through creation and upholding of practical identities (SN 4.3.7; FC 3.3.4), which then trickles down, shaping (to some extent) every one of the processes we have as living beings (FC 5.2.3). There are two options in interpreting what makes human action distinctive, then. Either it is a special form of animal action, whereby we, uniquely, choose the principles that other animals have to accept as naturally given, or it is something completely different from animal action – despite, accidentally, sharing their essential features. The first option seems both more parsimonious and more in line with Korsgaard's general outlook.

Moreover, even though we may bypass teleological requirements when convincing ourselves that *we* (that is, *I*) can act, since we have a first-person access to the contents of our wills, it is not so clear that we can navigate Korsgaard's sharply distinguished first and third-person standpoints in recognizing our fellow humans as agents without teleological judgment. In a sceptical point that should ring familiar to Kantians[12], there can be an abyss between our outward experience of other people and the unique experiential point of view we ascribe to them. We can bring in other famed responses to the other minds problem to solve this, but if we are already in the predicament of judging animals in a certain way, it would be rather strange to make an

[11] Indeed, the moral relevance of pleasure and pain stems form its being a part of functioning well (FC 9.4.4.)
[12] Saunders (2016) argues that transcendental idealism makes the problem of recognising other free agents intractable. Though Korsgaard's metaphysics, by virtue of being far less defined, makes her situation less dire, this problem resonates with her framework: "I am conscious of the moral law, which tells me that I ought to act in a certain way, and this reveals that I can. Once more, this is how my freedom is revealed to me. The problem concerns how this applies to the freedom of others" (2016: 170-1)

exception and ask for special help from other philosophical systems when dealing with the *animal rationale*.

Finally, a case can be made that not even our own identity is as clear-cut as we could imagine it to be. Korsgaard – rightly, in a way – takes it as a given that some of our actions are inscribed into and look towards overarching projects that presuppose a practical identity over time: "[s]ome of the things we do are intelligible only in the context of projects that extend over long periods" (CKE 13: 371). This, indeed, is a keystone in her argument for the universalizability of reasons (SC 9.7.1)[13]. But even if we agree that our practices are such that we must assume that our unity throughout time somehow obtains, it may be suspicious to simply establish by fiat what surely merits some explanation, especially when we have one available[14]. If we are animals, and we are teleologically arranged, what it means for us to be humans is, precisely, to maintain our forms in time using our principle-choosing faculty of reason.

In conclusion, given Korsgaard's idiosyncratic conception of action, the question about its possibility is not banal. It is not clear that there can be something in the world that brings changes forth according to the principles of efficacy and autonomy as described in her theory. Yet, if there exist teleologically arranged beings, and we count ourselves among them, we can understand how this is possible; and, in want for another explanation, since it is hard to deny that attributing action to other animals in the way required here relies on such a conception, (**BT1**) presents itself as the most plausible and parsimonious means to ground such a possibility.

b. Second bridge thesis: the necessity of action

Let us recall what peculiar 'must' we are after in elucidating the necessity that binds us to acting. It cannot be that we are morally bound to act without incurring in a gross *petitio principii.* On the other hand, it cannot be, either, that we are logically or metaphysically

[13] A precision needs to be made. Strictly speaking, it is not the temporal unity of the self that is at risk when failing to abide by universal reasons (SC 9.7.2), but what agency is about, anyway, is creating an identity that, by the nature of reasons, cannot be constrained to a single time slice in our lives (SC 9.7.4). But, of course, what "agency is about", self-constitution, can hardly be made sense of if not for our plight to make "wholes" of ourselves instead of dissolving into the hotchpotch of impulses and particularities that we are made up of at any given point. This urge to make ourselves whole runs the risk of verging into the mystical unless we have a specific understanding of why humans are rather more like wholes than like heaps. If we are like the other animals, self-maintaining activities, then the meaning is, if not clear, as clear as it can be.
[14] This is a point of caution that could be made, in general, about transcendental arguments, like the one that is implicit in this position. I will have more to say about them in the next section.

wont to act. There are many things that constrain us in that way. We cannot occupy several locations at the same time, we cannot exist and not exist simultaneously, and yet we would be hard-pressed to extract normative guidance from those facts[15]. The way in which we 'must' act is such that we can fail; not fail *to act*, but fail to act *properly*, be more or less successful at it (SC 1.4.8).

At this point, it will hardly need much arguing to sustain that this is precisely the role that functions play in Korsgaard's teleological conception. What it means to be a living being is to engage in such an activity for the duration of its life (SC 2.2.1). A parrot *must* persist in being one, in making himself one; if he stops doing it, we no longer call him a parrot – he becomes an ex-parrot. But he can certainly do it poorly; he might fail to get food, he might not be up to standard with his flight. If we are animals who maintain themselves by adopting a practical identity (SC 2.4.1), it becomes clearer in what sense we are "doomed" to act.[16]

This point may seem suspiciously similar to a common criticism of the constitutivist project, especially as expressed by David Enoch (2006), and then adopted by some moral error theorists (Streumer, 2017). If we have no (normative) reason to act, the argument goes, action cannot by itself provide us with reasons. Am I, then, forsaking Korsgaard's constitutivism? No. The 'must' of teleology is not of the kind that would please a realist, it is a description of the special activity that characterizes living beings. Moral oughts come entirely from within the purview of rational action and its own normativity, and animals, indeed, do not have duties (FC 4.2.6). But without a fact of the matter as to why and in what sense we are bound to act, all our talk about constitutive standards remains purely hypothetical. It is only because our nature is such that we need to act, and because we keep being what we are through acting, that we find ourselves bound by the internal rules of acting. This, nevertheless, is to say nothing about realist objections – the question, why not act half-heartedly (Enoch, 2006: 189) in rebellion against our animal inheritance, is still available to the realist.

[15] "If you're unconvinced, think of all the features plausibly considered nonoptional in this non-normative way: Do you think normativity emerges from all of them?" (Enoch, 2006: 190)

[16] "It is not as if you could simply subtract "rationality" from a human being, and you would be left with something that functions like a non-human animal. A non-rational animal, after all, functions perfectly well without understanding the principles of reason, since he makes his choices in a different way. He is "designed" by the evolutionary process to be guided by the ways he instinctively perceives the world, rather than by reason" (FC 5.2.3)

Of course, this bridge thesis applies without loss of generality to all animals, which are distinguished from the rest of living beings by action: "What is distinctive of animals, in other words, is that they carry out a part of their self-maintaining activities through action. They are alive in a further sense that plants, for they spend their lives *doing* things" (SC 5.4.1)[17]. That brings us back to the qualification to (**P2**) I left unexplained. Even though there is a broad sense in which the constitutive principles of efficacy and autonomy "are normative" for all animals, human or not, since we can evaluate the movements of animals (SC 5.4.4), it is only *self-conscious* animals that have anything to say about the principles on which they act (SC 6.1.7), and, thus, normativity strictly speaking – having *reasons* to act – can only apply to them, i.e., to us. Nonetheless, (**BT2**) establishes the specific type of unavoidability that makes action an ineradicable dimension of the human condition in just the right way to ground normativity.

4. Korsgaard's arguments for teleology

I have argued that Korsgaard's argument for the normativity of actions requires two "bridge theses" to be complete. They respond to the role teleology plays in her theory, grounding both the possibility and necessity of action in the relevant senses. Let us recap, then, the argument until now:

> (**P1**) If action is possible, then it is governed by the constitutive principles of efficacy and autonomy.
>
> (**P2**) If we, as self-conscious beings, must do X, the constitutive principles that X is governed by are normative for us.
>
> (**BT1**) If animals are teleologically constituted, action is possible.
>
> (**BT2**) If animals are teleologically constituted, they must act.
>
> (**C**) The constitutive principles of efficacy and autonomy are normative for us.

Evidently, the argument does not follow yet. Two crucial steps are still missing. First, and least importantly, we need to accept this statement:

> (**P3**) We are self-conscious animals.

[17] In this specific fragment, Korsgaard is presenting Aristotle's views, but in the context of the rest of the section (5.4) it is clear that she endorses it.

This should not be too controversial, but we must keep in mind that Korsgaard is making an Aristotelian use of the word "animal" here, but in a minimal Aristotelian understanding of such term, it should be acceptable enough that we do get by thanks to and by means of our representational abilities and capacity for local movement. But some may still harbour doubts regarding such a premise. The "we" that functions as subject here is conspicuously undefined, and it may already rest on a teleological understanding of biological judgments[18]. If that is so, since my argument has revolved around the implicit centrality of teleology in Korsgaard's edifice, I should have no issue in accepting it; nonetheless, I will leave the door open to the possibility of understanding this empirical proposition without appeal to teleology.

With that note out of the way, we can focus on the crux of the matter:

(**T**) Animals are teleologically constituted.

For the remainder of this section, I will present Korsgaard's arguments in favour of this thesis, and argue that they are especially weak against moral error theorists, who do not have to take on new commitments to reject (**T**) as argued for. If this is so, and my argument so far holds, moral error theorists do not even need to charge upfront against Korsgaard's substantive metaethics to reject it, since, without (**T**), her conclusion fails to obtain.

Now, I have said "arguments" in favour of teleology, but one of them may not quite qualify as one. In *Fellow Creatures* (2.2.4-6), Korsgaard considers the doubts that may arise with her introducing teleological parlance in a post-Darwinian era. More precisely, there are two problems: (i) teleological or functional talk may be rendered inapplicable by natural selection, (ii) even if it is not, it seems like the proper subject of teleological judgments should then be the species, rather than the individual. Korsgaard promptly disregards the first point, saying that "evolution does not show us that there are *no* functionally self-maintaining objects. Instead, it shows us how there can be such objects even if no one designed them" (FC 2.2.4). Of course, this is not an argument in favour of teleological descriptions of living beings; if anything, it just establishes that such an approach would not be *prima facie* incompatible with natural selection.

[18] Thompson (2008) makes this point precisely.

Regarding the second point, which is a matter of great controversy in its own right[19], Korsgaard argues that the development of consciousness marks a watershed in which *selves* acquire final goods through their valenced experience of the world (FC 2.2.5). However, these goods are not merely what is experienced as such, since we routinely regard harmful things as good-for-us (FC 2.2.6). So, once more, Korsgaard is not making an argument for teleology, but presupposing it. If there is a distinction between what we perceive as good and what *actually* is good-for-us, that is, what actually promotes the unity of our selves, then there must be a matter of fact as to our functional layout. But this is precisely what needs to be argued for, it is what (**T**) asserts, and it cannot be immediately derived from the fact of valenced experience.[20]

To find Korsgaard's actual argument for teleology, we must return to *Self-Constitution*. There, she appears to present a transcendental argument in favour of a teleological conception of the world: "We need the world to be organized into various objects in order to act […] An object is identified as a locus, a sort of force field, of particular causal powers, and the causal powers in question are identified as those we might either use or have to work against." (SC 2.3.2) When *we* act, furthermore, we regard ourselves as causes in the world, thus, teleologically as well (SC 2.3.3; 5.2.4). Now, the reasons behind this transcendental argument in particular are extremely problematic. First of all, the dichotomy between a view of the world where causality only occurs between whole states of the universe and one where objects are causes through their Aristotelian form is false (Mumford and Anjum, 2011) (Williams, 2019). Furthermore, some have argued against the requirement for hopes about the hospitality of the world as a condition of action (Freyenhagen, 2020). And even if we grant as much, it is hard to dispel the aspect of circularity behind this approach, since, as shown before, the plight for action is itself teleologically grounded through (**BT2**).

But let us concede all of this. Perhaps we ought not to expect a foundationalist construction on Korsgaard's part, and we can make sense of a certain reflective equilibrium between a teleological worldview and an internal awareness of our need for

[19] On the units of selection problem see, for example, (Sober, 2000).
[20] Korsgaard's view of pain as a reason, in general, seems to rely on a view of organisms where it can make sense to speak of objectively good- or bad-for-them effects. (SN 4.3.6.) See note 11.

action.[21] Even then, I claim, the argument will not stand scrutiny against a moral error theorist, whom, if we are to take the closing remarks of the *Sources of Normativity* seriously, should precisely be the target of her exposition. Let us say Korsgaard's transcendental argument is of this form:

(1) We can only act if the world is teleologically organized.

(2) We must act.

(3) Therefore, the world must be teleologically organized.

As Robert Stern points out, however, the conclusion of a transcendental argument can take different shapes (2000: 10). The strongest form it can take is that of a *truth-directed* argument. The problem is that such a strong version, especially for the case in point, seems outrageous (ibid.: 65). Either we commit to an extremely strong form of (1), where any movement of ours, and not Korsgaard's technical conception of action, requires teleological organization; or we take the Panglossian view that our projects cannot be frustrated by the world. Since Korsgaard draws directly from Kant's remarks on practical postulates in her argument (SC 5.2.5), we may amend it into a belief-directed one:

(3') Therefore, we must believe the world to be teleologically organized.

This is more plausible now. A last remark is in place, however. Let us remember that the kind of action that elicits teleological judgment from what Korsgaard has presented is the technical sense she attributes to it, that is, action governed by the constitutive principles of efficacy and autonomy. If we can accept (2) without circularity, as mentioned earlier, that must be motivated by some internal awareness of our plight to act and, presumably, it will not be the same 'must' of (**P2**) and (**BT2**). That 'must', we have established, had an irremovable teleological element to it. Then, what is (2) actually saying? The context in which the normative question arises is one of justifying the claims we know morality impinges on to us (SN 1.1.1). So, even if this is not to be read as a properly normative 'must', we are motivated to regard the world as teleologically organized insofar as the question of normativity arises and asks us for an answer about its sources.

[21] But this would not overrule the need for (**BT2**). Same as with (**BT1**), that we find an internal motivation for lending credence to a certain proposition does not absolve us from explaining its possibility.

The problem, then, is that the argument lacks any additional force for the moral error theorist. The core tenet of the moral error theory is that moral thinking involves systematically false beliefs (Olson, 2014: 8). If all we have got going for teleology is that a *belief* in it is entailed by our moral thinking, will the moral error theorist not discard it with the same ease as she does objective value? This is a supposed feature about the world that is meant to complete an argument for normativity, and which has nothing going for it except our purported need for normativity. The only cost a moral error theorist needs to pay for this is accepting that we incur in systematic error when we think morally – which is precisely what a moral error theorist already believes. Even if the point were pressed that the transcendental argument somehow proves we have no choice but to believe in a teleologically ordained world, moral error theorists have been happy to accept as a nice policy to follow that we stick to false beliefs (Olson, 2014: 193), or even that we cannot believe the moral error theory even if it is true (Streumer, 2017: 154), so that would not pose too grave a problem either.

Korsgaard's theory presents, then, a major vulnerability. For her argument for the normativity of our actions to work, she needs to give reasons in favour of the teleological constitution of animals such as ourselves; but the only reason she gives for that emanates from the very moral interests that motivate the query for normativity in the first place. By doing this, not only does she introduce a potential circle at the very core of her proposal, but she makes it so someone who denies her premises has nothing to lose but a theory they disagreed with in the first place. If somebody already thinks that we are legitimated in believing in the claims of our common moral practices, they may follow Korsgaard. J. L. Mackie, however, will simply quip: "this ingrained belief is false" (1977: 49).

5. The case for theoretical teleology

Is that it for Korsgaard's proposal, then? I would not think so. The thread holding this last point is that, when teleology is exclusively motivated by an already moral thinking, there are no added costs for the moral error theorist in denying it. Teleology plays, in Korsgaard's argument, a purely internal role, it is brought forward by a moral interest, and its only function is to ground normative claims. By not peeking outside of the practical standpoint, teleology lacks any anchor to which it can hold when someone is content with declaring morality a systematic source of false beliefs. But if teleology is

justified independently, the shockwave that would result from denying it for metaethical purposes may deter some from doing it.

A strongly metaphysical, Aristotelian-style[22] account of teleology will not be needed either. It suffices that there are theoretical[23] reasons that count in favour of accepting (**T**), which will imply three things: (i) that there is a sense of function that can be applied to living beings such that their movements can be evaluated; (ii) that there is a sense of function that can be applied to living beings as a whole rather than to their parts; (iii) that it, in particular, describes living beings as having their own self-maintenance as their function.

Whilst the debate is anything but settled, philosophers of biology with views on functionality that can be collegial with these requirements are not scarce. The classic proponent of this functional approach is Mayr (1974), whose idea of 'programs' as applied to living beings introduces an evaluative criterion for states of the organism. From a systems-theoretical point of view, Schlosser contends that functional explanation (as a two-ways causal dependence) can only be ascribed, in general, in the context of "complex self-re-producing" systems (Schlosser, 1998). If there is a sense in which a living being can be established to be a system, one of the features of which is complex self-re-production, perhaps by adding fitness pressures through natural selection can it be defended that self-maintenance itself is a function of the whole; though this will depend on the stance one takes on the units of selection debate. Lewens has argued that talk of functionality in biology is as justified as in artifacts because both kinds of entities undergo (qualifiedly) analogous processes of selection (Lewens, 2004). The analogy will not hurt Korsgaard insofar as action requires perception, and normativity requires self-

[22] The received interpretation of Aristotle's teleology, in any case. Some phenomenological interpretations of Aristotle's teleology have given less metaphysical import to it (Wieland, 1961). More recently, one could focus on one of the two kinds of mechanical inexplicability Ginsborg finds in Aristotle to characterize a less costly Aristotelian model (Ginsborg, 2004).

[23] Drawing on the classical theoretical/practical distinction, by this I simply mean reasons that are not motivated directly by our moral practice. Even though the borders may be fuzzy, I will focus on reasons given by the scientific pursuit of biological knowledge, and assume that it is a paradigmatic enough case to not cause too much trouble, without entering into the debate about the kind of normativity that governs scientific inquiry (Cf. Thorstad, 2022). Furthermore, my choice in seeking "reasons that count in favour of" teleology can be contested as well (Olson, 2014: 167). With that, I am not endorsing any specific view on the normativity of belief. If (**T**) (for instance) explains X set of facts, or states of affairs, or propositions, then in rejecting (**T**) one thereby rejects the explanations for X – this is a *pro tanto* cost that will need to be weighed with the alternatives. My argument in section 4 aimed to show that the cost of rejecting (**T**) for an already committed moral error theorist is, effectively, zero.

consciousness aside from teleology; and this, once again, introduces an evaluative understanding of an animal's doings against the background of its fitness. Moreno and Mossio (2015) have defended a notion of biological agency in terms of functionality. Finally, very recently, Novick (2023) has argued that structuralism and functionalism can coexist and complement each other as explanatory strategies. This sits well with Korsgaard's generally constructivist approach [24], while still providing an external motivation for teleology.

In summary, even though we are a long way from settling whether functional-teleological approaches to biology are sustainable, it is, at least, plausible that they are. If that is so, there can be hope for Korsgaard's charge against the sceptic. If biology tells us that animals are, indeed, somewhat "queer", the moral error theorist will not have as easy of a time toppling down the constitutivist project. This is not to say that biology could ever establish that organisms are *as queer* as moral error theorists would need them to be, this depends on whether we are on board with the rest of Korsgaard's arguments; but the project becomes more robust by not being exposed to having its vital premise discarded off-hand.

6. Conclusion

I have argued that teleology plays a central role in Korsgaard's argument for the normativity of actions. Without an understanding of animals by which they can be said to have self-maintenance as their function, neither is it clear that her conception of action could be *possible*, nor would we be able to say that we *must* act in the relevant sense. Though this link to teleology is especially pressing for non-human animals – which is of the utmost concern to Korsgaard's project –, I have contended that it extends to humans as well. Korsgaard then tries to argue for a teleological worldview by means of a transcendental argument, grounded on the conditions of possibility of our actions, which inflicts a fatal blow on her edifice: moral error theorists have no problem accepting that our moral practices entail false beliefs. Finally, I have shown that there can still be hope for Korsgaard's argument insofar as teleological approaches to organisms still have theoretical plausibility in the philosophy of biology.

[24] "[…] in order to conceive of ourselves as knowers, we have to conceive of the world as one public object; we have to *construct* our conception of it that way" (SC 9.7.5).

7. References

Álvarez, María and Hyman, John. (1998) "Agents and their actions", *Philosophy* 73(2), pp. 219-245.

Enoch, David. (2006) "Agency, Schmagency. Why Normativity Won't Come from What Is Constitutive of Action", *The Philosophical Review* 115(2), pp. 169-198.

Freyenhagen, Fabian. (2020) "Active Irrespective of Hope", *Kantian Review* 25(4), pp. 605-630.

Ginsborg, Hannah. (2004) "Two kinds of mechanical inexplicability in Kant and Aristotle", *Journal of the History of Philosophy* 42(1), pp. 33-65.

Kant, Immanuel. (1900-) *Akademieausgabe von Immanuel Kants Gesammelten Werken*.

Korsgaard, Christine M. (1996a) *Creating the Kingdom of Ends.* CUP

Korsgaard, Christine M. (1996b) *The Sources of Normativity.* CUP

Korsgaard, Christine M. (2008) *The Constitution of Agency*. OUP

Korsgaard, Christine M. (2009) *Self-Constitution. Agency, Identity and Integrity*. OUP

Korsgaard, Christine M. (2018) *Fellow Creatures. Our obligations to the other animals*. OUP

Lewens, Tim. (2004) *Organisms and Artifacts. Design in Nature and Elsewhere.* MIT Press.

Mackie, J. L. (1977) *Ethics: Inventing Right and Wrong.* Penguin Books.

Mayr, Ernst. (1974) "Teleological and Teleonomic, A New Analysis", *Boston Studies in the Philosophy of Science* XIV, pp. 91-117.

Moreno, Álvaro and Mossio, Matteo (2015) *Biological Autonomy. A Philosophical and Theoretical Enquiry.* Springer.

Mumford, Stephen and Anjum, Rani Lill. (2011) *Getting Causes from Powers*. OUP

Novick, Rose. (2023) *Structure and Function*. CUP (Cambridge Elements)

Olson, Jonas. (2014) *Moral Error Theory. History, Critique, Defence.* OUP

Saunders, Joe. (2016). “Kant and the Problem of Recognition: Freedom, Transcendental Idealism, and the Third-Person”, *International Journal of Philosophical Studies* 24(2), pp. 164-182.

Schlosser, Gerhard. (1998) “Self-re-production and Functionality. A Systems-Theoretical Approach to Teleological Explanation”, *Synthese* 116, pp. 303-354.

Sober, Elliott. (2000) *Philosophy of Biology*. Routledge.

Stern, Robert. (2000) *Transcendental Arguments and Scepticism.* OUP

Streumer, Bart. (2017). *Unbelievable Errors*. OUP

Timmermann, Jens. (2006) “Value without Regress: Kant’s ‘Formula of Humanity’ Revisited”, *European Journal of Philosophy* 14(1), pp. 69-93.

Thompson, Michael. (2008). *Life and Action. Elementary Structures of Practice and Practical Thought*. HUP

Thorstad, David. (2022) “There are no epistemic norms of inquiry”, *Synthese* 410.

Wieland, Wolfgang. (1961) “Das Problem der Prinzipienforschung und die aristotelische Physik”, *Kant-Studien* 52(1-4), pp. 206-219.

Williams, Neil E. (2019) *The Powers Metaphysic.* OUP