

# Perverse Reasons

FRANCESCO ORSI

## Abstract

For an agent to be motivated by a normatively perverse reason is to be motivated by a normative or evaluative thought as such which, if true, would count as such against the action that it motivates the agent to perform, or against the attitude that it motivates the agent to take. For example, that an action is morally wrong or prudentially bad counts, as such, against performing the action. When the thought that an action is morally wrong or prudentially bad (bad for me) motivates me as such to perform the action, my motivating reason is normatively perverse. If being motivated by normatively perverse reasons is possible, then what, if anything, is wrong about it? I present and reject some accounts of what may be wrong about normative perversity (wrong reasons, malfunctioning attitudes, practical irrationality, instability, evaluative ignorance). In the course of this discussion some desiderata emerge. Then I defend the suggestion that normative perversion is socially undesirable, in that it undermines certain valuable interpersonal and intrapersonal relations. Entering and maintaining these relations is constitutive of valuing people as beings to whom reasonable justification is owed. I show how this account satisfies the desiderata.

## 1. Introduction

In this paper I will discuss the phenomenon of acting on perverse practical reasons. My aim is to answer a question that, to my knowledge, has received casual rather than systematic attention in the otherwise rich literature on this topic, namely: What is peculiarly wrong with acting on perverse practical reasons?

In what follows, practical reasons are taken to cover reasons for actions and for attitudes. To understand the idea of perverse reasons, we need a distinction between motivating and normative reasons. Motivating reasons are considerations in the light of which an agent acts or desires something. Motivating reasons are contrasted with normative reasons, that is, considerations which count in favour of acting in a certain way or desiring certain things. Obviously enough, not all motivating reasons are normative reasons, and vice-versa. We can be motivated by considerations that do not count in favour of what we do, and some considerations that count in favour of doing a certain action might fail to motivate us.

I will be interested, specifically, in motivating reasons which are *normatively* perverse. As a first approximation, for an agent to be motivated by a normatively perverse reason is to be motivated by a *normative* or *evaluative* thought *as such* which, if true, would count *as such* against the action that it motivates the agent to perform, or against the attitude that it motivates the agent to take.<sup>1</sup> Here are some paradigmatic cases. That an action is morally wrong or prudentially bad counts, as such, against performing the action. When the thought that an action is morally wrong or prudentially bad (bad for me) motivates me as such to perform the action, my motivating reason is normatively perverse. That a certain outcome is overall bad (for example because it produces a lot of suffering) counts as such against desiring its occurrence. When the thought that an outcome is overall bad motivates one as such to desire its occurrence, the reason for one's desire is normatively perverse. That outcome A is worse than outcome B counts as such against preferring A over B. When the thought that A is worse than B motivates one as such to prefer A over B, one's preference is motivated by a normatively perverse reason. The same holds, *mutatis mutandis*, for 'positive' normative or evaluative thoughts. That an action is morally good, or prudentially good, counts intuitively as such in favour of performing it. But when the thought that an action is morally or prudentially good motivates one, as such, not to perform it, one's motivating reasons are likewise normatively perverse. This seems an initially plausible way of articulating the idea that a normatively perverse agent is attracted by the bad *qua* bad, and repelled by the good *qua* good.<sup>2</sup> (In sections 2.1 and 2.4 I will sharpen this initial characterisation.)

<sup>1</sup> 'As such' means taken in isolation from a given context. This is important, because e.g. the evaluative thought that an action x is prudentially bad (bad for me) can be a normatively perverse motivating reason for me to do x, even in a context where x being bad for me actually counts as a reason in favour of me doing x (for example, because I deserve to inflict some punishment on myself). In such a case my motivation happens to match the normative import of x being bad for me. But my motivation can still count as perverse, to the extent that I am regarding 'x is bad for me' in isolation from this context (I do not know or care that I deserve to self-inflict punishment) and that, plausibly, when isolated from such a context, x being bad for me does count against me doing x. Thanks to a reviewer for pressing this point.

<sup>2</sup> Normatively perverse reasons are usually discussed in the context of either defending the claim that we can only desire what appears in some respect good to us (the so-called guise of the good doctrine) (Gregory, 2013; Hawkins, 2008; Sussman, 2009; Tenenbaum, 2007, ch. 6; 2018; Raz, 2016)

In what follows, it will be useful to distinguish conflicted vs. unconflicted and impure vs. pure normatively perverse motivation. Perverse motivation is conflicted when a given normative or evaluative thought (say, *x* being bad) motivates the agent both to do or desire *x*, and to refrain from or be averse to *x*. It is unconflicted when *x* being bad motivates the agent only to do *x*. Perverse motivation is impure when e.g. *x* being bad is not the agent's only motivating reason for doing *x*, but it is accompanied by other, non-perverse reasons (say, the prospect of pleasure). It is pure when the perverse reason is the only motivating reason: the prospect of bad or wrong is the only attraction.

A perspicuous description of (as far as one can see, unconflicted and pure) normative perversion is provided by Augustine's autobiographical story:

'There was a pear tree near our vineyard, heavy with fruit, but fruit that was not particularly tempting either to look at or to taste. A group of young blackguards, and I among them, went out to knock down the pears and carry them off late one night ... We carried off an immense load of pears, not to eat – for we barely tasted them before throwing them to the hogs. Our only pleasure in doing it was that it was forbidden ... I was thus evil for no object, having no cause for wrongdoing save my wrongness. The malice of the act was base and I loved it – that is to say I loved my own undoing, I loved the evil in me – not the thing for which I did the evil, simply the evil ... seeking no profit from wickedness but only to be wicked'. (Augustine, 2006, p. 29)<sup>3</sup>

In this paper I will assume that normatively perverse agency is conceptually and psychologically possible. This is in part because arguing for this claim would require a separate article. But also, while I recognize it is a controversial assumption, I believe that the

---

or attacking it (Stocker, 1979; Velleman, 1992; Setiya, 2010). See also Hanisch (2018). In the text I will refer to such an agent as a 'normative pervert' or simply 'pervert', and to the phenomenon as 'normative perversion/perversity'. Finally, I will not discuss other kinds of normative perversity, for example theoretical perversity (believing that *p* on the ground that *p* is false/unsupported by evidence/contradictory etc.) or aesthetic perversity (appreciating a work of art purely on the ground of its negative aesthetic qualities, e.g. ugliness).

<sup>3</sup> Augustine, like Hume (2000, p. 270) and Kant (2009, p. 39), only discusses a moral type of normative perversion.

ensuing discussion actually strengthens the case in its favour: if perverse agency were not possible, a discussion over its rights and wrongs would sound vacuous, and we can be rather confident that the ensuing discussion at least does not sound vacuous. If normative perversion is possible, then what, if anything, is wrong or defective about it?<sup>4</sup> In section two I present and reject some accounts of what may be peculiarly wrong or defective about normative perversion. In section three I defend the positive suggestion that normative perversion is socially undesirable, in that it undermines certain valuable interpersonal and intrapersonal relations.

## **2. What is not wrong with perverse reasons**

In this section I will present, and reject, four initially plausible answers to the question of what is wrong with normative perversion. The point of this discussion is not merely to clear the ground for my own answer, but also to provide a number of desiderata for an acceptable answer to the question – desiderata that my account will be shown to satisfy better than other accounts.

### *2.1 The normative pervert acts for wrong reasons, or has malfunctioning attitudes*

It is rather obvious that a normative pervert acts for wrong reasons, even in those cases where she happens to do what she actually ought to do (morally or otherwise).<sup>5</sup> For example, she might believe that doing charity work for the poor is a bad thing to do – she might have even plausible-looking reasons for that belief, such as that charity worsens poor people's condition in the long run, and so on. However, given her perversion, the fact that doing charity

<sup>4</sup> The description given above might seem to already contain a sufficient answer to this question: it is wrong because it is perverse. If so, we can rephrase the question as: what sort of defect (rational, epistemic, moral, etc.) is normative perversion? The answer to *this* question is not self-evident. Any answer to this question will have to say something more in addition to the description given.

<sup>5</sup> For simplicity I will often only refer to acting for a normatively perverse reason, but similar considerations are meant to apply to desiring for a normatively perverse reason.

work is a bad thing to do motivates her *to do* charity work. She ends up doing what is – for the sake of argument – the right thing, but for an obviously wrong reason: you are not supposed to do anything for the reason that it is a bad thing to do.

However, acting for the wrong reasons is not what distinguishes the normative pervert. Suppose I want to do charity work only so that I will win the Good Guy award – I have no concern for the people I help. That I will win the Good Guy award is arguably a morally wrong reason for doing charity work, but it lies, intuitively, on a very different scale of wrongness to the pervert's consideration that doing charity work is a bad thing to do. What we want to say about perverse reasons is that they are not only wrong or bad from this or that point of view, but also *structurally inappropriate* – hence the air of paradox about someone who wants something on the grounds that it is bad, on top of a sense of their immorality (if they are motivated by thoughts of moral wrongness, say) or their imprudence (if they are motivated by the thought of something being bad for them). Therefore, a proper diagnosis must go beyond merely classifying perverse reasons as wrong reasons of this or that type.

In section 1 I said that for an agent to be motivated by a normatively perverse reason is for them to be motivated by a normative or evaluative thought as such which, if true, would count as such against the action that it motivates the agent to perform, or against the attitude that it motivates the agent to take. On the basis of this characterisation, there is room for different theorists to acknowledge different candidate normatively perverse reasons, depending on one's view of what counts as a normative or evaluative thought, and on one's view of whether that thought, if true, would count as such against the action that it motivates the agent to do. For example, if thoughts about thick evaluative properties count as evaluative or normative, and if thick properties count as such against or in favour of actions and attitudes, then these are also potential perverse reasons: doing something because it is cruel, dishonest, distasteful, stupid – or refraining from doing something because it is kind, honest, etc. Again, if thoughts about what is beneficial, harmful, healthy, unhealthy, are evaluative and not purely descriptive, and if the relative properties count at least to some extent as such in favour of or against actions and attitudes, then they will also be available as perverse reasons. Once there is sufficient agreement on the paradigmatic cases illustrated so far, nothing that I say in what follows will depend on taking a stand on this. What is important is that in all these cases the sense of structural inappropriateness is not lost: normative

perversity is desiring and acting on considerations that, though intelligible, are quintessentially defective *as reasons for acting or desiring* in the ways the normative pervert is motivated by them to act or desire, regardless of whether they are also defective as moral, prudential, or generically value-based reasons for acting or desiring in those ways.<sup>6</sup> In fact, the limiting case of normative perversity may be the following: being motivated to do *x* directly by the very thought that there is normative reason not to do *x*. In 2.4 I will argue that even this case – where the structural inappropriateness of the motivating reason couldn't be any more apparent – is a coherent possibility for the normative pervert.

The *normative* pervert, as understood here, thus contrasts with other agents who might also be thought of as being motivated by wrong or even perverse reasons, but for whom it is not normative or evaluative thoughts that do the motivating. For example, Mr. Sadik is motivated to tell unfounded scary news to people for the reason that doing so will cause them pain, but he gives no thought at all to any normative or evaluative fact, to the effect that causing pain to others is wrong or bad. Mr. Sadik's motivating reason is morally repugnant, but there isn't an air of paradox about it – it does not strike us as structurally inappropriate. Here is another example. Ms. Masok often does things in the light of the fact that they will minimize her own pleasure. On certain accounts of reasons, Ms. Masok acts on a wrong, even perverse, reason. But Ms. Masok never considers whether minimizing her own pleasure is bad, or even just bad for her. What attracts her is only the fact that her pleasure is minimized. No matter how prudentially misguided Ms. Masok may appear, there is not something structurally inappropriate about her motivation. Mr. Sadik and Ms. Masok are not *normative* perverts. (If, once they consider that causing pain to others or minimizing one's own pleasure are bad things to do, they get extra motivation to act in those ways, then they become what I have called normative perverts of the *impure* kind.)

So far, then, the first desideratum is that a satisfactory account of what is wrong with normative perversity should capture the way in

<sup>6</sup> Generically value-based perverse reasons are perverse reasons which are not obviously classified as moral or prudential. I gave one example above: that outcome A is worse than outcome B is a perverse reason for preferring A over B. For someone with consequentialist inclinations, this may be a perverse reason of the moral type. But one need not take any stand on this to recognize it as a perverse reason.

which perverse motivation is structurally, and not just morally or prudentially, inappropriate.<sup>7</sup>

Before leaving this section, I need to mention and set aside an answer to our question, which might seem attractive in the light of the first desideratum as just stated: People who act intentionally or desire on normatively perverse grounds possess *malfunctioning attitudes*. The idea would be that intentional action and desire have a compass-like function – namely, to orient us towards, at least, the perceived good and away from the perceived bad. Normatively perverse reasons (no matter whether morally or otherwise bad reasons) would be structurally inappropriate, insofar as they frustrate the function of the responses (actions or desires) they motivate.<sup>8</sup>

Whatever the theoretical merits of the idea of a function of intentional action or desire, the problem with this answer is that, as shown in later sections, a pervert's actions and attitudes need not be malfunctioning. The full explanation of this will have to wait until 2.4, but I can anticipate it by saying that a normative pervert's responses are correctly oriented, once this is (plausibly) understood as being oriented in accordance with perceived reasons *for one* to do or desire something. To this extent, her actions and desires are well functioning.

The first desideratum can now be restated: a satisfactory account of what is wrong with normative perversity should capture the way in which perverse motivation, though not malfunctioning, is structurally inappropriate, and not simply morally or prudentially wrong.

<sup>7</sup> I prefer not to describe normative perversion as a case of acting or desiring for *the wrong kind of reasons*. While it is true that 'x is bad' just *cannot* be a good reason to do x, this has nothing to do with the way that, for example, 'believing that p makes me happy' is (according to many) a wrong kind of reason for believing that p. In normative perversity there is no 'trespassing of normative domains', as it were. Nor is 'x is bad' unintelligible as someone's reason to do x in the way that, for example, the love of Sophocles (without further elaboration) is an unintelligible reason for drinking tea (to use an example of Joseph Raz's).

<sup>8</sup> This is different from the idea that normative perverts fail a *constitutive* norm or aim of intentional action or desire (see McHugh and Way, 2018). If a constitutive norm for an activity A is such that, if you disregard it (often enough), you don't count as A-ing, then the pervert (at least the unconflicted and pure type) would count as not really acting intentionally or desiring, rather than acting or desiring inappropriately. But this violates the assumption I made: normative perversity is conceptually and psychologically possible.

2.2 *The normative pervert is practically irrational*

Consider a version of internalism about wrongness judgments:

*Rational Internalism.* It is necessary and a priori that, if A judges that x is wrong, then either A has a defeasible motivation to avoid doing x, or A is practically irrational.<sup>9</sup>

Practical irrationality is in this sense a form of incoherence between one's normative judgments and one's motivation. Perverse agents in the 'conflicted' version *are practically rational* as far as Rational Internalism goes: they do have some motivation to avoid doing what they judge to be wrong. So obviously this view cannot capture what is wrong with conflicted normative perversion. What about unconflicted perverts? By Rational Internalism, they would be practically irrational, as they lack even any defeasible motivation to avoid doing what they judge to be wrong. However, cases of practical irrationality are normally thought to involve an extreme form of apathy, or depression, or in general weakness of will: one judges that x is wrong, but just cannot be brought to be motivated not to do x, and so ends up being motivated (or remains unmotivated, as the case may be) *against one's better judgment*. But the case of unconflicted normative perverts is markedly different: there is no sense of inner struggle or going against one's better judgment, since one's normative judgment *is* followed through, albeit in a direction opposite to what it should be. And while it usually makes sense for a weak-willed agent to wish that she were motivated in accordance with her better judgment (and to be unhappy with herself at the thought that she is not motivated accordingly), it would not make sense for a pervert *qua pervert* to wish that she were motivated by her judgment 'as she should' (i.e. in a non-perverse way). Such a higher-order attitude would mean a serious crack in, if not the end of, her unconflicted normative perversion.

Rational internalists could dig in their heels and insist that, even though different both phenomenologically and in other ways from weakness of will, normative perversion still is a form of practical irrationality. However, this diagnosis becomes less plausible once we place Rational Internalism next to a similar, but different view:

<sup>9</sup> This is inspired by Michael Smith's view about judgments of rightness (1994). Of course Rational Internalism only addresses perverse agents motivated by thoughts (indeed, judgments) of moral wrongness. It is not obvious that there are plausible internalist views about the whole range of normative and evaluative thoughts that may perversely motivate. But I won't press this issue.



*Overall Reason Rational Internalism.* It is necessary and a priori that, if A judges that A has overall *reason* to do x, then either A has a defeasible motivation to do x, or A is practically irrational.<sup>10</sup>

When judged by this theory, at least some perverse agents are practically *rational*. True, such agents make the sort of judgments ('x is wrong') which imply that there is (strong, if not overall) reason *against* doing certain actions. But this doesn't mean that they judge that *they have* overall reason not to do those actions. In ways to be explained in section 2.4, such agents can resist the inference from 'x is wrong' to '*I have* overall reason not to do x'. On the contrary, they may well judge that *they have* overall reason to do x, a reason provided by the fact that x is wrong (or by another normative or evaluative fact). And if they do so judge, then they are practically rational, since they are indeed motivated to do x.<sup>11</sup>

Notice that I am not assuming that perverse agents *must* be guided by a judgment of overall reason. What is essential to a perverse agent is acting, perversely, in the light of certain considerations. This might or might not also involve making the normative judgment that these considerations amount to pro tanto or overall reasons for action. If it does involve such normative judgment, then perverse agents are practically rational. If it does not involve such a judgment, then perverse agents are neither rational nor irrational, by the lights of Overall Reason Rational Internalism. Should they then be regarded as practically irrational by the lights of Rational Internalism? The point made above about the lack of inner struggle suggested that, if perverse agents turn out to be practically irrational, then we have reason to doubt Rational Internalism as stated there. Perverse agents may or may not be in some way incoherent, but they are not incoherent *by their own lights* in the way akratics typically are. It seems then that an improved version of Rational Internalism would have to include a clause excluding perverse agents from its scope. (A different suggestion, to be discussed in 2.4, is that perverse agents are *already* excluded from the scope of Rational Internalism, if their

<sup>10</sup> This is inspired by Michael Ridge's internalism about normative judgments (2014).

<sup>11</sup> Has such an agent thereby become a 'sappy' perverse agent (see Velleman, 1992) by producing an overall normative judgment and then acting *in conformity* with it rather than in opposition to it? I don't think so. Like I said above, there is a clear sense in which unconflicted perverse agents face no inner struggle or weakness of will. The point made here simply builds on this insight. Perverse agents can afford to think of their motivating reasons as being normative *for them*.

understanding of wrongness and similar normative concepts is too deviant to attribute them genuine judgments of wrongness.)

In any case, we can at least conclude on a second (negative) desideratum: what is wrong with perverse agents should not be a matter of lack of internal coherence.<sup>12</sup>

### *2.3 The normative pervert undermines her own perversity*

David Sussman (2009) argues that perverse action has ‘a point, but that point cannot be addressed to anyone occupying a different point of view ... if I do something just because it is bad, how am I supposing that others will respond when I tell them this?’ (2009, p. 627). The only way, for example, for someone like the young Augustine to address non-perverse others (defend himself from their challenges, advise them, criticize them, etc.) would be to present his point of view as ‘an aberrant conception of what is good, and so undo its own character as perverse’ (ibid.). In other words, as soon as a normative pervert is faced with an interpersonal request for justification, he can only present his actions as something in favour of which there is at least some reason. But, according to Sussman, this is tantamount to undermining one’s own perversity, since Augustine’s perversity crucially depended on his understanding ‘that the moral wrongness of an act constitutes or entails a compelling reason against so acting ... he stole from an appreciation of the wrongness or wickedness of theft *as* a compelling reason against stealing’ (ibid., p. 614, p. 616). *If* perversity requires thinking of one’s acts as something there are compelling reasons against, and none in favour of, and *if* in interpersonal exchange one cannot but appeal to reasons in favour of one’s actions, then in interpersonal exchange the pervert will have to give up his perversity. This is not so much a charge of practical irrationality as one of inherent instability.

However, the first antecedent above can be resisted. As seen in the previous section, a normative pervert *can* think of her actions as what she has overall reason to do – a reason provided by badness, wrongness, and the like. That is how she can turn out to be practically rational. While there is a sense in which she does think of, for example, wrongness as counting against doing what she does (or else her understanding of ‘wrongness’ would be too deviant), that

<sup>12</sup> In this connection, it is noteworthy that none of the cases labelled by Pettit and Smith (1993) as cases of ‘practical unreason’ are anywhere close to normative perversity as understood here.

thought will have to be compatible with the fact that perceived wrongness *guides* rather than hinders her actions. Thus it is false that the pervert must think of her actions as something *she has* overall compelling reasons against doing, and none in favour of doing. If this point is sound, then in interpersonal exchange the pervert is free to present badness and the like as reasons in favour of what she does, without thereby undermining her own perversity.<sup>13</sup>

Sussman might at this point reply that presenting badness as a reason in favour of doing bad things is not likely to win many adherents (among non-perverts, that is). This is probably true. However, this is not going to stir up much trouble for the pervert. First, winning many adherents may or may not be an instrumentally desirable goal for the normative pervert. Even if she were intent on maximizing evil for its own sake, the best strategy may not be to convert as many as possible to the cause of evil for evil's sake. (If you want to maximize virtue, the best strategy may not be to convert as many as possible to the cause of *virtue for virtue's sake*.) Second, if winning adherents is indeed an instrumentally desirable goal, she surely doesn't have to present badness as a reason in favour of doing bad things. She can simply *deceive* about her reasons for doing what she does, if that promises to be the best long-term strategy – and if she regards deceiving as wrong, she will be all the more attracted to this option. For example, in order to get people to steal, she can cite all sorts of non-perverse considerations that may weigh with her audience (fun, adventure, sense of liberation from conventions etc.), even though these are not *her* reasons to steal. Her hope would be to attract people to evil first, and only afterwards get them to appreciate evil 'as its own reward', so to speak.

Sussman might then insist that, if the alternatives at the pervert's disposal are either to fail to share her viewpoint, or to present it in deceitful ways (if temporarily), there is still something deeply flawed in perversity. This might be true, but it is not clear why the pervert *qua* pervert should be worried. Not sharing her viewpoint, or deceiving

<sup>13</sup> Perhaps Sussman believes that the very activity of engaging in interpersonal justification presupposes a non-perverse motivation: one has to think of one's perversity as *worth* justifying to others, and act in conformity with, rather than in opposition to, *this* evaluative belief. In this sense a 'socially engaged' normative pervert would undermine herself regardless of what she actually does or says in addressing others. However, as discussed in 2.2, it is not at odds with normative perversion to act in conformity with one's judgments about what one has reason to do, assuming that such judgments reflect a perverse motivational structure.

others, are not strategies that undermine her ‘integrity’ as a pervert. They might cause a kind of social isolation, but it is not clear how that would be bad for the pervert as such.

The upshot of this subsection is mixed. On the one hand, normative perverts *can* defend themselves in the face of non-perverts without undermining their own perversity. On the other hand, it is likely that in social address the pervert will often resort to deceiving others. Thus, even if, *pace* Sussman, the pervert herself is not the victim of normative perversity, it does seem promising to pick out *critical exchange with others (non-perverts)* as the context in which to look for what might be wrong with perversity. The third desideratum therefore is that what is wrong with normative perversity must have something to do with how the pervert must relate to others, non-perverts.

#### *2.4 The normative pervert is ignorant about the nature of evaluative properties*

Consider a buck-passing account (BPA) of the evaluative and normative features that motivate the normative pervert:

*BPA.* For  $x$  to be bad (wrong) is for  $x$  to have features that provide normative reason to oppose (not to do)  $x$ .<sup>14</sup>

BPA may seem to be *the* account that makes it particularly clear why perverse reasons are indeed perverse. If for  $x$  to be bad is for  $x$  to have features that provide normative reason to oppose  $x$ , then badness, as far as it goes, *rules out* reasons to favour  $x$ . And if badness as far as it goes rules out reasons to favour  $x$ , then badness just *cannot* be a reason to favour  $x$ . In other words, while there can be reasons to favour something bad, these reasons cannot just be provided by the fact that it is bad. And this gives us a possible diagnosis of what is wrong with perverse agents. As noted in 2.2, a normative pervert may be motivated to favour or to do  $x$  simply in the light of  $x$ 's

<sup>14</sup> See Scanlon (1998, p. 97). For an introduction see Suikkanen (2009). A book-length defence is Rowland (2019). Scanlon only ‘passes the buck’ in the case of value properties:  $x$  being good or  $x$  being bad are not *themselves* reasons to favour or oppose  $x$ . But he allows that wrongness can itself be a reason not to do  $x$ . To keep things simple, I will assume that both bucks are passed –  $x$  being bad and  $x$  being wrong do not themselves provide reasons to oppose  $x$ , though both rule out, as far as they go, reasons to favour  $x$ .

badness, or they may also be further motivated by the normative belief or judgment that badness provides a reason to favour or do *x*. In the first case, in favouring *x* in the light of *x*'s badness, that is, in treating badness as a reason to favour *x* (without necessarily having the belief that badness is a reason to favour *x*), the pervert manifests a kind of ignorance of what for BPA is a necessary property of badness (the property of ruling out, as far as it goes, reasons to favour what is bad). In the second case, since the pervert favours *x* upon the normative belief that *x*'s badness provides a normative reason to favour *x*, then she has a belief which is necessarily false by the lights of BPA (there is no possible world where *x*'s badness as such provides a normative reason to favour *x*).<sup>15</sup> Either way, normative perverts would be guilty of some epistemic fault with respect to the nature of badness: they do not get badness right (and likewise if they act perversely on other evaluative features for which a BPA-style account is true).<sup>16</sup>

However, this BPA-based diagnosis seems mistaken, and understanding why sheds further light on normative perversity. In treating badness and the like as reasons to favour *x*, or in being motivated by the belief that badness is a reason for them to favour *x*, the normative pervert need not manifest any epistemic fault, even by the lights of BPA. The pervert, in fact, needs to possess an adequate understanding of the normative or evaluative features she is motivated by, or she would not count as a pervert. Suppose the young Augustine understands 'wrong' as a purely descriptive or inverted commas concept, for example 'disapproved by the elderly'. Being motivated by 'wrongness' in this sense does not amount to normative perversion, but, say, only to rebellion against the elderly. Again, imagine someone who has a seriously deviant understanding of 'wrong', for example, 'reducing

<sup>15</sup> The distinction between *treating p as a reason* and *believing p to be a reason* is important on certain accounts of motivating reasons (Schlosser, 2012; Alvarez, 2017), but also in the context of understanding theoretical reasoning (McHugh and Way, 2016). But here both possibilities deserve to be mentioned, because motivation by perverse reasons can be exhibited in both ways. For example, a charitable interpretation of the Satanic motto 'evil, be thou my good' would ascribe to Satan the *belief* that evil is a reason for him to do certain things (Gregory, 2013). Satan's perversity goes beyond simply his being disposed to act in the light of perceived evil – it expresses a considered normative stance towards evil.

<sup>16</sup> If they are conflicted perverts, then they are also motivated to oppose *x* in the light of *x*'s badness. To this extent, they 'get badness right'. But since they are also motivated to favour *x* in the light of *x*'s badness, they still do not *fully* get badness right.

the quantity of nitrogen in the universe'. Someone who (barely intelligibly) is motivated to do something because it is 'wrong' in this sense is no normative pervert. Finally, the normative pervert cannot quite *mean* by 'wrong' something like the opposite of BPA, for example 'such that there is reason to favour it'. This would amount to a *perverse understanding* of 'wrong', but the air of paradox about her *motivation* would thereby disappear. In acting motivated by wrongness in her sense, she would act on a merely idiosyncratic notion. We lose the sense of something structurally inappropriate about her motivation – after all, she would do what she thinks there is only reason to favour, and no reason not to favour, as far as its 'wrongness' goes. What distinguishes the normative pervert is not a semantic inversion of normative terms, but the way her otherwise adequate understanding of normative terms plays out perversely in desire and action.

Perhaps surprisingly, normative perverts can possess a sufficiently adequate understanding of normative terms by actually embracing BPA (as formulated above) without any inconsistency. Augustine may well understand 'x is wrong' as 'x is such that there is (sufficient) reason against doing x'. Given this understanding, can Augustine consistently (a) treat the wrongness of stealing as a reason *for him* to steal? And can he consistently even (b) have the belief that wrongness provides *him* with a reason to steal? If he cannot, then this would show that normative perverts with an adequate understanding of the normative property that motivates them are caught in a contradiction – yet another epistemic fault.

Now it is entirely possible that many real-life normative perverts are inconsistent in just this way. However, there is logical space for normative perverts to avoid such contradictions, and indeed avoid the alleged epistemic faults altogether. Augustine can draw a distinction between

- (1) x is such that *there is* (sufficient) reason against doing x, and
- (2) x is such that *I have* (sufficient) reason against doing x.

In other words, Augustine may deploy an understanding of normative reasons for action which does not immediately assign reasons to any particular agent – and in particular, not to *him*. For example, Augustine may plausibly believe that the reasons against doing x which have to do with x's wrongness are neither part of his actual motivational set, nor part of the motivational set he would have after sound deliberation, and conclude that these are not reasons *for him*, albeit being genuine reasons against doing x. At any rate,

Augustine may take this as an open question: There are reasons against doing *x*, but do I have those reasons?

Armed with this distinction between what may be called *unowned* and *owned* normative reasons, Augustine can (a) consistently believe that the wrongness of theft rules out unowned reasons to steal, while treating the wrongness of theft as a reason for him to steal (a reason owned by him).<sup>17</sup> Likewise, he can (b) consistently believe that the wrongness of theft rules out unowned reasons to steal, while believing that the wrongness of theft is a reason for him to steal (a reason owned by him). It may be replied that, if you believe that wrongness rules out unowned reasons to steal, then you must believe that wrongness also rules out owned reasons to steal, on pain of contradiction. But Augustine (or any normative pervert) can resist this reply. Owned and unowned reasons may be held to be altogether different classes of normative reasons, for example corresponding to the distinction between internal and external reasons.<sup>18</sup>

<sup>17</sup> It might seem that case (a) does not even have the appearance of logical inconsistency, because Augustine does not have two beliefs to start with. But treating *p* as a reason for *x* while believing *p* to be a reason against *x* could still be seen as a kind of incoherence, if *p* is treated as a reason for *x* of the same kind as the reason against *x* that *p* is believed to be. Thanks to a reviewer for pressing this point.

<sup>18</sup> To be clear: unlike Bernard Williams (1981), the normative pervert in question would believe that *there are* genuine external normative reasons alongside internal ones. In addition to the external/internal reasons distinction there are other theoretical options to substantiate the unowned/owned reasons distinction. For example, unowned reasons might be ‘reasons of fittingness’, which do not address any agent – they simply express a normative or evaluative relation of fittingness between a type of object or situation and a type of response. A wrong action is such that it has properties which make it fitting not to do it (or unfitting to do it). Owned reasons on the other hand require a different structure, such that the agent is a term of the relation. The idea that properties like badness or wrongness should be understood in terms of fittingness goes back to the way Franz Brentano and A.C. Ewing understood value, though it is not clear whether they saw any contrast between fittingness and owned reasons. See Rabinowicz and Rønnow-Rasmussen (2004) and Jacobson (2011) for possible differences between a fitting attitude account of value and BPA, and Howard (2018) for an overview of recent literature on fittingness. Finally, I should note that my use of ‘unowned’ and ‘owned’ reasons does not correspond to Errol Lord’s distinction between objective normative reasons and ‘possessed’ normative reasons (2018). (See also Hawthorne and Magidor, 2018.) For Lord, a possessed normative reason is an objective normative reason for an agent to do something, which meets certain subjective conditions. But the owned

I am not saying that Augustine would be right in drawing such distinctions. I am only arguing that nothing in BPA itself dictates that reasons be automatically owned by someone – any such debate belongs to the theory of normative reasons, not to the theory of the nature of value or other normative properties. Moreover, the very existence of the internalism vs. externalism debate about normative reasons is enough evidence that an understanding of certain normative reasons as unowned is not itself a deviant understanding. Augustine can have an adequate understanding of both value properties *and* normative reasons.

Summing up, even when judged by the lights of BPA, the normative pervert need not manifest any ignorance, false belief, or contradictory belief about normative facts. What is defective is the structure of the pervert's motivation – how she is attracted or guided by normative and evaluative properties – not her understanding of those properties, which must (and can) be adequate enough for her to count as normatively perverse. The fourth desideratum is thus that a satisfactory account of what is wrong with normative perversion should not seek the pervert's defect in any epistemic fault with respect to the nature of properties like badness or wrongness.

### **3. What is wrong with normative perversion**

Here is a summary of the four desiderata emerged so far. A satisfactory answer to 'what is wrong with normative perversion'

1. should capture the way in which perverse motivation, though not malfunctioning, is structurally inappropriate, and not simply morally or prudentially wrong;
2. should not locate the fault in practical irrationality or lack of internal coherence;
3. should have something to do with how the normative pervert must relate to others, non-perverts;
4. should not seek the pervert's defect in some epistemic fault with respect to the nature of properties like badness or wrongness.

---

reasons that the pervert needs are not already objective reasons – that is why the pervert can coherently understand x's wrongness as ruling out unowned reasons to favour x while treating x's wrongness as an owned reason (owned by him) to favour x. If he had to believe that x's wrongness is also an unowned reason to favour x, then this belief would contradict his (otherwise adequate) understanding of wrongness.



In this section I defend an alternative account of what is wrong with normative perversion in terms of its social undesirability (in a sense to be explained). First I spell out the account, then I show how this account satisfies the four desiderata above.

We can start by remarking that there is a slightly different view to BPA, the *acceptance* of which has great social importance:

*BPA+*. For x to be bad (wrong) is for x to have features that provide *anyone* with normative reason to oppose (not to do) x.

Unlike BPA, BPA+ establishes an immediate connection between ‘x is wrong’ and ‘there is reason *for me* not to do x’ (and likewise for other normative, deontic, or evaluative properties for which an account like BPA+ is true). In the terms introduced above, on BPA+ the normative reasons having to do with wrongness and the like are by definition *owned* by agents (or at least, by agents who *can* respond or act in the relevant ways). Now BPA+ may or may not be true. But I am going to argue that accepting BPA+ as one’s (implicit or explicit) understanding of wrongness and the like is socially desirable. When you point out to A that her proposed course of action is wrong, and she agrees, it is desirable that she *ipso facto* gets to believe that there is at least some reason *for her* to refrain from it, or that she *ipso facto* gets to treat certain features (the wrong-making ones) as reasons *for her* to refrain from it. In other words, it is desirable that she accepts that no space is left to wonder ‘so what?’.<sup>19</sup> But the consistent normative pervert sketched so far precisely rejects the idea that, if what she proposes to do is wrong, there is *ipso facto* reason *for her* not to do it. She creates (or at least can create) for herself, and in fact for any other agent, the space to wonder ‘so what?’ – because only by creating such a space can she then consistently treat wrongness as a reason *for her to do* the action. If creating such a space is socially undesirable, the normative pervert’s fault lies, in the first instance, in creating such a space.

Why and in what sense is it socially desirable that, upon believing that x is bad or wrong, we *ipso facto* get to believe that there is at least some reason *for us* to oppose or refrain from x? And why is it socially undesirable not to automatically have such a belief?

When normative and evaluative concepts are understood along the lines of BPA+, people are placed into certain relations with each other, or intrapersonally with themselves as well. If you and I believe that I did something wrong, and you and I *ipso facto* believe that there was (sufficient) reason *for me* not to do that action, you can hold me accountable for what I did (in particular, but not only,

<sup>19</sup> See Stratton-Lake and Hooker (2006).

if you are the victim or you care for the victim of my action), and I am in a position to accept your holding me accountable. We recognize each other in our respective roles. The same point applies even if other people are not involved. If I believe I did something that was, prudentially, overall bad (overall bad for me), and I *ipso facto* believe there was overall prudential reason *for me* not to do that action, I hold myself accountable – in this case, towards myself – for what I did. I establish a relation of accountability with myself.

Now, when the normative pervert creates the space for a ‘so what?’ reaction to her normative beliefs, she undermines such relations, both in the interpersonal and the intrapersonal case. Remember that the pervert must take wrongness, in its ordinary role as a consideration that counts against doing *x*, as a real but unowned reason against doing *x*, at least not owned by herself. A non-pervert who accepts BPA+, by contrast, takes wrongness as a reason *for whatever agent* – i.e. owned by anyone – against doing *x*.<sup>20</sup> Given this difference, a pervert and a non-pervert cannot enter the relation of mutual recognition sketched above (at least as far as wrong actions are concerned). The non-pervert may hold the pervert accountable, but the pervert’s ‘so what?’ reaction entails that she is refusing accountability for her own wrong actions, because she does not treat wrongness-related reasons against doing *x* as owned by herself. She only treats perverse wrongness-related reasons *in favour of* doing *x* as owned by herself. The point of course is not that the pervert is actually off the hook – the point is rather that a certain kind of relation with non-perverts fails to get established.<sup>21</sup>

Again, the same applies even if other people are not involved. The prudential normative pervert treats the fact that an action or outcome

<sup>20</sup> For simplicity of exposition I write here as if wrongness ‘keeps’ the buck.

<sup>21</sup> I use the cases of morally and prudentially wrong actions for illustration, but similar considerations apply to bad states of affairs and other evaluative facts which are amenable to a BPA+ account. For example: Mr. Ignoramus prefers ignorance over knowledge for the perverse reason that ignorance is worse than knowledge. A non-pervert may hold Ignoramus accountable for this preference – or hold a similar attitude towards him. But Ignoramus, though believing that if ignorance is worse than knowledge, there is reason to prefer knowledge, takes this as an unowned reason – at least not owned by himself. Ignoramus, *qua* pervert, is not in a position to accept certain non-perverts’ attitudes towards him and his preferences, and thus fails to maintain a relation of mutual recognition with them (at least as far as the value of ignorance and knowledge is concerned).

is prudentially bad (bad for her) as a reason for her *to do it* or *to favour it*, and to do so coherently, she must think that something being prudentially bad only provides *unowned* reason *not to do or favour it* – that is, she must create the logical space for a ‘so what?’ reaction in this case as well. But in doing so, she ensures that whatever prudential advice or reproach her non-perverse self addresses to her perverse self does not get a grip. Again, the point is not that she, as a pervert, actually manages to ‘get off the prudential hook’. However, she, as a pervert, fails to enter a relation of accountability with her non-perverse self.<sup>22</sup>

Entering and maintaining such inter- and intrapersonal relations of recognition is a valuable thing. This thought can be articulated in a number of ways, but the basic insight is that entering and maintaining such relations is constitutive of valuing each other (and ourselves) as beings to whom a reasonable justification is owed.<sup>23</sup> Let’s return to the young Augustine stealing pears for the reason that it is a wrong thing to do. In acting on such a reason, Augustine must, on pain of inconsistency, reject the shopkeeper’s claim that there is a very strong reason *for him* not to steal the pears. In doing so, Augustine refuses to acknowledge accountability towards the shopkeeper. He breaks the relation of mutual recognition (if ever there was one). And breaking this relation means that he treats the shopkeeper as unworthy of a reasonable justification. Even though Augustine, as argued above, must recognize that there was something to be said against stealing the pears – or else his perversity wouldn’t be genuine – the only consideration that Augustine could honestly offer as a ‘justification’ for his action is that stealing the pears was a wrong thing to do. Assuming that the shopkeeper is not himself a normative pervert, such a ‘justification’ is as unreasonable as can be

<sup>22</sup> Can the pervert establish relations of mutual recognition with similarly minded normative perverts? If Augustine and his similarly minded friend take the wrongness of theft as a reason owned by both of them in favour of stealing, then are they going to hold each other accountable *for failing to steal* – i.e. for failing to do their perverse duty? The answer to this question requires full theorizing on the conditions for the relation of mutual recognition. In the text I have only assumed that, if I don’t see wrongness as a reason *for me* not to do x, and you do, a relation of mutual recognition between us is undermined. Thus I have only spoken of a necessary condition for mutual recognition, not a sufficient one. I do not have the space to discuss potential further conditions.

<sup>23</sup> The recent classic for this is Scanlon (1998). However, as far as I am aware, he does not draw the connection between relations of mutual recognition, the acceptance of a BPA+ account, and normative perversity.

from the shopkeeper's (or anyone's) point of view. In general, the very demand for a reasonable justification to non-perverse others, and to one's own non-perverse self, cannot even *begin* to be met by perverse agents.

In fact, now we can see that normative perverts undermine relations of mutual recognition not only by creating the space for a 'so what?' reaction. In principle one could create such a space without being a normative pervert: one could simply remain indifferent to thoughts about good and bad or right and wrong. But the normative pervert goes further: she is repelled by the good or the right, and attracted by the bad or the wrong. She exploits the 'normative void' she creates in order to let her perverse reasons operate undisturbed. And mustering perverse reasons (or at least being disposed to do so) is a direct affront against the very ideal of reasonable justification, again assuming non-perverts to be on the receiving end of such a justification. The normative pervert thus adds insult to injury against that ideal.<sup>24</sup> By contrast, the normatively indifferent is by definition neither against, nor in favour of what they regard as wrong. And if they do go on and act in ways they regard as wrong, at least they will act on non-perverse reasons – say, they might steal because stealing is fun or exciting. While such considerations would still be unreasonable as a justification – the shopkeeper would rightly reject them as in any way justifying stealing – at least they do not dishonour the very ideal of reasonable justification; they just fail to be reasonable considerations under the circumstances.

With this, I have sketched an account of why normative perversity is socially undesirable. To be sure, I haven't defended the idea that people should be valued as beings to whom reasonable justification is owed. This would require a much longer work. However, what I can do here is to argue that this account satisfies the four desiderata for what is wrong with normative perversity.

First, we need to capture the way in which perverse motivation, though not a case of malfunctioning action or desire, is structurally inappropriate, and not simply morally or prudentially wrong. At first sight it might seem that my account in terms of undermining the ideal of reasonable justification condemns perverse reasons from a moral point of view, and is thus unable to meet this desideratum. But this would be a mistake. The desideratum of structural

<sup>24</sup> As hopefully clear throughout the essay, the pervert's 'insult and injury' must be understood at the level of her motivating reasons – what she ends up doing may not actually be wrong, at least on some views of intentional action.

inappropriateness requires finding a diagnosis of what's wrong with normative perversion that encompasses normatively perverse reasons of moral, prudential and other types, showing them to be inappropriate *as reasons*. But what makes them inappropriate as reasons may well be that they fail to match some substantive ideal that reasons should match regardless of what type of reasons they are. On the present account, this is the ideal of reasonable justification. As illustrated, whether one acts as a moral pervert (i.e. treats *x* being wrong as a reason in favour of *x*) or as a prudential pervert (i.e. treats *x* being bad for her as a reason in favour of *x*), in both instances the ideal of reasonable justification is equally in force, and is equally violated by both perverts, even if prudential perversity may well be nobody's business but the pervert's. This should not surprise: the damage done to the ideal of reasonable justification starts from creating the space for what I have called the 'so what?' reaction to certain normative and evaluative thoughts (while perversity aggravates this damage as compared with mere indifference), and this space can always be created, no matter which domain one is considering – no matter what type of normative reasons one is being confronted by.<sup>25</sup> Thus, the present account satisfies the first desideratum, regardless of whether we should classify the ideal of reasonable justification as 'moral' or otherwise.

Second, we should not locate the fault of perversity in practical irrationality or lack of internal coherence. My account satisfies this desideratum, because it identifies the source of the problem in what perversity does to certain relations with people. And this is a substantive failure, in the sense just described. Even when what is undermined is a certain intrapersonal relation between the perverse and the non-perverse self, the fault lies in the pervert dishonouring an ideal rather than in the pervert finding herself, *qua* pervert, in a state of weakness of will or incoherence. But what about the conflicted pervert, who is both to some extent attracted and to some extent repelled by thoughts about the bad? Even if, as argued above, she turns out to be rational by the lights of Rational Internalism, because she is at least to some extent motivated to avoid doing *x* by the thought that *x* is bad (2.1), she is overall (unsurprisingly) in a state of motivational incoherence, because she treats '*x*

<sup>25</sup> I thus suspect a similar diagnosis can be given of epistemic perversity – we are owed, by others and by ourselves, reasonable justification for beliefs. The epistemic pervert (who believes that *p* on epistemically perverse grounds, e.g. that there is no evidence for *p*) dishonours this demand. But I will not go into this issue.

is bad' as both (implying) a reason for her to avoid doing x, and as a reason for her to do x. To the extent that she treats 'x is bad' as (implying) a reason for her to avoid doing x, she *could* enter relations of mutual recognition with non-perverts. But her perverse side must keep pulling her away from such relations – she cannot expect her perverse reasons to be acceptable as reasonable justification any more than the unconflicted pervert can. Thus our diagnosis applies also to the conflicted pervert, as it should.

Third, a proper account should have something to do with how the normative pervert must relate to others, non-perverts. This is the positive lesson learned from Sussman's argument. The current proposal has pretty much everything to do with how the pervert must relate to others, non-perverts: being a consistent normative pervert commits her to a 'so what?' reaction to normative beliefs addressed to her, and this reaction – aggravated by the outrageous 'justification' she, as a pervert, is disposed to offer to non-perverts – undermines both relations of mutual recognition with others, and relations of accountability with oneself. Unlike in Sussman's diagnosis, however, the fault does not lie in a tendency to undermine one's own perversity when confronted by others – the normative pervert is a threat for non-perverts (to the extent that they are owed reasonable justification) and not, or not necessarily, a threat to herself *qua* pervert.

Fourth, we should not seek the pervert's defect in the epistemic fault of ignoring the nature of properties like badness or wrongness, or merely having false or contradictory normative and evaluative beliefs. My account satisfies this desideratum. The fault of undermining certain valuable relations with others and with oneself is squarely practical, and perhaps even 'moral', but not epistemic. It is true that, at least for the consistent pervert, this fault depends on taking a certain 'theoretical' stance towards claims about normative reasons – i.e. on making, at least implicitly, the suggested distinction between unowned and owned normative reasons. But what I have pointed out here is the social undesirability of making this distinction. I have not at all suggested that the pervert's mistake is that such a distinction is false.

The pervert, as such, may naturally remain unmoved by this diagnosis – it may even suggest her further perverse reasons ('I will do x because my reasons violate the ideal of reasonable justification'). But the problems found in accounts which lay the blame on something the pervert may – in her residual rationality – care about (malfunctioning attitudes, practical irrationality, self-undermining dispositions, false or contradictory beliefs) indicate that a satisfactory account will offer no particular hopes for reform.

## 4. Conclusion

In this paper I have described normative perversity and rejected several answers to the question of what is wrong with it. In the course of the discussion a few desiderata for a satisfactory answer have emerged. This is the kind of spadework that, to my knowledge, has not been systematically carried out before. Finally, I have suggested that normative perversity undermines certain inter- and intrapersonal relations which require honouring ideals of reasonable justification, and that this account satisfies the desiderata. To be sure, this is meant as a first pass at a satisfactory answer. I did not have the space to examine whether there may be alternative proposals that even better satisfy the desiderata.<sup>26</sup>

## References

- Alvarez, Maria, 'Reasons for Action: Justification, Motivation, Explanation', *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/>>.
- Augustine, *Confessions* (II edition), edited by M. P. Foley (Indianapolis/Cambridge: Hackett, 2006).
- Alex Gregory, 'The Guise of Reasons', *American Philosophical Quarterly*, 50 (2013), 63–72.
- Christoph Hanisch (ed.), *Ethical Theory and Moral Practice* 21. Special Issue on 'Acting Under the Guise of the Bad' (2018).
- Jennifer Hawkins, 'Desiring the bad under the guise of the good', *Philosophical Quarterly*, 58 (2008), 244–64.
- John Hawthorne and Ofra Magidor, 'Reflections on the Ideology of Reasons' in Daniel Star (ed.), *Oxford Handbook of Reasons and Normativity* (Oxford: Oxford University Press, 2018).
- Christopher Howard, 'Fittingness', *Philosophy Compass* 13 (2018), e12542.
- David Hume, *A Treatise of Human Nature* (Oxford: Oxford University Press, 2000).
- Daniel Jacobson, 'Fitting Attitude Theories of Value', *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2011/entries/fitting-attitude-theories/>>.

<sup>26</sup> I thank two anonymous reviewers for this journal for their careful and helpful comments. I thank audiences in Tartu, Helsinki, Bucharest, L'Aquila, and Munich for feedback on previous versions. This article was written with the support of grants PUT1630 (Estonian Research Council) and SHVHV16145 T (TK145) (SA Archimedes).

## Francesco Orsi

- Immanuel Kant, *Religion Within the Bounds of Bare Reason* (Indianapolis/Cambridge: Hackett, 2009).
- Errol Lord, *The Importance of Being Rational* (Oxford: Oxford University Press, 2018).
- Conor McHugh and Jonathan Way, 'Against the Taking Condition', *Philosophical Issues*, 26 (2016), 314–31.
- Conor McHugh and Jonathan Way, 'What is Reasoning?', *Mind*, 127 (2018), 167–96.
- Philip Pettit and Michael Smith, 'Practical Unreason', *Mind*, 102 (1993), 53–79.
- Wlodek Rabinowicz and Toni Rønnow-Rasmussen, 'The strike of the demon: On fitting pro-attitudes and value', *Ethics*, 114 (2004), 391–423.
- Joseph Raz, 'The Guise of the Bad', *Journal of Ethics and Social Philosophy*, 10 (2016), 1–15.
- Michael Ridge, *Impassioned Belief* (Oxford: Oxford University Press, 2014).
- Richard Rowland, *The Normative and the Evaluative: The Buck-Passing Account of Value* (Oxford: Oxford University Press, 2019).
- T. M. Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998).
- Markus E. Schlosser, 'Taking Something as a Reason for Action', *Philosophical Papers*, 41 (2012), 267–304.
- Kieran Setiya, 'Sympathy for the devil' in Sergio Tenenbaum (ed.), *Desire, Practical Reason, and the Good* (Oxford: Oxford University Press, 2010), 82–110.
- Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994).
- Michael Stocker, 'Desiring the bad: An essay in moral psychology', *Journal of Philosophy*, 76 (1979), 738–53.
- Philip Stratton-Lake and Brad Hooker, 'Scanlon versus Moore on goodness' in Terry Horgan and Mark Timmons (eds.), *Metaethics After Moore* (Oxford: Oxford University Press, 2006).
- Jussi Suikkanen, 'Buck-passing accounts of value', *Philosophy Compass*, 4 (2009), 768–79.
- David Sussman, 'For Badness' Sake', *Journal of Philosophy*, 106 (2009), 613–28.
- Sergio Tenenbaum, *Appearances of the Good: An Essay on the Nature of Practical Reason* (Cambridge University Press, 2007).
- Sergio Tenenbaum, 'The Guise of the Bad', *Ethical Theory and Moral Practice*, 21 (2018), 5–20.
- David Velleman, 'The Guise of the Good', *Noûs*, 26 (1992), 3–26.
- Bernard Williams, 'Internal and External Reasons' in *Moral Luck* (Cambridge: Cambridge University Press, 1981), 101–113.

FRANCESCO ORSI ([orsi@ut.ee](mailto:orsi@ut.ee)) is Associate Professor of Practical Philosophy at the University of Tartu, Estonia. His recent publications include *Value Theory* (London: Bloomsbury, 2015) and a special issue of *Philosophical Explorations* on 'The Modern Guise of the Good' (forthcoming).