

## Daniel Eggers, The motivation argument and motivational internalism

This is the peer-reviewed manuscript of the article “The motivation argument and motivational internalism”, published in *Philosophical Studies*, 172, 2015, pp. 2445–2467. The final publication is available at: <https://link.springer.com/article/10.1007/s11098-014-0420-y>

The *motivation argument* is widely considered to be the most important and most powerful argument for metaethical non-cognitivism.<sup>1</sup> There seems to be general agreement that the general idea behind this argument stems from David Hume – even though the question of whether Hume himself embraces a non-cognitivist view of moral judgement is a disputed one.<sup>2</sup> The *locus classicus* for Hume’s version of the argument is Book III, chapter 1, of the *Treatise of Human Nature*, in which Hume tries to prove, against moral rationalists such as Samuel Clarke, that our moral distinctions are not derived from reason.

[M]orality is [...] suppos’d to influence our passions and actions, and to go beyond the calm and indolent judgments of the understanding. And this is confirm’d by common experience, which informs us that men are often govern’d by their duties, and are deter’d from some actions by the opinion of injustice, and impell’d to others by that of obligation.

Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be deriv’d from reason; and that because reason alone, as we have already prov’d, can never have any such influence. Morals excite passions, and produce or prevent actions. Reason itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason. (Hume 1740/2000: 294)<sup>3</sup>

As can clearly be seen from the above passages, Hume’s argument works by way of two crucial claims, the one emphasising the motivational force of moral judgements (or morality in general), the other denying the workings of reason this very force: According to Hume, our moral distinctions cannot be the product of reason since they have an effect on our actions of which reason is incapable – for reasons already set out by him in the famous chapter “Of the influencing motives of the will” in Book II.

---

<sup>1</sup> For the term “motivation argument”, see Cohon 1997: 251; Zangwill 2009: 416; and Smith 2009: 105. The argument is also referred to as the “argument from motivation” (Brown 1988: 69) or the “non-cognitivist argument” (Shafer-Landau 2003: 119) and classified as an “internalist argument” (Strandberg 2011: 342), “antirealist argument” (Brink 1989: 37) or “internalist antirealist” argument (Brink 1989: 42).

<sup>2</sup> For a discussion of this issue, see, for instance, the contributions in Pigden 2009.

<sup>3</sup> Hume makes basically the same point in Section I of the *Enquiry concerning the Principles of Morals*: “The end of all moral speculations is to teach us our duty; and, by proper representations of the deformity of vice and beauty of virtue, beget correspondent habits, and engage us to avoid the one, and embrace the other. But is this ever to be expected from inferences and conclusions of the understanding, which of themselves have no hold of the affections nor set in motion the active powers of men? They discover truths: but where the truths which they discover are indifferent, and beget no desire or aversion, they can have no influence on conduct and behaviour.” (Hume 1751/1975: 172).

A modern, more formal rendering of the motivation argument that comprises these two claims under the names of *motivational judgement internalism* and *motivational Humeanism* has recently been provided by Russ Shafer-Landau:

1. Necessarily, if one sincerely judges an action right, then one is motivated to some extent to act in accordance with that judgement. (*Motivational Judgement Internalism*)
2. When taken by themselves, beliefs neither motivate nor generate any motivationally efficacious states. (*Motivational Humeanism*)
3. Therefore moral judgements are not beliefs. (*Moral Non-cognitivism*) (Shafer-Landau 2003: 121)

As Shafer-Landau rightly emphasises, the argumentative strategy characteristic of the motivation argument is employed by numerous modern metaethicists, even if these writers may not lay out the argument in exactly the way Shafer-Landau does. Moreover, the specification of the argument provided by Shafer-Landau, and his characterisation of its two premises, seem to be securely grounded in the terms of the contemporary metaethical debate. Thus it is widely accepted that the discussion about non-cognitivism is a discussion about whether moral judgements are beliefs or rather desire-like states, and it also seems uncontroversial that the two premises of the motivation argument amount to what is now usually being discussed as motivational internalism and motivational Humeanism. In fact, the view that motivational internalism and motivational Humeanism provide the two premises of the motivation argument is the reason why so much in contemporary metaethics revolves, in one way or the other, around these two positions.<sup>4</sup>

Yet, although Shafer-Landau's interpretation of the motivation argument is in line with the terms of the current metaethical debate, it still seems worth questioning. If we compare Shafer-Landau's version of the argument with the passage from Hume, two things spring to our attention. The first is that where Hume is relatively unspecific about the relation between morality and motivation, motivational internalism makes quite a strong claim, positing a necessary connection between making a moral judgement and being motivated to act in accordance with it. The second, related point is that where Hume seems to be concerned with empirical observations of a psychological kind, the purpose of contemporary defenders of the argument is usually to make a conceptual point, a point about our concept of moral

---

<sup>4</sup> References to motivational internalism as one of the premises of the motivation argument can be found, for instance, in Brink 1989: 37f.; Zangwill 2009: 416f.; Smith 2009: 106; and Schroeder 2010: 9-11. As already noted, Brink and Strandberg even refer to the motivation argument as an 'internalist argument' (see Brink 1989: 42; and Strandberg 2011: 342). See also Björnsson et al. who explicitly emphasize that the vast attention motivational internalism has attracted in the past is mostly due to its being part of the motivation argument (Björnsson et al. 2012: 125). The second premise of the argument is usually referred to not under the label 'motivational Humeanism', but under the label 'Humean theory of motivation'. It is important to note, however, that the latter term is used quite differently and that only the strong orthodox variant of this theory is equivalent to the second premise of the argument (for a discussion of different versions of the Humean theory of motivation, see Barry 2010).

judgement. Of course, the problem is not that Shafer-Landau's formulation of the argument is not an exact reproduction of Hume's argument, since this is not Shafer-Landau's intention. However, we may still ask whether the argument is given the form it needs to have if its purpose is to establish the non-cognitivist conclusion, and it seems that there are reasons for doubt here, reasons that have to do with the two aspects indicated above.

In what follows, I will try to show that, as it stands, Shafer-Landau's version of the motivation argument is flawed, and for two main reasons. First, it does not sufficiently distinguish between conceptual claims and empirical psychological claims. Without a clear distinction between these two types of claims, however, it is not clear whether the argument is a valid argument and really able to provide support for non-cognitivism. Secondly, and more importantly, the first premise, though widely accepted, is given a formulation that is, in crucial aspects, too strong. In order for the argument to work, the non-cognitivist does not need to defend the version of motivational internalism that Shafer-Landau suggests, but may settle for a more modest position that conceives of moral judgements not as necessarily motivating, but as potentially motivational. This second point is of special importance, given that Shafer-Landau's refutation of motivational internalism is based on rejecting it in favour of exactly this latter view. Therefore, if my criticism of Shafer-Landau's version of the motivation argument is correct, we may accept Shafer-Landau's arguments against motivational internalism as he defines it and still be able to defend the link between moral judgement and motivation that is needed for the non-cognitivist argument to go through.

It is important to emphasise that the application of this point goes beyond a mere critique of Shafer-Landau's objections to non-cognitivism. As has been suggested above, there can hardly be any doubt that the vast attention motivational internalism has attracted in the past is due to its role as one of the premises of the motivation argument. Therefore, if I am correct in thinking that the first premise is usually given an inappropriately strong formulation, this attention would seem to be somewhat misguided, even though the version of motivational internalism discussed in the metaethical literature might still represent an interesting metaethical claim in its own right. The least we should do, then, is to give much more attention to the weaker motivational claim that, in my view, provides the appropriate premise for the motivation argument and is readily accepted by Shafer-Landau and other opponents of non-cognitivism. Even if this claim may not be motivational internalism in the orthodox sense, it may be all the motivational internalism that a non-cognitivist defender of the motivation argument is ever going to need.

## 1. The Motivation Argument: Conceptual vs. Empirical

There are two slight problems with Shafer-Landau's rendition of the motivation argument that have not been addressed above because they can relatively easily be dealt with. The first problem is that the object of the conclusion, i.e., moral judgement, is not mentioned in any of the premises. Now it might seem that we can make up for this by simply adding the uncontroversial premise that judgements to the effect that an action is right are moral judgements. However, the conclusion that moral judgements are not beliefs only follows if judgements to the effect that an action is right are the *only* kind of moral judgements whatsoever, and this premise is not only controversial, but simply false. A better way to amend to argument, therefore, is to reformulate the first premise such as to make use of the term moral judgement and to thereby introduce the kind of generality which characterises the conclusion. The revised version of the first premise, which we will subsequently refer to as P1, would then reads as follows: "Necessarily, if one makes a moral judgement, then one is motivated, to some extent, to act in accordance with that judgement."

A second problem is that Shafer-Landau's rendition of the motivation argument presupposes that there is no such thing as a 'besire', that is, mental states that are, at the same time, belief- and desire-like. Once we allow such states, we may argue, again, that the argument's conclusion does not follow from its premises: For all the two premises tell us, moral judgements might be 'besires' and hence *also* beliefs – even if they may not be *just* beliefs. There are two ways in which we might make up for this presupposition. We could explicitly add a further premise to the effect that beliefs and desires are what Hume refers to as 'distinct existences'. Or we might slightly revise the conclusion so as to make clear that by "beliefs" we mean to refer to mental states that are just beliefs and not, at the same time, desires. Since I do not intend to explicitly discuss the widespread rejection of 'besires' in this paper – as worthwhile and important as such a discussion may be –, I consider the second option to provide a sufficient remedy for the issue at hand. The revised version of the conclusion, which I will subsequently refer to as K, would therefore read as follows: "Therefore moral judgements are not (just) beliefs."

In virtue of these preliminary revisions, Shafer-Landau's rendition of the motivation argument would assume the following form:

### **Motivation Argument (MA)**

- P1 Necessarily, if one makes a moral judgement, then one is motivated to some extent to act in accordance with that judgement.
- P2 When taken by themselves, beliefs neither motivate nor generate any motivationally efficacious states.
- K Therefore moral judgements are not (just) beliefs.

Now, a more serious worry with this argument, and one that applies not only to Shafer-Landau's version but to many modern renditions, has to do with the fact that these renditions turn what originally seems to have been empirical observations into conceptual claims. The problem here is not just that contemporary metaethicists tend to reinterpret Hume's argument as an argument about moral concepts, as opposed to an argument about psychological fact. It is that this reinterpretation is not carried out consistently, or at least not with sufficient clarity. In virtue of the necessary connection posited between moral judgement and motivation, the first premise, i.e. motivational internalism, is usually read as a claim about the concept of moral judgement: If a judgement does not give rise to motivation for action, it may not to be described as a moral judgement because such a complete absence of motivation for action is incompatible with our concept of moral judgement.<sup>5</sup> However, while a conceptual reading seems appropriate for the first premise, it does not seem the most natural reading of either the second premise or the conclusion. Given that P2 does not appeal to the notion of necessity, it seems to be better interpreted as a claim about (contingent) psychological fact. The same applies to the conclusion that moral judgements are not (just) beliefs: For all the conclusion tells us, our concept of moral judgement may allow some or all moral judgements to be beliefs – even if, as a matter of contingent psychological fact, they are not.<sup>6</sup>

The problem with the ambiguity of the second premise and the conclusion is that it casts doubt on whether the motivation argument can really provide support for non-cognitivism as it is usually understood. To be sure, the combination of conceptual and empirical claims does not necessarily result in an invalid argument. While this could happen easily, it does not do so in the present case: Even under the partly empirical reading sketched above, it is impossible for both premises of the argument to be true without the conclusion being true as well. If we admit that our concept of moral judgement commits us to conceive of

---

<sup>5</sup> That motivational internalism makes a conceptual claim is emphasized, for instance, by David Brink and Michael Smith (see Brink 1989: 42; and Smith 1994: 61). See also Björnsson et al. who emphasize the conceptual nature of motivational internalism, but also note the recent tendency to discuss internalism as an *a posteriori* claim (see Björnsson et al. 2012: 126 and 132).

<sup>6</sup> The only way to argue that, even as they stand, the second premise and the conclusion need to be read as conceptual claims, or that they at least imply such claims, were to generally exclude the possibility of our concepts being erroneous and in the need of modification. We could do so either by presupposing some kind of pre-established harmony between our concepts and the world or by employing a normative understanding of concepts according to which our concepts are constituted by true assumptions only. However, this take on concepts is far from being generally accepted, and there seem to be strong arguments that speak against it.

moral judgements as motivationally efficacious and that, as a matter of psychological fact, beliefs are not motivationally efficacious in this sense, we are logically forced to admit that, as a matter of psychological fact, moral judgements are not (just) beliefs. In denying this psychological fact, we would either contradict the second premise or employ the concept of moral judgement in a way that is incompatible with our acceptance of the first.

The real problem is that, in its given form, the argument suggests that non-cognitivism is an empirical-psychological thesis while in the current metaethical debate, non-cognitivism is usually understood as a conceptual thesis, being primarily concerned with the meaning of moral terms. Recently, Richard Joyce has strongly emphasised the difference between metaethical non-cognitivism and current approaches in empirical psychology (Haidt, Greene) to which he refers as examples of “psychological emotivism” (Joyce 2009: 34). Joyce defines non-cognitivism as the view that moral utterances do not express beliefs but rather desire-like states and stresses that the notion of *expression* employed by non-cognitivists is a *non-causal* notion that refers to linguistic convention rather than empirical psychological fact: It is possible, according to this notion, to express a mental state without actually being in that state, just as one may express regret by saying ‘sorry’ without sincerely feeling it (see Joyce 2009: 32). Metaethical non-cognitivism, then, is neither identical to the kind of causal or empirical theories of moral judgement Joyce labels “psychological emotivism”, nor does it presuppose, strictly speaking, that any such theory is true.

Much in line with this, most commentators take the non-cognitivist thesis to be concerned not with empirical aspects of human psychology, but with semantic or pragmatic aspects of moral language and moral discourse. Examples include Michael Smith, who refers to non-cognitivism as a claim about the functional role of moral sentences (see Smith 2001: 93), and Michael Ridge, who conceives of non-cognitivism as a view about what moral sentences are conventionally taken to express (see Ridge 2007: 53). Similarly, John O’Leary-Hawthorne and Huw Price, though explicitly distinguishing between a semantically defined form of non-cognitivism and a psychologically defined form, take even the latter to amount to a linguistic claim about the nature of moral discourse (see O’Leary Hawthorne/Price 1996: 276f.). In fact, in his more extensive discussion of non-cognitivism, Shafer-Landau himself presents non-cognitivism as a claim concerning the purpose of moral language and the point of moral discourse (see Shafer-Landau 2003: 20), thereby suggesting that the debate between cognitivism and non-cognitivism is not so much about the empirical nature of moral judgements as about how we conceptualise them within the framework of our linguistic practices. It should also be noted that expressivism, which is widely considered to be the most

important variant of non-cognitivism,<sup>7</sup> is generally understood as a theory about the meaning of moral words and not as an empirical psychological theory. In line with this, Strandberg, who discusses the motivation argument as an argument for *expressivism* rather than non-cognitivism, explicitly interprets the conclusion of the motivation argument as a conceptual claim (see Strandberg 2011: 346).

Therefore, even though the exact relationship between metaethical theories such as non-cognitivism and expressivism on the one hand and empirical psychological theories of moral judgement on the other has been neglected in the past and stands in need of clarification, it seems that in order to do justice to the previous debate over non-cognitivism, we need to conceive of the conclusion of the motivation argument for non-cognitivism as a conceptual rather than an empirical psychological thesis. Yet, once we do so and rephrase the conclusion of Shafer-Landau's version of the argument accordingly, the argument ceases to be a valid argument:

**Motivation Argument\* (MA\*)**

- P1 Necessarily, if one makes a moral judgement, then one is motivated to some extent to act in accordance with that judgement.
- P2 When taken by themselves, beliefs neither motivate nor generate any motivationally efficacious states.
- K\* Necessarily, moral judgements are not (just) beliefs.

Even if it is combined with the conceptual claim that moral judgements necessarily motivate, the psychological fact expressed by P2 does not suffice to establish the conceptual conclusion that, *necessarily*, moral judgements are not (just) beliefs. Since, in principle at least, P2 allows for the conceptual possibility of there being motivating or motivationally efficacious beliefs, it cannot commit us to principally denying the possibility of beliefs being moral judgements. This can perhaps best be seen if we enrich P2 such as to explicitly mention the conceptual possibility of motivationally efficacious beliefs and enrich Shafer-Landau's original conclusion such as to explicitly mention the conceptual possibility of beliefs being moral judgements:

- (a) Necessarily, if one makes a moral judgement, then one is motivated to some extent to act in accordance with that judgement.
- (b) As a matter of psychological fact, when taken by themselves, beliefs neither motivate nor generate any motivationally efficacious states (but there could be beliefs that motivate or generate motivationally efficacious states).

---

<sup>7</sup> See, for instance, Schroeder 2010: 65. See also Joyce's claim that most metaethicists treat the terms 'expressivism' and 'non-cognitivism' synonymously (Joyce 2009: 34).

- (c) As a matter of psychological fact, moral judgements are not (just) beliefs (but there could be moral judgements that are beliefs).

Obviously, all three claims can be true at the same time. Yet, given that the first two claims are equivalent to P1 and P2 and that the third claim is equivalent to a denial of K, it follows that even if both premises of MA\* are true, its conclusion may still be false.

In order to salvage the argument, then, we need to turn the second premise into a conceptual claim as well. In my view, both the fact that motivational Humeanism may be taken as either an empirical or a conceptual claim and the fact that it supports non-cognitivism as usually understood only if we conceive of it in the latter way have been given far too little attention in the past. If we acknowledge both of these facts, we are left with the following situation. Either we stick to the view that non-cognitivism is a conceptual thesis. In this case, the motivation argument meant to support non-cognitivism needs to be a wholly conceptual argument, consisting of three unambiguously conceptual claims:

**Motivation Argument, conceptual (MA<sub>C</sub>)**

- P1 Necessarily, if one makes a moral judgement, then one is motivated to some extent to act in accordance with that judgement.
- P2 Necessarily, when taken by themselves, beliefs neither motivate nor generate any motivationally efficacious states.
- K Necessarily, moral judgements are not (just) beliefs.

Alternatively, we may part company with the metaethical debate and conceive of non-cognitivism as an empirical psychological thesis (of course, we may also be interested in this empirical psychological thesis without labelling it ‘non-cognitivism’ in the first place). Even then, however, we should not stick to Shafer-Landau’s original version of the motivation argument, valid as it may be. One reason for reformulating the original argument consists in the slight lack of clarity that results from the fact that conceptual and empirical claims are not explicitly identified and distinguished. However, a further and more serious problem is that, although the conceptual claim expressed by P1 clearly supports the argument’s conclusion, it is not really *needed* for establishing the conclusion. We may just as well settle for a weaker, non-conceptual version of P1 without affecting the validity of the argument. There are two kinds of considerations, therefore – considerations of clarity and considerations of parsimony –, which suggest that if the conclusion we are interested in is the empirical-psychological claim that moral judgements are not just beliefs, we should rephrase the motivation argument in the following manner:



### **Motivation Argument, empirical (MA<sub>E</sub>)**

- P1 As a matter of psychological fact, if one makes a moral judgement, then one is motivated to some extent to act in accordance with that judgement.
- P2 As a matter of psychological fact, when taken by themselves, beliefs neither motivate nor generate any motivationally efficacious states.
- K As a matter of psychological fact, moral judgements are not (just) beliefs.

In what follows, I will exclusively focus on the conceptual variant of the motivation argument because only this variant of the argument is concerned with non-cognitivism as usually understood. My aim will be to more precisely examine and characterise the premises that are needed for this argument to work and to demonstrate that, even for this variant, the first premise is given a formulation that violates the principle of parsimony.

### **2. The motivation argument and motivational internalism**

Considerations of parsimony not only suggest that we should further revise P1 if it is to function as the first premise of MA<sub>E</sub>. They also suggest that we should further revise it if it is to function as the first premise of MA<sub>C</sub>. If MA<sub>C</sub> is meant to demonstrate that moral judgements cannot be beliefs, and if one premise of the argument claims that beliefs cannot be motivationally efficacious at all, then what is needed in order to establish the intended conclusion is the claim that moral judgements *can be* motivationally efficacious, that is, possess the very property that beliefs, when taken by themselves, necessarily lack. Of course, it will not do to just claim that *some* moral judgements are or can be motivationally efficacious because this would leave us with the possibility that some moral judgements, namely the ones that are *not* motivationally inefficacious, *are*, in fact, beliefs. What the non-cognitivist defender of MA<sub>C</sub> needs to claim, therefore, is that *all* moral judgements can be motivationally efficacious, that is, that all moral judgements have the capacity of motivating the person making them. As P2 suggests, there are two ways in which we may think of this capacity. Moral judgements may either have it in virtue of being motivational states themselves or in virtue of being able to produce, by themselves, such motivational states. What the defender of the non-cognitivist argument cannot allow is only that the motivational capacity is explained by recourse to a prior motivational state whose existence does not in any way depend on the moral judgement – because under this interpretation, beliefs would obviously possess the capacity as well.

If, for lack of a better title, we refer to these two ways in which moral judgements may be motivationally efficacious by saying that moral judgements are *potentially motivational*, we should expect the first premise of MA<sub>C</sub> to amount to the claim that, necessarily, moral judgements are potentially motivational. In fact, it seems that this also the appropriate conceptual counterpart to Hume's empirical observation: While Hume clearly thinks that our moral judgements (or our moral distinctions) can actually motivate us, and that they do on a regular basis, there is no reason to attribute to him the view that our moral judgements must always succeed in motivating us. However, while P1 may be taken to be just another way of saying that, necessarily, moral judgements are potentially motivational, this is not the interpretation that is usually applied to motivational internalism, and quite decidedly, it is not what Shafer-Landau has in mind.

Shafer-Landau concedes that moral judgements have a "deep pull" (Shafer-Landau 2003: 119) on us and that any theory of moral judgement, such as Shafer-Landau's own version of moral realism, must be capable of explaining this fact. Shafer-Landau's attempt to provide such an explanation without, at the same time, opening the doors for non-cognitivism is based on distinguishing two different claims: the claim that moral judgements are "intrinsically motivating" (Shafer-Landau 2003: 122) and the claim that moral judgements necessarily motivate. According to Shafer-Landau, the first premise of the motivation argument, i.e. motivational internalism, is equivalent with the latter claim, a claim which, in his view, we can perceive to be false once we confront us with examples of the so-called amoralist: a person who makes a moral judgement but is completely unmoved by her judgement. Thus, while Shafer-Landau admits that moral judgements are "intrinsically motivating", he denies that they necessarily motivate and hence claims the first premise of the motivation argument to be false.

Now there are two problems with Shafer-Landau's distinction. The first is that it is not clear what the distinction actually amounts to and whether we have reason to accept it. The lack of clarity is partly due to ambiguities in the way Shafer-Landau sets up the distinction. Shafer-Landau's definition of what it means for a mental state to be "intrinsically motivating" seems relatively straightforward and, in fact, equivalent to our idea of a mental state's being potentially motivational or having the capacity to motivate: a mental state is "intrinsically motivating", according to Shafer-Landau, if it motivates "in virtue of its own nature and content" (Shafer-Landau 2003: 147), either entirely without the aid of other motivationally

efficacious states or by means of generating such states by itself.<sup>8</sup> Yet, as Shafer-Landau is keen to emphasise, to say that a mental state is “intrinsically motivating” in this sense is not to say that it actually motivates in a given situation: Owing to competing mental states, to physical exhaustion or to severe depression, an “intrinsically motivating” state may exert “no motivating force whatsoever” (Shafer-Landau 2003: 147), which means that mental states may be “intrinsically motivating” without being, necessarily, motivating. In order to support this claim, Shafer-Landau distinguishes two ways to make sense of what he calls a ‘motivational defeater’:

There are two relevant options. The first claims that a defeater operates by overriding an existing motivation, which nevertheless remains to some small degree. Following a similar debate about reasons and values, we can call such motivation *pro tanto*. Such motivation, when defeated, does not disappear completely. Some remnant remains, however weak or actually causally inefficacious. Alternatively, we might conceive of a defeater as extinguishing the motivation that, in other contexts, would exist were it not for the presence of the defeater. Again, we can borrow terms about practical reason and value and call such motivations *prima facie*. (Shafer-Landau 2003: 148)

Now while we may well find the distinction between *pro tanto* motivation and *prima facie* motivation plausible in its own right, it tends to somewhat obscure the distinction it is actually meant to enlighten, namely the distinction between states that are “intrinsically motivating” and states that necessarily motivate. Thus, the fact that a certain motivation is “extinguished” by a motivational defeater does not seem to provide a sufficient basis for saying that the mental state in question did not actually motivate or did not exert any motivational influence whatsoever. The extinction metaphor – the fire extinguisher being a case in point – rather suggests that the mental state did succeed, at first, in exerting motivational influence, but that the resulting motivation was completely removed once the motivational defeater exerted its opposed influence. To say that motivation was extinguished and to say that motivation did not exist in the first place seem to be two different things, and it is only the latter fact that seems incompatible with a mental state being necessarily motivating: The fact that the motivation produced by a certain mental state is quickly extinguished and extinguished completely, does not refute the claim that the state in question is a necessarily motivating state; it only contradicts the claim that the motivation exerted by such a mental state necessarily *remains* motivationally efficacious, at least partly, in the presence of possible motivational defeaters.

---

<sup>8</sup> My reason for not following Shafer-Landau’s usage is that some contributors to the internalism debate employ the term ‘intrinsically motivational’ quite differently and take it to only refer to mental states which motivate without the help of other mental states (see, for instance, Svavarsdóttir 1999: 163; see also Zangwill 2009: 416). It seems less confusing, then, to refer to the two possible ways in which a mental state may exert motivational influence by saying that a mental state is *potentially motivational*.

Shafer-Landau's account of *prima facie* motivation, therefore, is not really helpful when it comes to understanding the difference between "intrinsically motivating" and necessarily motivating states (or, for that matter, between potentially motivational and actually motivating states).<sup>9</sup> It seems, then, that something more, and, in fact, something different, needs to be said in order to make sense of the distinction Shafer-Landau is concerned with. However, one initially might want to express some doubts as to whether this distinction can be made sense of at all. What we need is the idea of a motivational defeater that hinders a potentially motivational state from exerting any actual motivational force whatsoever – not by extinguishing motivation, but by barring motivation from coming into existence in the first place. Yet, it is not clear how we are to think of this kind of motivational defeater because it is not clear how we are to conceive of a potentially motivational state that does not in the least motivate: in what, if not in the exertion of some kind of motivational influence, be it ever so slight, can the property of being a potentially motivational state consist? After all, what we are looking for is a certain state of mind, and since motivation, or the exertion of motivation, happens in the mind as well – how can a potentially motivational state be instantiated in the mind without motivation being instantiated at the same time?

However, that there is at least one way of solving this apparent paradox can be seen if we remind ourselves of how we defined the concept of a potentially motivational mental state. If one of the two ways in which a mental state may be potentially motivational is that it has the capacity of generating a second, motivationally efficacious state, then it seems that the kind of motivational defeater we are looking for could just be something that prevents this kind of generation from taking place. Following Shafer-Landau, we may, for example, conceive of the potentially motivational state as some kind of evaluative belief. Once we assume that, in order to actually motivate, evaluative beliefs need to produce desires, all we have to do is to additionally assume that under certain conditions, evaluative beliefs may fail to do so. In this case, we would have the instantiation of a potentially motivational state but not the instantiation of any actual motivation – because the independent mental state by which motivation is constituted is not instantiated.

Now if this is one way to make sense of Shafer-Landau's distinction between 'intrinsically' and 'necessarily' motivating states, then it might seem that his way of

---

<sup>9</sup> The same can be said about Shafer-Landau's definition of motivational externalism as the "contradictory" (Shafer-Landau 2003: 145) to motivational internalism. According to this definition, motivational externalism amounts to the claim "that it is at least conceptually possible for a sincere normative judgement to entirely lack motivational power" (Shafer-Landau 2003: 145). Yet, it seems that we may already attribute motivational *power* to a normative judgement if it is "intrinsically motivating", that is, has the capacity to give rise to motivation, which means that one need not subscribe to the supposedly different and stronger claim that normative or moral judgements are necessarily motivating in order to claim motivational externalism to be false.

interpreting the first premise of the motivation argument could be biased against the non-cognitivist and towards some kind of moral realism or moral cognitivism. However, the opposite is true. If we want the motivation argument to *establish* the truth of non-cognitivism, and not just to presuppose it, we should strive to formulate each premise of the argument in a way that, if taken by itself, is compatible with non-cognitivism being false and cognitivism being true. Therefore, if moral cognitivism allows us to make sense of the distinction between potentially motivational and actually motivating mental states, we ought not to set up P1 in a way that makes this distinction collapse. Moreover, it is far from being clear that the non-cognitivist cannot make sense of the distinction as well. As long as the non-cognitivist does not take moral judgements to be ordinary first-order desires – as few if any modern non-cognitivists seem to do –, but rather conceives of them as higher-order desires or states of norm-acceptance, it may make good sense for him to view moral judgements as mental states that motivate somewhat indirectly – for example, by generating ordinary first-order desires –, and to allow for the possibility that, under certain circumstances, this generation may be foreclosed. Furthermore, there are, arguably, other ways to conceive of a potentially motivating state that fails to actually motivate for which this question does not arise in the first place. One could, for example, conceive of it as a *disposition* which is blocked by certain external factors, where this disposition is not thought of as one that, in the absence of such hindering factors, motivates by creating independent motivational states but as one that motivates in its own right.<sup>10</sup>

Shafer-Landau seems to be right, then, in emphasizing the distinction between potentially motivational and actually motivating states, even though his way of elaborating on the distinction may not be wholly satisfactory. The real problem with his analysis, however, is that the distinction undermines Shafer-Landau's own critique of motivational internalism. Once we follow Shafer-Landau and accept the distinction, it seems that P1 no longer provides an appropriate formulation of motivational internalism, at least where motivational internalism is thought of as the first premise of the motivation argument. Given the strength of P2, which, in effect, rules out *any* possibility of a belief being potentially motivational, the defender of the motivation argument does not have to claim that, necessarily, moral judgements motivate, in order to establish that, necessarily, moral judgements are not (just) beliefs. All he has to do is to claim that, necessarily and unlike beliefs, moral judgements are potentially motivational.

---

<sup>10</sup> I am grateful to an anonymous reviewer for pointing out this possibility to me.

What we may conclude from our discussion, therefore, is the following: If we accept the systematic distinction between potentially motivational and actually motivating mental states (as we seem to have good reason to do), then taking motivational internalism in its usual meaning as the first premise of the motivation argument means to violate the principle of parsimony and to burden the non-cognitivist with claims he may rightfully reject. It seems, then, that in order to provide an appropriate rendition of the motivation argument, we should give up P1 in its original form and to rephrase it such as to present moral judgements as potentially motivational.

However, before we revise Shafer-Landau's rendition of the motivation argument accordingly, one further question needs to be discussed, namely what kind of motivation it is that the defender of the motivation argument needs to postulate. Shafer-Landau's relatively unspecific, but commonplace reference to the motivation "to act in accordance" with one's judgment leaves open what the exact content or object of motivation is. More specific formulations of the internalist claim usually posit a direct link between the content of the moral judgement and the content of motivation, for example, by defining motivational internalism as the claim that "Necessarily, if a person judges that she ought to  $\phi$ , then she is, at least to some extent, motivated to  $\phi$ ." (Strandberg 2011: 342).<sup>11</sup>

As I have already emphasised above, if the internalist claim is to function as the first premise of the motivation argument, then the antecedent of this latter formulation does not set the bar high enough: Since non-cognitivists want to claim that all, and not only some, kinds of moral judgements are desire-like states, they cannot rest with making a conceptual point about certain forms of moral judgements, such as first-person ought judgements, but need to claim that *all kinds* of moral judgements are capable of giving rise to motivation. What is equally important, however, is that the consequent of the definition sets the bar too high: If the second premise of the motivation argument is the strong claim that, necessarily, beliefs are not potentially motivational at all and therefore unable to exert any motivational influence whatsoever, then all the first premise needs to claim is that, necessarily, moral judgements are capable of giving rise to *any kind of motivation whatsoever* – because this claim is sufficient, in combination with the second premise, to exclude the possibility of moral judgements being (just) beliefs. The non-cognitivist defender of the motivation argument, therefore, is not committed to thinking that the moral judgement 'It is morally right to  $\phi$ ' is necessarily capable of motivating the person making the judgement to  $\phi$ . He may just as well think that the judgement is only capable of motivating the person to get others to  $\phi$ , or not to suffer

---

<sup>11</sup> For similar definitions, see Brink 1989: 40; Smith 1994: 61; Mele 1996: 727; Dreier 2000: 619; and Björnsson et al. 2012: 126.

other people's  $\phi$ -ing without protest – as long as he admits that all moral judgements are equally capable of giving rise to some motivation of this kind.

There are two conclusions to be drawn from this. One is that we should not embrace a formulation of the kind “Necessarily, if a person judges that she ought to  $\phi$ , then she is, at least to some extent, motivated to  $\phi$ .” in order to specify the content of the motivation associated with moral judgements, but rather stick to Shafer-Landau's original formulation. A further conclusion, however, is that we need to give the phrase “in accordance with” a broad interpretation: In order to be motivated to act in accordance with one's moral judgements, in the sense required for the motivation argument to go through, one need not necessarily be motivated to carry out the exact kind of action that the judgement prescribes – it is sufficient to have *some* kind of judgement-related motivation.

In light of the preceding considerations, then, it seems that an appropriate rendition of the motivation argument should have the following form:

#### **Revised Motivation Argument, conceptual (rMA<sub>C</sub>)**

- P1' Necessarily, moral judgements are potentially motivational, i.e., they can motivate the person making them to act in accordance with the judgement or generate judgement-related motivationally efficacious states by themselves.
- P2 Necessarily, when taken by themselves, beliefs neither motivate nor generate any motivationally efficacious states.
- K Necessarily, moral judgements are not (just) beliefs.

P1' is all the defender of the motivation argument is committed to, and unlike P1, it is immune to Shafer-Landau's objections and to other objections launched against motivational internalism as well. It seems, therefore, that rMA<sub>C</sub> is what not-cognitivist defenders of the motivation argument should employ because it salvages the argument's main point while avoiding unnecessary theoretical baggage.

### **3. The truth about internalism?**

In revising the first premise of the motivation argument in the above way, we have obviously come quite some way from how motivational internalism has traditionally been defined. There are a couple of questions, therefore, which we need to address. One question is whether there is any independent rationale for P1': Is there a reason, apart from trying to formulate the most parsimonious and strongest version of the motivation argument, for why non-cognitivists (or, for that matter, motivational internalists) would want to subscribe to P1', or do they perhaps have reasons to prefer the stronger claim P1 to P1', even if this makes their

argument for non-cognitivism more vulnerable? Another question to be asked is whether we should still think of the revised first premise of the motivation argument as a statement of motivational internalism.

Now the first thing to be emphasised in response to the first question is that P1' is entailed by P1. In order to be able to claim, with P1, that moral judgements necessarily motivate to act in accordance with that judgement, one is obviously committed to the claim that moral judgements have the potential to motivate. It would be irrational, therefore, to endorse P1, but to want to reject P1'. However, it would be a bit premature to simply conclude from this that we must have a sufficient rationale for P1' wherever we have a sufficient rationale for P1. If all the reasons that speak in favour of P1' were in fact reasons that also speak in favour of P1, then one could at least argue that the substitution of P1' for P1, which we have advocated above, is, in a sense, futile. After all, the idea behind trying to find an appropriate formulation for the first premise of the motivation argument is not merely to formulate a valid argument and to comply with the dictates of parsimony. It is also to formulate a premise whose truth can independently be defended. If, however, the truth of P1' cannot be defended without recourse to P1, then it seems that the non-cognitivist defender of the motivation argument may just as well stick to P1.

Yet, the fact that Shafer-Landau's objection to motivational internalism creates problems for P1, but not for P1', already suggests that we do not have this kind of perfect justificatory equivalence here. Moreover, we would quite generally expect a logically weaker claim to be easier to justify than a logically stronger one and think that there should at least be *some* reasons that speak in favour of the weaker claim, but not, at the same time, in favour of the stronger. Now, while the aim of this paper is not to provide an independent evaluation of motivational internalism and a thorough comparison of the respective merits P1 and P1', I would like to develop at least a tentative idea of what such independent reasons in favour of P1' could look like. In order to do so, let us have a look at the kind of argument which is usually employed in defences of motivational internalism and which I will refer to as the 'argument from social practice'. As the name indicates, the argument starts from certain aspects of our common moral experience, which are neatly summarized in the following example taken from Mark Schroeder's book *Non-Cognitivism in Ethics*:

Suppose that you and your friend have been discussing whether she ought to donate money to CARE, a highly rated international poverty-fighting organization. She thinks not. Maybe she thinks that her money works more effectively to fight poverty if given to Oxfam, or maybe she thinks that it is more important to donate to the political campaigns of the party she believes will make a larger difference than she can with her donation. Or maybe she simply thinks that it is her right to spend her money as she pleases, and prefers to spend it on soy lattes and



sugar-free biscotti.

Whichever of these is the case, part of the point of engaging in this discussion with her is probably that you expect it to make a difference, if you convince her. Suppose, for example, you really do convince her that you are right, and that she ought to donate money to CARE. If the next thing that happens is that a representative of CARE comes knocking on the door soliciting donations, you will expect that she will not be indifferent. Having decided that donating is what she ought to do, you will expect her to at least feel some motivation to donate. Before you convinced her, maybe she felt indifferent, but after you convince her, you expect her to feel indifferent no longer. If your friend really feels no such motivation, you are likely to wonder whether she was really just being insincere in agreeing with you, perhaps just hoping that you would get off her back. (Schroeder 2010: 9f.)

There are three crucial, but separate aspects of our common moral practice which we may infer from Schroeder's example: a) we expect other people to act in accordance with their moral utterances, b) we know from experience that people frequently do not act in accordance with their moral utterances, and c) in such cases, we tend to question the sincerity of their moral utterances.

There are four theoretical positions regarding the connection between moral judgement and motivation which we ought to distinguish at this point: Motivational Externalism, Motivational Internalism I (equivalent to P1'), Motivational Internalism II (equivalent to P1), and Motivational Internalism III (also known as 'strong motivational internalism'): the claim that, necessarily, if we make a moral judgement, we act in accordance with that judgement. Now it seems hard to deny that Motivational Externalism can provide a plausible explanation for why we often witness people acting in accordance with their moral utterances. Externalists may, for example, reduce such actions to self-interest, altruism or an independent and prior desire to do the morally right thing. In virtue of this, externalists can also provide some explanation for why we *expect* other people to act in accordance with their moral utterances, that is, give an explanation for (a). Moreover, Motivational Externalism does not seem to have any problems to account for (b), either. After all, the demands of morality need not always converge with our interests – as a matter of fact, we would rather expect them not to do so –, and not all individuals need to have the desire to do the right thing or the same amount of altruistic motivation in the first place. However, it seems that the externalist cannot, on the basis of this, give a very plausible explanation for (c): If we take moral behaviour to be ultimately based on self-interest, altruism or the desire to do the right thing, then why do we tend to think that an individual who acted contrary to her moral utterance may have been insincere? Why do we not just think, in accordance with what we have just said, that the person lacked the desire to do the right thing or had strong self-interested reasons not to act in accordance with her sincere moral utterance? That we tend to

think, instead, that the person did not honestly subscribe to her moral utterance in the first place suggests that our expectation might not solely be the result of experience or custom, but in some way related to our conception of what a moral judgement is – which is the fundamental idea of motivational internalism.

It seems, then, that motivational internalism may be in a better position to simultaneously account for all the three aspects of our social practice described above. The important thing to note, however, is that, as it stands, our argument suggests quite a strong form of motivational internalism: the view that, necessarily, if we make a moral judgement, we act in accordance with that judgement (Motivational Internalism III). After all, it is this variant of motivational internalism that provides the most straightforward explanation for (c), which means that the argument from social practice would not provide any genuine support for either P1 or P1'. However, few if any non-cognitivists today would want to commit themselves to an endorsement of Motivational Internalism III which is considered implausibly strong by virtually every contributor to the internalism/externalism debate. It seems, therefore, that the argument from social practice needs some kind of supplement that allows us to rule out Motivational Internalism III, and this supplement is often provided by what I will subsequently refer to as the 'argument from introspection'.

One important aspect of our moral self-experience is that, when faced with a moral duty or obligation, we sometimes give in to temptation, as when eat the second chocolate bar instead of keeping it for our absent little brother – like the donator of the two chocolate bars asked us to do and we think we ought to do. In such cases, we do not act according to our moral judgement. However, we do not usually think in such cases that we never really believed it was our duty to keep the chocolate bar in the first place, either. Rather, we have the feeling that our motivation to act in accordance with our moral judgement was simply overridden by some other, stronger motivation, in this case the motivation to enjoy the second chocolate bar. However, Motivational Internalism III seems to have no place for this kind of phenomenon.

The most plausible interpretation of the three aspects of our social practice (a), (b) and (c) and the above aspect our moral self-experience taken together, therefore, seems to be provided by an internalist claim that is weaker than Motivational Internalism III and allows for the possibility that the motivation associated with a moral judgement may be overridden. Yet, it is important to note that Motivational Internalism II (P1) and Motivational Internalism I (P1') fare equally well in this regard. Not only can both claims allow for our tendency to question the sincerity of people who fail to act in accordance with their moral utterances,

given that they both conceive of the connection between judgement and motivation as a conceptual connection and not just an empirical regularity. They also both allow for moral judgements which fail to issue in action because of overriding non-moral motivation. The only difference is that the defender of Motivational Internalism II (P1) *needs to* interpret our failure to act in accordance with our moral judgement in terms of such overriding non-moral motivation, whereas the defender of Motivational Internalism I (P1') has one further option: he may either argue that the motivation associated with the moral judgement was overridden by some other motivation or that the moral judgement failed to actually give rise motivation, in one of the ways suggested in section 2.

It seems, therefore, that the argument often appealed to in defence of motivational internalism, namely the argument from social practice supplemented by the argument from introspection, does not provide us with any obvious reasons for preferring Motivational Internalism II (P1) over Motivational Internalism I (P1'). However, we may rightly wonder whether the reverse also holds. Thus, we might think that the difference just described speaks in favour of Motivational Internalism I (P1'), at least when we enrich the argument from introspection with one further aspect of our moral self-experience. The case of temptation sketched above is one where we face a negative moral duty (in this case the duty not to eat the second chocolate bar), but fail to comply with the duty. For such cases, an explanation in terms of overriding non-moral motivation seems most plausible. However, we also know of cases where we face positive moral duties which require us to actually do something, and it is often claimed that in such cases, moral motivation is even more of a problem than in cases of purely negative duties. Yet, very little in the way we experience ourselves when we fail to comply with such positive moral duties suggests that the problem must *always* be one of overriding non-moral motivation. Quite often, we do not feel the kind of positive urge to violate our moral duty that we feel in the chocolate bar example at all. For all our self-experience tells us, we may simply not have been motivated by our moral judgement in the first place. While Motivational Internalism I (P1') allows us to try to defend this aspect of our common moral experience, Motivational Internalism II (P1) requires us to substantially re-interpret it and to assume that there is overriding motivation at work where we feel no motivation at all.

It might seem, therefore, that the rationale for Motivational Internalism II (P1) might be somewhat unstable. If the defender of Motivational Internalism II (P1) generally refuses to admit evidence from moral self-experience and introspection and confines himself to what I have referred to as the argument from social practice, then it seems that what he is in effect

supporting is Motivational Internalism III rather than Motivational Internalism II. If, on the other hand, he does admit evidence from moral self-experience and introspection in order to fend off Motivational Internalism III, he seems to open the door for an enhanced argument from introspection that supports Motivational Internalism I (P1') rather than Motivational Internalism II.

Now, as already indicated, the purpose of the above discussion is not to take a stand with regard to the internalism/externalism debate or the question of whether the most plausible form of motivational internalism conceives of moral judgements as actually motivating or potentially motivating. What needed to be shown was only that we have reasons to assume that there is an independent rationale for P1' which can somewhat underwrite our argument for why a non-cognitivist defender of the motivation argument should choose P1' as the first premise of his argument. If nothing else, the kind of provisional argument for P1' developed above suggests that there are, indeed, reasons that speak in favour of P1' without, at the same time, speaking in favour of P1. In fact, it seems that these reasons do even conform to the spirit of the standard defence of motivational internalism. Nothing in what I have referred to as the argument from social practice and the argument from introspection is incompatible with P1', or supports P1 at the expense of P1'. Moreover, it is the very appeal to introspection necessary to argue for preferring P1 over 'strong motivational internalism' which promises a further independent argument in support for P1'. There seem to be good reasons, therefore, for the non-cognitivist endorser of the motivation argument not to stick to P1, but to give it up in favour of P1'. The decision for P1' as the first premise of the argument not only results in a more parsimonious argument that is immune to the objection raised by Shafer-Landau. It also promises additional resources for defending the truth of motivational internalism in its own right.

Our considerations also suggest why, apart from trying to be faithful to the original Humean argument, the non-cognitivist has reasons to stick to the strong orthodox version of the Humean theory of motivation as the second premise of the argument. By embracing a weaker form of the Humean theory – like, for example, Michael Smith's version which takes desires to be a necessary component or constitutive of motivation, but allows beliefs to give rise to new desires and hence to motivate without the aid of prior motivational states –, the non-cognitivist defender of the motivation argument commits himself to conceiving of moral judgements as actually motivating: If he conceives of them only as potentially motivational and endorses the weaker version of the Humean theory, according to which *both* beliefs *and* desires are potentially motivational, then the non-cognitivist needs an independent argument

for why moral judgements are not just beliefs, given that the motivation argument would then fail to provide any such argument.

Moreover, even if the non-cognitivist bites the bullet and takes moral judgments to actually motivate, this does not mean that he can happily stick to the weaker version of the Humean theory. A version of the argument that has P1 as its first premise and the weaker version of the Humean theory as its second premise is not a valid argument anymore:

- P1      Necessarily, if one makes a moral judgement, then one is motivated to some extent to act in accordance with that judgement.
- P2'     Necessarily, beliefs are potentially motivational, but may fail to actually motivate.
- K\*      Necessarily, moral judgements are not (just) beliefs.

It is compatible with the truth of P1 and P2' that moral judgements are a sub-set of beliefs, namely all or some of those beliefs that actually motivate. It is impossible, therefore, to infer from P1 and P2' that, necessarily, moral judgements are not (just) beliefs – just as it is impossible, for instance, to infer that (c) necessarily, mothers are not women, from the fact that (a) necessarily, mothers have at least one child, and the fact that (b) women are capable of having children, but need not have any children.

Conclusion K only follows if, in addition to conceiving of moral judgements as actually motivating (P1), we also assume that they motivate in a certain way, namely without the help of any further mental states or by way of a motivational force that is intrinsic to the moral judgement itself in a more narrow sense of the term. Only if we make this further assumption and attribute to moral judgements a property that beliefs lack even according to the weaker version of the Humean theory, we may stick to the weaker version of the Humean theory and still conclude that moral judgements cannot just be beliefs.<sup>12</sup> However, interpreting the first premise of the motivation argument thus restrictively seems problematic for two reasons. The first is that by subscribing to the restrictive interpretation of P1, one rules out important variants of both cognitivism *and* non-cognitivism from the start. The second is that it is not clear how the non-cognitivist would want to argue for the truth of the enhanced version of P1. Thus, it is not clear whether the traditional defence of motivational internalism we have sketched above provides any basis for a claiming that we conceive of the motivational efficacy of moral judgements as the single result of one unitary motivational

---

<sup>12</sup> Caj Strandberg has recently argued that the non-cognitivist defender of the motivation argument needs to interpret motivational internalism in this way (Strandberg 2011: 346f.). Strandberg does not discuss the relevant differences between the stronger and the weaker version of the Humean theory of motivation in this context. Yet, his suggestion that the weaker, and not the stronger, version of the Humean theory provides the second premise for the motivation argument, seems challengeable and in need of independent justification – be it only because it is the stronger, and not the weaker, version that is part of Hume's original argument.

state. This understanding of the conceptual connection between moral judgement and motivation only seems compelling if we have already made up our mind that quite a specific version of cognitivism or non-cognitivism must be true. A version of the motivation argument that interprets the second premise in terms of the orthodox variant of the Humean theory of motivation, therefore, not only seems more faithful to the original argument. It also allows the non-cognitivist to start from a premise that does justice to the debate over cognitivism and non-cognitivism as a debate about the nature of moral judgements, because it does not presuppose much in the way of how we should conceive of this nature.

Finally, we need to take up the second question formulated at the beginning of this section, namely the question of whether we should still think of the revised first premise of the motivation argument, P1', as a statement of motivational internalism. Given that the literature on motivational internalism is dominated by the definition employed by Shafer-Landau, that is, P1, one might think that this question can only be answered in the negative. This is also suggested by the fact that those authors who take moral judgements to be potentially motivational – as opposed to actually motivating – do not seem to think of their view as an example of motivational internalism.<sup>13</sup>

Yet, such evidence notwithstanding, there are at least two reasons for viewing P1' as a statement of motivational internalism. The first reason, which has already been implicit in our discussion of the argument from social practice, is that even this weaker claim conceives of the connection between moral judgement and motivation as a conceptual connection. It can, therefore, provide an explanation for the fact that we question the sincerity of people who do not act in accordance with their moral utterances that is not available to motivational externalists. The second reason is that there is at least circumstantial evidence for thinking that what really lies at the heart of the previous internalism/externalism debate, even where this debate is not explicitly concerned with the motivation argument, is the question of whether moral judgements are potentially motivational.

The reason for wondering whether the debate was ever a debate about whether moral judgements are necessarily motivating in the first place is the striking fact that a considerable number of self-professed internalists defends versions of what is often referred to as “conditional internalism” (Björnsson et al. 2012: 126) or “restricted” (Miller 2008: 234) internalism. Conditional internalists, such as Michael Smith, Jamie Dreier or Simon Blackburn, explicitly concede that there may be cases in which the motivation adhering to

---

<sup>13</sup> This is not only true of Shafer-Landau, but also of Jonathan Dancy. Yet, while Shafer-Landau takes the view to be an example of externalism, Dancy sees it as “neither internalist nor externalist” and even claims that “there is a lot more about it that is internalist than there is externalist” (Dancy 1993: 25).

moral judgements is defeated – cases of depression or weakness of will, for example. Their analysis of moral motivation, therefore, is one that ties the motivation resulting from a moral judgement to further conditions, such as practical rationality or psychological normalcy (see Smith 1994; Dreier 1990; and 1998). The resulting position, then, amounts to the claim that, necessarily, if a person makes a moral judgement, *and* if she is practically rational/psychologically normal, then she is, at least to some extent, motivated to act in accordance with that judgement.

Now what is striking about these views is that they are forwarded as examples of *internalism*, and that this classification is accepted by many followers and critics. If motivational internalism is the claim that, necessarily, moral judgements motivate, then – one would think – any concession that moral judgements may fail to actually motivate should be equivalent to a denial of internalism and to an acknowledgement of the truth of externalism. In line with this, at least some externalists have suggested that the restrictions on internalism allowed by Smith, Dreier and Blackburn amount to nothing other than the victory of externalism.<sup>14</sup> Yet, Smith and Dreier do not seem shaken by such worries, and the most Blackburn is willing to acknowledge is that “externalists can win individual battles”, while “internalists win the war for all that” (Blackburn 1998: 61).

Now one way to explain this striking fact is to make use of Shafer-Landau’s distinction of *pro tanto* and *prima facie* motivation. Thus, one might think that what defenders of conditional internalism want to allow is that the motivation adhering to a moral judgement may not only be overridden by some other motivation, but also be quickly, and wholly, extinguished. According to this interpretation, even conditional internalists would want to deny that, on occasion, moral judgements may fail to give rise to any motivation whatsoever. However, there is no evidence that Smith, Dreier and Blackburn think of the possible defeat of moral motivation in this way. It seems, therefore, that another explanation is asked for, and one explanation that seems worth considering is that these authors, although accepting the ‘official’ definition of motivational internalism, are actually concerned not with the question of whether moral judgements are actually motivating, but with the question of whether moral judgements are potentially motivational. Once we think of motivational internalism in terms of the idea that moral judgements are capable of giving rise to motivation by themselves, without the help of any prior motivational states, it makes perfect sense to concede that moral judgements may not be motivationally efficacious at all, but to still think that internalism is

---

<sup>14</sup> See, for instance, Svavarsdóttir’s critique of Blackburn in Svavarsdóttir 2001: 22-24. Svavarsdóttir’s critique is taken up by Zangwill (see Zangwill 2009: 419ff.). See also Brink’s suggestion that the dispute about externalism and Smith’s version of conditional internalism may, at least partly, be just a dispute about words (Brink 1997: 7).

true. However, for those taking the ‘official’ definition of motivational internalism at face value and making the distinction between potentially motivational and actually motivating mental states, it makes just as perfect sense to see this concession as incompatible with the truth of internalism.

We may, then, end our discussion of the motivation argument and motivational internalism with the cautious and somewhat tentative conclusion that what lies at the heart of many discussions of motivational internalism may be the question of whether moral judgements are capable of giving rise to motivation by themselves, even where this is not explicitly acknowledged. This conclusion, of course, raises the question of why the stronger claim that moral judgements actually motivate has then been widely accepted as a definition of motivational internalism. It does not seem wholly impossible, however, to come up with an answer to that question. The reason may just be that the distinction between mental states that, by conceptual necessity, are potentially motivational, and mental states that, by conceptual necessity, are actually motivating has not sufficiently been addressed in the past or has not been taken to amount to much by the contributors to the debate. Shafer-Landau can take credit for having drawn attention to this distinction and its possible implications. Yet, instead of providing a refutation of motivational internalism as the first premise of the motivation argument, it seems that, in the end, Shafer-Landau offers the non-cognitivist defender of the motivation argument a way out – by pressing him to make the internalist claim more precise and to give up on the ‘official’ definition of motivational internalism. As already indicated, it may well be that the view that moral judgements are actually motivating remains an interesting metaethical question in its own right. Yet, it is not wholly clear what the theoretical import of this claim should consist in, unless one gives up on the orthodox version of the Humean theory of motivation, but still wants to defend a version of the motivation argument. As argued above, however, the non-cognitivist defender of this argument has good reasons to stick to the stronger version of the Humean theory of motivation, and as long as he does so, the claim that moral judgements are potentially motivational is all the internalism he is ever going to need.

### *Acknowledgements*

For helpful criticisms and suggestions, I am indebted to Wilfried Hinsch and an anonymous referee for *Philosophical Studies*. I would also like to express my gratitude to the *Deutsche Forschungsgemeinschaft* (German Research Foundation) for funding the project on expressivism and moral sentimentalism (HI 376/6-1) from which this paper has emerged.

### **References:**



- Barry, M. (2010). Humean theories of motivation. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics. Vol. 5* (pp. 195–223). Oxford: Oxford University Press.
- Björnsson, G., et al. (2012). Recent work in motivational internalism. *Analysis*, 72, 124–137.
- Blackburn, S. (1998). *Ruling passions*. Oxford: Clarendon Press.
- Brink, D. O. (1989). *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.
- Brink, D. O. (1997). Moral motivation. *Ethics*, 108, 4–32.
- Brown, C. (1988). Is Hume an internalist? *Journal of the History of Philosophy*, 26, 69–87.
- Cohon, R. (1997). Is Hume a noncognitivist in the motivation argument? *Philosophical Studies*, 85, 251–266.
- Dancy, J. (1993). *Moral reasons*. Oxford: Blackwell.
- Dreier, J. (1990). Internalism and speaker relativism. *Ethics*, 101, 6–26.
- Dreier, J. (2000). Dispositions and fetishes: Externalist models of moral motivation. *Philosophy and Phenomenological Research*, 61, 619–638.
- Hume, D. (1740/2000). *A Treatise of Human Nature*. Ed. by D. F. Norton and M. J. Norton. Oxford: Oxford University Press.
- Hume, D. (1751/1975). An enquiry concerning the principles of morals. In D. Hume, *Enquiries concerning human understanding and concerning the principle of morals*. Ed. by L. A. Selby-Bigge and P. H. Nidditch. Oxford: Clarendon Press.
- Joyce, R. (2009). Expressivism, Motivational Internalism, and Hume. In C. Pigden (Ed.), *Hume on motivation and virtue* (pp. 30–56). Basingstoke: Palgrave.
- Mele, A. (1996). Internalist moral cognitivism and listlessness. *Ethics*, 106, 727–753.
- Miller, C. (2008). Motivational internalism. *Philosophical Studies*, 139, 233–255.
- Pigden, C. (Ed.) (2009). *Hume on motivation and virtue*. Basingstoke: Palgrave.
- O’Leary-Hawthorne, J., & Price, H. (1996). How to stand up for non-cognitivists. *Australasian Journal of Philosophy*, 74, 275–92.
- Ridge, M. (2007). Ecumenical Expressivism: The best of both worlds? In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics. Vol. 2* (pp. 52–76). Oxford: Oxford University Press.
- Schroeder, M. (2010). *Noncognitivism in ethics*. London; New York: Routledge.
- Shafer-Landau, R. (2003). *Moral realism. A defence*. Oxford: Clarendon Press.
- Smith, M. (1994): *The moral problem*. Oxford: Blackwell.
- Smith, M. (2001). Some not-much-discussed problems for non-cognitivism in ethics. *Ratio*, 14, 93–114.
- Smith, M. (2009). The motivation argument for non-cognitivism. In: C. Pigden (Ed.), *Hume on motivation and virtue* (pp. 105–120). Basingstoke: Palgrave.
- Strandberg, C. (2011). The pragmatics of moral motivation. *Journal of Ethics*, 15, 341–369.
- Svavarsdóttir, S. (1999). Moral cognitivism and motivation. *The Philosophical Review*, 108, 161–219.
- Svavarsdóttir, S. (2001). On Blackburn’s *Ruling Passions*. *Philosophical Books*, 42, 18–26.
- Zangwill, N. (2009). Non-Cognitivism and motivation. In C. Sandis (Ed.), *New Essays on the Explanation of Action* (pp. 416–424). Basingstoke: Palgrave.