

THE ARGUMENT FROM MORAL PSYCHOLOGY

Abstract: *The argument from moral psychology is one of the strongest arguments that non-cognitivists use against cognitivism—the metaethical position according to which our moral judgements express beliefs. According to this argument, once we put together the Humean theory of motivation and motivational internalism, we yield the conclusion that cognitivism cannot represent the correct view about the semantic function of moral discourse. I will first attempt to show that a neurological syndrome, called pain asymbolia (a rare condition caused by lesions to the posterior insula that produces complete, thoroughgoing indifference to pain), can be used in order to formulate a persuasive defence of the Humean theory of motivation. After that, I will consider motivational internalism and argue that, without additional empirical evidence, neither internalism nor externalism can provide a plausible explanation of the connection between moral judgements and motivation. In conclusion, I intend to defend the claim that, until more persuasive arguments in favour of internalism are presented, non-cognitivists should not rely on the argument from moral psychology in their attempts to refute cognitivism.*¹

Keywords: *Moral psychology, pain asymbolia, the Humean theory of motivation, internalism.*

Introduction

Arguably, the debate between cognitivism and non-cognitivism plays a central role in contemporary metaethics. Both positions attempt to explain the nature of moral thought and the semantics of moral language. According to the most widely accepted account, cognitivists argue that moral sentences are capable of being true or false, because they express psychological states (e.g. beliefs or seemings) that are apt for assessment in terms of truth and falsity. Non-cognitivists typically deny this. They argue that moral sentences cannot be true or false, because they do not express truth-assessable psychological states, but rather some non-cognitive states such as emotions, desires, or sentiments of approval

1 I would like to express my sincere gratitude to the anonymous reviewers for their suggestions and constructive comments. Following their suggestions, I included several important improvements in the manuscript.

This work was supported by the project “Dynamic Systems in Nature and Society: Philosophical and Empirical Aspects” (no. 179041) financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

and disapproval. There are a number of serious arguments that are specifically designed to refute cognitivism. But perhaps one of the strongest arguments that non-cognitivists use against cognitivism is the so-called 'argument from moral psychology'. This is the argument in which the key premises are the subject matter of moral psychology, the field of metaethics that, broadly speaking, seeks to explain the nature of our moral motivation. While various versions of this argument can be found in the philosophical literature (e.g. Smith, 1994; Miller, 2003; Shafer-Landau, 2003), the main idea of the argument runs as follows: the combination of the predominant views in moral psychology, most of which take their inspiration in writings of the famous Scottish philosopher David Hume, seemingly entails the falsity of cognitivism.

Let us now have a closer look at the structure of the argument from moral psychology. Suppose that moral judgements express beliefs, as cognitivists claim. Now, it is a well-known fact that there is a reliable connection between our moral judgements and our motivation to act in the manner prescribed by these judgements. Many philosophers believe that the best explanation of this fact is to say that there is an internal and conceptual connection between moral judgements and motivation. This means that if an agent sincerely judges that, say, it is morally impermissible to eat meat, then, as a matter of conceptual necessity, she will be motivated to refrain from eating meat. This is motivational internalism. But what exactly does it mean to say that an agent is motivated to do something or to pursue a course of action? According to the Humean theory of motivation, our being motivated is always a matter of having a relevant desire and a means–end belief. Beliefs cannot by themselves motivate someone to act, and neither can desires. But Humeans argue that beliefs and desires can together constitute our motivation: the desire would tell us how the world should be, and the belief would tell us whether and how we might change the world in order to make it that way (Smith, 1994, p. 92; Miller, 2003, p. 219).

However, the enormity of the problem the cognitivist faces becomes clear when we combine motivational internalism with the Humean theory of motivation. Namely, it seems that if the semantic function of moral discourse were to be explained along cognitivist lines, the belief expressed by our moral judgement would have to be a belief that sustained an internal and necessary connection to a desire. That is to say, it would be a necessary or conceptual truth that an agent who possessed the belief would also possess the desire. Yet no belief is necessarily connected to a desire, because beliefs and desires are distinct existences and, according to Hume's dictum, it is impossible to have a necessary connection between distinct existences (Stoljar, 2010, p. 149). So there are no beliefs that sustain an internal and necessary connection to a desire. But then moral judgements cannot express beliefs. More formally, the argument can be put like this:

- 1) Suppose that moral judgements express beliefs.
- 2) It is an internal and necessary fact about an agent that if she sincerely judges that *X* is good, she is motivated to pursue the course of action *X*.

- 3) But being motivated to do something or to pursue a course of action is always a matter of having a belief and a desire.
- 4) So if a moral judgement expresses a belief, it would have to be a belief that sustained an internal and necessary connection to a desire.
- 5) But no belief is necessarily connected to a desire, because beliefs and desires are distinct existences, and it is impossible to have a necessary connection between distinct existences.
- 6) Therefore, it cannot be the case that moral judgements express beliefs.

The argument is valid: if all the premises are true, then the conclusion must be true. But we also need to know whether the premises really are true. The crucial premises of the argument are (2), (3), and (5), and I will deal with these premises in reverse order. So, I shall first attempt to show that a neurological syndrome, called *pain asymbolia*, can be used in order to defend the view that beliefs and desires are distinct existences. I will then argue that the claim according to which our motivation requires suitable belief-desire complexes is also supported by recent findings in the study of the neuroscience of pain. After showing how the premises (5) and (3) might get empirical support, I will consider motivational internalism, expressed by premise (2). I intend to show that if non-cognitivists decide to use the original, internalist version of this premise, the argument becomes unconvincing, because without additional empirical evidence motivational internalism cannot provide a plausible explanation of the connection between moral judgements and motivation. But if non-cognitivists decide to modify this premise along externalist lines, they end up with an invalid argument. I shall, therefore, attempt to defend the claim that even if it is possible to supply a line of reasoning that makes the premises (5) and (3) plausible, the non-cognitivist case against cognitivism is pretty shaky. It is worth noting, however, that in this paper I do not want to suggest that cognitivism is not threatened by other, more successful, arguments, or that we should ultimately accept the cognitivist interpretation of moral discourse. All I shall seek to establish is that, until more persuasive arguments in favour of internalism are presented, non-cognitivists should not rely on the argument from moral psychology in their attempts to refute cognitivism.

The Humean theory of motivation

Michael Smith defines the Humean theory of motivation as follows:

R at t constitutes a motivating reason of agent A to ϕ iff there is some ψ such that R at t consists of an appropriately related desire of A to ψ and a belief that were she to ϕ she would ψ , where the beliefs and desires in question are 'distinct existences'. (Smith, 1994, p. 92)

On this definition, the Humean theory of motivation can be viewed as the conjunction of the following two claims: (a) beliefs and desires are distinct existences, and (b) motivation requires suitable belief-desire complexes. Now,

both of these claims play an important role in the argument from moral psychology: the premise (5) says that no belief is necessarily connected to a desire, because beliefs and desires are distinct existences, and the premise (3) says that our motivation to do something or to pursue a course of action has its source in the presence of the relevant desire and means-end belief. In the following two sections I intend to show that these claims are supported by recent findings in the neuroscience of pain. In that sense, my defence of the Humean theory of motivation will be shaped by empirical facts.

Premise (5): Beliefs and desires are distinct existences

Let us first focus on the Humean claim according to which beliefs and desires are conceived of as metaphysically-distinct states of affairs. As Michael Smith notices, for any imagined pair of metaphysically-distinct states of affairs it is always possible to pull that pair apart, at least modally (1994, p. 7). So, if *x* and *y* are metaphysically-distinct states of affairs that typically occur together, then it would have to be possible to pull *x* apart from *y*. That is to say, it would have to be possible to show that on some occasion *x* occurred, while *y* did not occur, or vice versa. In this sense, it is not very difficult to show that beliefs and desires are, in fact, metaphysically-distinct states of affairs.

According to analytical functionalism, beliefs and desires can be established as distinct states of affairs because they have different and separable dispositional or functional roles (Jackson and Pettit, 1988; Smith, 1994, pp. 111–116). For beliefs, the functional role is popularly explained by saying that beliefs have a world-mind direction of fit—the contents of one's beliefs change depending on how one takes the world to be. By contrast, desires have a mind-world direction of fit—the content of one's desires affect behaviours to make the world conform to them. Mark Platts summarizes the idea of the different direction of fit in the following way:

Beliefs aim at the true, and their being true is their fitting the world; falsity is a decisive failing in a belief, and false beliefs should be discarded; beliefs should be changed to fit with the world, not vice versa. Desires aim at realisation, and their realisation is the world fitting with them; the fact that the indicative content of a desire is not realised in the world is not yet a failing in the desire, and not yet any reason to discard the desire; the world, crudely, should be changed to fit with our desires, not vice versa. (Platts, 1979, pp. 256–7)

Beliefs and desires also differ with regard to their truth-aptness. Since beliefs are psychological states that purport to represent how the world is, they are assessable in terms of truth and falsity. Whether our beliefs will turn out to be true or false depends on whether or not the world actually is as they represent it. On the other hand, since desires do not have the same representative function as beliefs and do not purport to represent the way the world is, they are not assessable in terms of truth and falsity. As a result of these differences, we can always pull beliefs and desires apart, in the sense that “for any belief and desire

pair that we imagine, we can always imagine someone having the desire but lacking the belief, and vice versa” (Smith, 1994, p. 7). And, based on the possibility of always pulling any imagined belief and desire pair apart, it seems that we can successfully defend the claim that beliefs and desires are metaphysically-distinct existences and that we can, accordingly, accept premise (5) of the argument from moral psychology.

Some philosophers will, however, find this conclusion hard to swallow. It is possible to object that there are evident cases in which a belief cannot be pulled apart from a desire. For instance, take the case of our beliefs about pain. Until recent neurological investigations, there was a widely accepted opinion that pain is not merely the response to noxious stimuli or disease but is, by definition, an unpleasant experience (Baier, 1958, p. 26; IASP, 1986). According to this opinion, the fact that we dislike pains is not a contingent fact—there is a *necessary* connection between the belief that a person feels pain and her desire to avoid that pain. But then, the objection continues, the claim that beliefs and desires are metaphysically-distinct existences is false. And, if this is so, premise (5) of the argument from moral psychology must be rejected.

Yet it is not clear at all that this objection is compelling. The neurological syndrome, called pain asymbolia, shows that the objection does not actually succeed in disproving the possibility of pulling apart our beliefs about pain from our desires to diminish pain. Let us get a little clearer on how such a neurological syndrome is possible. In contrast to the popular understanding, a number of philosophers and scientists endorse the view according to which pain is a composite mental state (e.g. Dennett, 1978; Hardcastle, 1997; Price, 2000; Grahek, 2007). On this view, a typical human pain experience comprises at least two distinct components: the sensory/discriminative component and the affective/motivational component. These components are normally linked together, but they can become disconnected.

Pain asymbolia is a neurological syndrome characterized by the complete dissociation of the sensory component of pain from its affective component. In this sense, pain asymbolia is often regarded as “the only clear-cut case in which severe pain is not experienced as unpleasant, and in which there are no traces of any other aversive attitude toward it” (Grahek, 2007, p. 3). This claim is supported by the fact that patients with this syndrome display striking behavioural reactions to pain stimulation. Pain asymbolics fail to respond with appropriate motor and emotional reactions to painful stimuli applied anywhere on their bodies. In addition, they typically show significantly greater values for pain tolerance and pain endurance (Berthier, Starkstein, and Leiguarda, 1988, p. 47). Finally, asymbolics not only fail to display normal reactions to painful stimuli, but actually behave in the exact opposite way: they typically smile or laugh during the pain testing procedure (Ramachandran, 1998, p. 1857). In the first reported case of pain asymbolia, Schilder and Stengel note that:

The patient displays a striking behaviour in the presence of pain. She reacts either not at all or insufficiently to being pricked, struck with hard objects, and pinched. She never pulls her arm back energetically or with strength. [...] Pricked

on the right palm, the patient smiles joyfully, winces a little, and then says, 'Oh, pain, that hurts.' She laughs, and reaches the hand further toward the investigator and turns it to expose all sides. (Schilder and Stengel 1928, p. 147)

When it comes to understanding the relevant neural systems that appear to be implicated in the syndrome, the insular cortex was invariably damaged in all patients diagnosed with pain asymbolia. Neurological findings suggest that insular damage may play a critical role in the development of the syndrome by interrupting connections between sensory cortices and the limbic system (Berthier et al. 1988). Thus, pain asymbolia might be described as a sensory-limbic disconnection syndrome (Grahek, 2007, p. 52). In fact, in his famous article on disconnection syndromes in animals and humans, Norman Geschwind proposes the sensory-limbic disconnection hypothesis as an explanation of pain asymbolia. Geschwind suggests that since the somatosensory cortical areas responsible for the detection of sensory features of noxious stimuli are spared in pain asymbolics, they are able to discern the modality, qualities, intensity, and location of these stimuli. On the other hand, the damage to the insula and parietal operculum may disrupt the connections between sensory and limbic structures, impairing their ability to attach appropriate emotional significance to painful stimuli (Geschwind, 1965, p. 270).

Regardless of its great importance to behavioural neurology, however, Geschwind's hypothesis is far from being uncontroversial. It has been criticized for assuming an entirely simplistic, feed-forward picture of brain function, which has fallen out of favour (Catani et al. 2005, p. 2233; Klein, 2015, p. 500). But even if we agree with this criticism and decide to jettison Geschwind's hypothesis, the claim that asymbolics can feel pain, while being completely unmotivated by it does seem to be correct. As a matter of fact, neurological and psychological testing of pain asymbolics showed conclusively that their ability to recognize pain upon noxious stimulation is fully preserved, but that they consistently fail to display any affective or motor responses to painful stimuli (Grahek, 2007, p. 43). Based on their ability to fully recognize pain upon noxious stimulation, pain asymbolics are able to form true beliefs about the pain they feel, but they do not have any desire whatsoever to diminish that pain. Since there is no necessary or conceptual connection between a person's *believing* that she feels pain and her *desiring* to diminish the pain that she feels, it follows that the claim according to which beliefs and desires are distinct existences is not decisively refuted by the objection. So, at least until we are presented with some more persuasive objections to this claim, I think we should concede premise (5) of the argument from moral psychology.

Premise (3): Motivation requires suitable belief-desire complexes

According to the premise (3), being motivated to do something or to pursue a course of action is always a matter of having a belief and a desire. In other words, the relevant desire and means-end belief are individually necessary and jointly sufficient conditions for a state's constituting someone's motivating reason

(Smith, 1994, p. 94). Given the already outlined difference between beliefs and desires in terms of direction of fit, this view certainly has some initial plausibility. That is, beliefs cannot on their own motivate someone to act, because although they tell us how the world is and how it could be changed, they do not tell us how the world is to be changed. On the other hand, desires cannot on their own motivate someone to act, because they tell us how the world should be changed, but they don't tell us how the world actually is and, more importantly, they don't tell us how the world has to be changed in order to be the way they tell us that it should be (Miller, 2003, p. 218). But, according to the Humean view, beliefs and desires can together constitute our motivation to act: the desire would tell us how the world should be and the belief would tell us whether and how we might change the world in order to make it that way.

Not all philosophers, however, accept the Humean view about motivation. According to anti-Humeanism, a person may be motivated to act morally simply on the basis of her moral beliefs. In this sense, moral beliefs can and do motivate people to act even in the absence of desires or any other affective states. Despite the fact that a number of philosophers advocate anti-Humeanism (e.g. Nagel, 1970; McDowell, 1978; McNaughton, 1988; Shafer-Landau, 2003), I think there are persuasive reasons to doubt that this theory represents a correct view about moral motivation. Recall that the examination of pain asymbolia revealed that pain asymbolics are able to form true beliefs about the pain they feel, but do not have any desire whatsoever to diminish that pain. Because of the complete dissociation of the sensory dimension of pain from its affective component, pain asymbolics typically form beliefs about pain that are deprived of all accompanied affective states. It turns out, however, that these beliefs are unable to motivate: pain asymbolics are unmoved by the painful stimulation, and they remain in a state of complete indifference and inaction (Grahek, 2007, p. 48). Thus, in contrast to anti-Humeanism, results from the examination of pain asymbolia strongly suggest that beliefs, or even more widely, cognitive states that are deprived of all accompanied affective states are motivationally inert.

This suggestion is in accordance with the Humean view. Humeans claim that beliefs are, on their own, insufficient for motivation, and that motivation always requires, in addition to a belief, the presence of a desire or some other affective state (Shaw, 1989, p. 163). The Humean claim that motivation requires suitable belief-desire complexes is thus supported by recent findings in the study of the neuroscience of pain. Based on that, I think we can conclude that there are good reasons for accepting premise (3) of the argument from moral psychology. So it seems that everything hinges on whether motivational internalism, expressed by premise (2), represents the correct account of the connection between moral judgements and motivation. It is to this question that we turn our attention in the section that follows.

Premise (2): Motivational internalism

Motivational internalists attempt to provide an explanation of the reliable connection between the formation of a moral judgement and the motivation to act as that judgement prescribes. More precisely, internalists argue that there is an internal (i.e. necessary, conceptual) connection between our moral

judgements and the will. This means that a person who judges that a type of action is morally right but consistently claims that she sees no reason to perform actions of that type obviously doesn't grasp the concept of moral rightness; she betrays some sort of conceptual confusion. Motivational internalism is not the only view about the nature of the connection between our moral judgements and motivation—there is also motivational externalism.

Motivational externalists tell us that moral judgements have no necessary or conceptual connection with motivation; rather, the connection is contingent and external. In that sense, a person who judges that a type of action is morally right but consistently claims that she sees no reason to perform actions of that type betrays no conceptual confusion. Suppose, for instance, that John is a good and strong-willed person, and that he judges that it is right to refrain from eating meat. According to externalists, good and strong-willed people must have as their primary source of moral motivation a desire to do the right thing (Smith, 1996, p. 182). So externalists would attribute to John the non-derivative desire to do what he believes to be right, and it is precisely from this non-derivative desire and the moral belief that it is right to refrain from eating meat that John acquires his motivation to refrain from eating meat. On the externalist view, an agent's motivation to act in accordance with her moral judgements is *derived* both from that agent's belief that something is morally right and her non-derivative desire to do what she believes to be right (Miller, 2003, p. 225). An agent's moral motivation is, thus, essentially derivative; it does not follow directly from the content of her moral judgements, as internalists maintain.

The important thing to bear in mind is that 'internalism' is an ambiguous label: it is used to refer to several different views about the connection between moral judgements on the one hand, and being motivated on the other (Björklund et al. 2012). Indeed, when it comes to understanding the nature of the conceptual connection that holds between our moral judgements and motivation, it is possible to distinguish two main versions of internalism, strong (or direct, unconditional) and weak (or indirect, conditional). Let us consider these two versions in turn.

Strong internalism is the view according to which if an agent judges that it is right for her to ϕ in circumstances C , then, as a matter of conceptual necessity, she will be motivated to ϕ in C (Smith, 1994, p. 61). This is a very strong claim; perhaps even too strong. As externalists often object, strong internalism is such a strong claim that it commits us either to denying that there are cases of practical irrationality (cases like the weakness of will, apathy, despair, etc.), or to accepting that there are such cases but that they may not defeat an agent's moral motivation while leaving intact her appreciation of moral reasons. But both of these two commitments of strong internalism seem to be unacceptable. The first commitment is unacceptable because various cases of practical irrationality (i.e. cases in which agents fail to be motivated by their own moral judgements) are scientifically established. The second is unacceptable because it is quite conceivable that a person may judge it morally right to ϕ in C , but that some form of practical irrationality may frustrate the connection between her

moral judgement and motivation (Smith 1994, p. 61). Based on this, externalists conclude that strong internalism provides us with an implausible explanation of moral motivation; the reliable connection between moral judgements and motivation is not correctly explained by the strong internalist thesis. But is this externalist objection convincing? Let us see if we can answer this question by considering the paradigmatic case of practical irrationality—the phenomenon the philosophical tradition calls “the weakness of will”. There are four ways in which this phenomenon might be understood:

- a) Weakness of the practical will with respect to action: I judge that *x* is the best all-things-considered option available to me (taking into account not just moral concerns but also selfish, prudential concerns) and yet I fail to do *x*.
- b) Weakness of the practical will with respect to motivation: I judge that *x* is the best all-things-considered option available to me (taking into account not just moral concerns but also selfish, prudential concerns) and yet I am not motivated to do *x*.
- c) Weakness of the moral will with respect to action: I judge that *x* is the most *moral* or *right* option available to me, and yet I fail to do *x*.
- d) Weakness of the moral will with respect to motivation: I judge that *x* is the most *moral* or *right* option available to me, and yet I fail to be motivated by *x*.

Since (A) and (C) are views about the connection between an agent’s judgements and her *actions*, it seems they are irrelevant to whether any version of *motivational* internalism is true. Essentially, (A) and (C) say that there is no necessary connection between an agent’s judgements and her actions. Motivational internalists can agree, without damaging their position: this view does not entail the existence of the necessary connection between an agent’s judgements and actions. To say that an agent is motivated to pursue some course of action is to say that she has at least some tendency or temptation to pursue that course of action. But this tendency or temptation does not have to be overriding. Thus, someone can be motivated to *y* but never get around to doing anything about it. In addition, an agent can wilfully do *y*, while still having *some* motivation to do *x*. The fact that she did *y* does not tell us anything about whether she was or was not motivated to do *x*. Therefore, (A) and (C) cannot be used to undermine strong motivational internalism. This leaves us with (B) and (D).

With respect to (B), it is true that the internalist must deny that this kind of weakness of the will is a genuine phenomenon. At first glance, that does not seem to be so hard to do. Try to imagine, for instance, the person who really does judge that her absolute *best* (not necessarily best *moral*) course of action is *x* and yet remains entirely unmotivated to do *x*. Admittedly, it is very hard to imagine such a person. It almost seems tempting to say that the denial of the existence of any such person comes very close to being a tautology. However, I am sceptical about whether the internalist can actually use this line of reasoning in order to show that (B) is not a genuine phenomenon. In particular, it is not especially

difficult to show that the internalist denial of (B) is not at all a tautology. In fact, it is a significant empirical claim that may turn out to be at odds with the scientific understanding of the connection that holds between moral judgements and motivation.

Namely, though important empirical advances have been made in attempts to understand the nature of our moral judgements, we are currently limited by a lack of information about what particular neurological structures are implicated in making a moral judgement. It remains an open possibility that moral judgements arise from the coordinated interaction of *several* neurological structures. It is thus possible that there could be various dissociations between our sincere making of moral judgements and our motivation to act in accordance with them; e.g. there could be people who judge some course of action to be best available to them and who are, because of some neurological disorder, completely unmotivated to bring that action about. The externalist seems better placed to capture the sense in which the behaviour of such people is intelligible. In addition, the externalist could full-heartedly accept the internalist point that it is very difficult to imagine such people. But it is not more difficult to imagine such people than to imagine people who feel pain while not being motivated to diminish it. And, just as the latter difficulty does not lead us to deny the existence of pain asymbolics, the former should not lead us to deny that (B) could represent the genuine phenomenon, as internalists maintain. Still, it would be unjustified to conclude that the internalism/externalism debate is resolved in favour of externalism. For this conclusion to be justified, the externalist would have to deliver up at least one clear-cut example of the dissociation between moral judgements and motivation. The present clinical literature, as will be discussed below, provides no evidence that there is such an example.

Finally, let us consider the fourth interpretation of the weakness of will—the interpretation (D). The main problem with this interpretation is that it is nothing more than a statement of externalism. Evidently, to say that an agent could judge that *x* is the most moral or right option available to her and fail to be motivated by *x* is to say that moral judgements have no necessary or conceptual connection with motivation. So if the externalist simply claims that (D) is a genuine phenomenon, she thereby begs the question against the internalist. The most straightforward way for the externalist to show that (D) is a genuine phenomenon would consist in showing that there are genuine cases of people who sincerely judge that a course of action is right, but who are completely unmotivated to perform that action. It is very important to bear in mind, however, that there is no well-documented case of the dissociation between moral judgements and motivation. The recent clinical literature provides no support for externalist claims. Rather, it consistently supports an association between *deficient* moral behaviour and *deficient* moral understanding (Kennett & Fine, 2008). Thus as things stand the externalist is in no better shape than the internalist.

Based on the analysis of the four possible interpretations of the weakness of will, the conclusion seems to be that, without additional empirical evidence, neither strong internalism nor externalism can provide a satisfactory explanation

of the connection between moral judgements and motivation. But without such empirical evidence, the non-cognitivist is not justified in using the strong internalist thesis in her argument from moral psychology.

I shall now consider if the weak internalist thesis can fare better than the strong thesis. Weak internalism is the view that although there is a conceptual connection between moral judgement and motivation, the connection involved is the following defeasible one: if an agent judges that it is right for her to ϕ in circumstances C , then either she is motivated to ϕ in C or she is practically irrational (Smith, 1994, p. 61). Thus, according to this disjunctive version of internalism, if a person judges that it is morally right to refrain from eating meat, then as a matter of conceptual necessity she will be motivated to refrain from eating meat, just as long as she is not suffering from some form of practical irrationality, such as weakness of will and the like. However, the weak internalist thesis is by no means without problems.

The most serious problem has to do with the objection, expounded by Alexander Miller, that the thesis faces a very real danger of lapsing into triviality (2003, p. 221). The objection is complex, but here I will simplify in order to convey the general idea. It runs as follows. Weak internalists require some substantial characterization of 'practical rationality'. But in order to provide such a characterization, it is not enough to say that practical rationality involves freedom from "weakness of will and the like", as does, for instance, Michael Smith (1994, p. 120). In Miller's view, the phrase 'and the like' seems to amount to some condition along the lines of "and of any other condition which is such as to frustrate the connection between moral judgement and motivation" (2003, p. 221). And, to be sure, if the condition is interpreted in this way, the weak internalist thesis comes out as trivial. Yet it turns out to be exceedingly difficult to come up with any other non-trivial characterization of practical rationality. Miller concludes that unless much more is said, we cannot be sure that the internalist position is so much as formulable, let alone capable of representing an appropriate account of the connection between moral judgements and motivation.

Miller's argument strikes me as unconvincing. Note that to say that it is exceedingly difficult to come up with some non-trivial characterization of practical rationality is not to say that it is *impossible* to come up with such a characterization. However, if the externalist wants to argue that the weak internalist thesis is false, this is exactly what she would have to show. So, Miller's objection is not a very good objection after all, because it does not prove that weak internalism is in any way conceptually defective or, for that matter, incapable of being formulated in a satisfactory way. The only direct implication of this objection is that the weak internalist needs to come up with some substantial characterization of practical irrationality. The objection thus leaves open the possibility that weak internalists may fulfil this task and end up with a plausible account of moral motivation.

Nevertheless, there does seem to be something undeniable in this objection. Note that without some substantial characterization of practical irrationality the weak internalist thesis comes out as something along these lines: If an agent

judges that it is right for her to ϕ in circumstances C and she is free of *any condition* which is such as to frustrate the connection between moral judgement and motivation, then, as a matter of conceptual necessity, she is motivated to ϕ in C . This is still a version of internalism, because the connection between moral judgements and motivation is conceived of as conceptual. Yet, if *any condition* whatsoever could make an exception to this conceptual connection, then weak internalism comes dangerously close to externalism. In fact, weak internalism would be extensionally equivalent to externalism. But then the more natural option would be to accept externalism and modify the premise (2) of the argument from moral psychology along externalist lines. This modification, however, would not help the non-cognitivist to improve her argument, since the externalist version of the argument from moral psychology is even more problematic than the internalist version. To make this claim as convincing as possible, let us consider the modified externalist version of the argument:

- 1) Suppose that moral judgements express beliefs.
- 2) There is a connection between a moral judgement and motivation, but the connection is altogether *external* and *contingent* (externalism).
- 3) Being motivated to do something or to pursue a course of action is always a matter of having a belief and a desire.
- 4) So, if a moral judgement expresses a belief, then for an agent to be motivated there has to be an appropriate desire (e.g. a non-derivative desire to do what she judges to be right).
- 5) But no belief is necessarily connected to any desire, because beliefs and desires are distinct existences, and there can be no necessary connection between distinct existences.
- 6) Therefore, it cannot be the case that moral judgements express beliefs.

Clearly the argument is invalid: the conclusion does not follow from the premises. All we know from the premises is that there is an external and contingent relation between moral judgements and motivation, that necessary and sufficient conditions for being motivated consist in a relevant desire and an appropriate means–end belief and, finally, that no belief is necessarily connected to a desire. There is nothing in the premises that implies that moral judgements cannot express beliefs. Therefore, the combination of the premises used in the modified version of the argument from moral psychology does not entail the conclusion.

Where does this result leave us? We have seen that it is possible to supply a line of reasoning that makes premises (5) and (3) plausible. With regard to premise (2), however, the case is much more complex. The debate between internalism and externalism is very much a live one, and much work remains to be done before we can make an informed choice between them. It may be that the advocates of these two positions will eventually come up with a plausible explanation of moral motivation, but the important point is that both the philosophical arguments and the science relevant to this issue are

still inconclusive. By using the argument from moral psychology, however, non-cognitivist presuppose a particular – internalist – resolution of a highly controversial philosophical debate which is nowhere near to being resolved. This should not lead us to conclude that the best account of moral judgement is cognitivist. Cognitivism could be seriously threatened by other non-cognitivist arguments. But, I think we are entitled to conclude that, at least until more persuasive arguments in favour of internalism are presented, non-cognitivists should not rely on the argument from moral psychology in their attempts to refute cognitivism.

References

- Baier, K. (1958). *The Moral Point of View*. Ithaca, Cornell University Press.
- Berthier, M., Starkstein, S., & Leiguarda R. (1988). "Asymbolia for Pain: a Sensory-Limbic Disconnection Syndrome." *Annals of Neurology*, 24, pp. 41–49.
- Björklund, F., Björnsson, G., Eriksson, J., Olinder, R. F., and Strandberg, C. (2012). "Recent Work on Motivational Internalism." *Analysis*, 72 (1), pp. 124–137.
- Catani, M., and Dominic H. ffytche (2005). "The Rises and Falls of Disconnection Syndromes." *Brain*, 28(10), pp. 2224–39.
- Dennett, D. (1978). "Why You Can't Make a Computer that Feels Pain." *Synthese*, 38, pp. 415–56.
- Geschwind, N. (1965). "Disconnexion Syndromes in Animals and Man." Part I. *Brain*, 88, pp. 237–294.
- Grahek, N. (2007). *Feeling Pain and Being in Pain*. Cambridge, Mass: MIT Press.
- Hardcastle, V. G. (1997). "When a Pain is Not." *The Journal of Philosophy*, 94, pp. 381–409.
- IASP (1986). "Pain Terms: a List with Definitions and Notes on Pain." *Pain*, 3 (Suppl), pp. 216–21.
- Jackson, F. and Pettit, P. (1988). "Functionalism and Broad Content." *Mind*, pp. 381–400.
- Kennett J. & Fine C. (2008). "Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths.'" In Sinnott-Armstrong, W. (ed.) *Moral Psychology: The Neuroscience of Morality: Emotions, Brain Disorders, and Development*. MIT Press.
- Klein C. (2007). "An Imperative Theory of Pain." *Journal of Philosophy*, 104(10), pp. 517–32.
- Klein C. (2015). "What Pain Asymbolia Really Shows." *Mind*, pp. 493–516.
- McDowell, J. (1978). "Are Moral Requirements Hypothetical Imperatives?" *Proceedings to the Aristotelian Society*, 52, pp. 13–29.
- McNaughton, D. (1988). *Moral Vision*. Oxford: Blackwell.
- Miller, A. (2003). *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press.

- Nagel, T. (1970). *The Possibility of Altruism*. Princeton University Press.
- Platts, M. (1979). *Ways of Meaning*. London: Routledge & Kegan Paul.
- Price, D. (2000). "Psychological and Neural Mechanisms of the Affective Dimension of Pain." *Science*, 288, pp. 1769–72.
- Ramachandran, V. S. (1998). "Consciousness and Body Image: Lessons from Phantom Limbs, Capgras Syndrome, and Pain Asymbolia." *Philosophical Transactions of the Royal Society of London—Series B: Biological Sciences*, 353, p. 1851–1859.
- Schilder, P., and Stengel, E. (1928). "Schmerzasympolie." *Zeitschrift für die gesamte Neurologie und Psychiatrie*, 113, p. 143–148.
- Shafer-Landau, R. (2003). *Moral Realism: a Defence*. Clarendon Press: Oxford.
- Shaw, D. (1989). "Hume's Theory of Motivation." *Hume Studies*, 15, pp. 163–83.
- Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.
- Smith, M. (1996). "The Argument for Internalism: Reply to Miller." *Analysis*, 56, pp. 175–83.
- Stoljar, D. (2010). *Physicalism*. New York: Routledge.