Normative Judgments, Motivation, and Evolution

Suikkanen, Jussi

Normative Judgments, Motivation, and Evolution

JUSSI SUIKKANEN

Final Author Copy: To be Published in *Filosofiska Notiser* (http://filosofiskanotiser.com/filosofiskanotiser.htm), special issue on Metaethics.

ABSTRACT: This paper first outlines a new taxonomy of different views concerning the relationship between normative judgments and motivation. In this taxonomy, according to the Type A views, a positive normative judgment concerning an action consists at least in part of motivation to do that action. According to the Type B views, motivation is never a constituent of a positive normative judgment even if such judgments have, due to the kind of states they are, a causal power to produce motivation in an agent. Finally, according to the Type C views, a normative judgment can produce motivation only with the help of a third mental state or a distinct substantial local disposition. This paper then outlines a novel evolutionary argument for the Type B views. If we assume that normative judgments' ability to shape our motivations enabled efficient planning and co-operation, the psychological mechanism responsible for this adaptation should be understood as a proximal mechanism. This paper argues that it is then more likely that we evolved to make normative judgments that have direct causal powers to influence our motivations because such Type B mechanisms are more reliable than the Type C mechanisms. It also suggests that the Type A views are either empirically false or collapse into the Type B views.

Key words: normative judgments; moral psychology; motivation; evolution; motivational internalism

1. Introduction

Human beings can form and execute complicated and often changing plans and co-operate in many sophisticated and creative ways. These abilities were an evolutionary advantage as they were needed in hunting, farming, and many other important human activities, and so the ability to carry out these activities successfully helped us to survive as a species.[1] In order to plan and co-operate effectively, we needed concepts that enabled us to shape our motivations via the judgments in which we employed these concepts. They included various normative concepts that

[1] For a recent account of the selective benefits of our ability to co-operate and the kind of social attitudes it requires, see Sterelny (2013) and Kitcher (2006, §3). For selective benefits of planning, see Gibbard (2003, ch. 13).

are today conventionally expressed in English with 'ought,' 'should,' 'must,' 'being a reason for,' 'good,' and the like.[2]

Consider a situation in which you have multiple needs. You want to sleep, drink, eat, and get to safety, but you cannot satisfy all these needs simultaneously. It helps here to consider what you should do first, what second, and so on. By using normative concepts in deliberation in this way, you can create a coherent plan that enables you to satisfy all your needs in an efficient sequence. It also helps if you are *motivated* to follow the plan that you judged to be best as this makes you actually follow the plan to satisfy all your needs. If the normative concepts you employed in planning shape what you are motivated to do, your motivations will conform to your normative judgments, which makes you an efficient agent.[3]

Likewise, consider co-operation. Imagine that a certain way of hunting provides us with food most reliably. It involves first locating animals, then chasing them to a specific location, and then finally releasing the traps. Because locating the animals, however, takes a long time and chasing is hard work whereas releasing the traps is easy, we might all want to be trap-releasers but that would not work. By relying on normative concepts, we can negotiate together what we ought to do – how we should rotate the roles or choose the most suitable individuals for them. If these concepts again shape our motivations through the judgments that employ them, after we come to an agreement we will all have the corresponding motivations that lead to effective co-operation.

For these reasons, I will assume in this paper that our ability to use normative concepts in normative judgments is *an adaptation*.[4] The use of these concepts in normative judgments arguably shapes our motivational states at least in planning and co-operation contexts and presumably in many other contexts too, and this makes us better planners and co-operators and more effective agents generally as well. Because these abilities provided selective benefits and having normative concepts that could shape our motivations in the relevant judgments gave us those abilities, there is some reason to assume that having that kind of concepts was an adaptation.

There is, however, a fundamental metaethical disagreement over *how* exactly our normative judgments are connected to motivation. How do the normative concepts *shape* our motivational states when we employ them to make normative judgments? This controversy is usually formulated in terms of motivational internalism and externalism.[5] Roughly, the internalists believe that, necessarily, if an agent makes a genuine normative judgment, she is motivated to act accordingly. The externalists, in contrast, reject the previous internal, modal connection between the normative judgments and motivation. They think that you can make a genuine normative judgment and yet not be motivated at all to act accordingly.

---

[2] I focus only on general normative concepts and not on moral concepts. For discussions of whether moral cognition is an adaptation, see, for example, Hauser (2006), Joyce (2006), and Kitcher (2006).

[3] For the selective benefits of normative guidance, see Gibbard (1990) and Kitcher (2006), who also used this idea as an argument for expressivism. Joyce (2006, 175–6) argues that, even if normative guidance is an adaptation, that does not fix whether that guidance is non-cognitive or cognitive in nature. This fits my conclusion below: what we can learn about normative guidance based on evolutionary considerations is neutral between expressivism and cognitivist views (§3 below).

[4] This assumption is shared by Schroeter (2005, §4).

[5] For an overview of the recent contributions, see Björklund et al (2012). For an outline of the historical origins of this debate, see Kauppinen (2007, 111–2).

There are, however, two reasons not to focus on the internalism versus externalism debate in the present context. Firstly, both 'internalism' and 'externalism' are labels for very broad families of views. For example, when we formulate internalism, different theoretical choices lead to different forms of internalism. We must first decide between strong and weak internalism. The former requires that, necessarily, when you judge that you ought to φ, you have *overriding* motivation to φ whereas the latter requires you to have only *some* motivation.[6] Secondly, unconditional internalism requires that, necessarily, *whenever* you judge that you ought to φ, you are motivated to φ, whereas conditional internalism claims only that, necessarily, *either* you are motivated to φ *or* you are, for example, psychological abnormal or practically irrational.[7] Finally, direct forms of internalism require that any agent who makes a normative judgment must be motivated *by that judgment* whereas deferred forms of internalism grant that it is enough if she has been motivated by her previous judgments or if the other members of her community are motivated by their judgments.[8]

Similarly, most externalists accept that normative judgments are reliably connected to motivation, because ordinary agents usually do what they think they should. Yet, externalists rely on different psychological processes to explain that connection. They have made use of, for example, (i) *de dicto* desires to do whatever you should do, (ii) prior substantive *de re* desires that have either a cultural or a biological origin, (iii) local motivational dispositions that connect normative judgments to motivation, (iv) higher-order desires to have desires that match one's normative judgments, (v) virtues, and so on.[9] There are thus as many versions of externalism as there are of internalism.

Because of this, it is difficult to see how any argument could vindicate either internalism or externalism generally. If the externalists, for example, rely on their intuitions concerning which agents make genuine normative judgments, *some* forms of internalism will be likely to fit those intuitions.[10] Likewise, if the internalists argue that externalism makes ordinary agents normative fetishists, some versions of externalism will probably not have that consequence.[11] It is thus difficult to make progress at the general level in this debate. We could, alternatively, investigate every position separately, but, given the number of positions, this approach too quickly becomes unmanageable.

The second reason why formulating the debate in terms of internalism and externalism is not helpful is that this distinction focuses on the modality of the connection between normative judgments and motivation – on how these states co-vary across worlds. However, different psychological processes, some more internal and others more external, could in principle support

---

[6] The strong form of internalism is rarely defended (but see Stevenson (1937, 16 and 19)). For a list of discussions of weak internalism, see Miller (2008, 234 fn. 3).

[7] For defenders of conditional and unconditional internalism, see Björklund et al (2012, 126–8).

[8] For defenders, see Björklund et al (2012, 128–9).

[9] (i) was introduced as an externalist target by Smith (1994, 71–6) and then defended by Lillehammer (1997) and Shafer-Landau (1998, 357 and 2003, 159). For (ii), see Svavarsdottir (1999, 198–9) and Shafer-Landau (1998, 356 and 2003, 158–60); for (iii), see Copp (1997, 50) and Dreier (2000, 623–9); for (iv), see Dreier (2000, 629–38); and for (v), see Cuneo (1999).

[10] For externalist arguments based on intuitions about cases, see Miller (2008, 235–236 fns. 8–10). For a version of internalism compatible with these counterexamples, see, for example, Björnsson (2002).

[11] For this objection to externalism, see Smith (1994, 74–75). For an externalist response, see Dreier (2000, 634–638).

a given degree of modal co-variation and what we want to know is not the degree of co-variation but rather the psychological process responsible for it.[12]

For these reasons, it is better to classify the different views based on the psychological processes they rely on in explaining the connection between normative judgments and motivation. §2 therefore proposes that different views should be understood as different versions of what I call Type A, Type B, and Type C views corresponding to different types of psychological mechanisms.

The rest of this paper then outlines a new kind of a defeasible argument for the Type B views. §3 argues that evolutionary considerations support these views more than the Type C views. Both views are competing solutions to the same *design problem*. In evolutionary biology, general principles are then used to predict which of the multitude of possibilities in this kind of situations evolve. These principles focus on the *availability*, *reliability*, and *efficiency* of the proximate mechanisms (Sober 1994 and 1999; Sober and Wilson 1998). §3 uses these principles to argue that it is more likely that we evolved to have a Type B rather than a Type C system.

§4 first explains how we might think that, for similar reasons, we evolved to have a Type A system. It then argues that some Type A views rely on psychological states that were not available as ancestral variants and others are either empirically untenable or collapse into Type B views. I will thus conclude that evolutionary psychology supports the idea that we evolved to have a Type B system – normative concepts and judgments that do not themselves consist of motivational states but ones that can causally produce such states in us without the help of any other states.

2. What Kind of a Process?

2.1 Type A Views

Natural languages contain different normative expressions. In English, they include 'ought,' 'good,' 'reasons,' and the like. Consider then a simple sentence containing one of these expressions, for example the sentence 'I should tell the truth', which could be uttered either sincerely or insincerely. We should then think that, to count as someone who can utter this sentence sincerely, you must be in a certain psychological state. After all, someone who utters the sentence insincerely does not really think that she should tell the truth.

We can then use 'normative judgment' as a neutral label for the psychological state in which you must be to count as someone who can sincerely utter the corresponding normative sentence – whatever that state is like.[13] As a neutral term, it refers to the judgments that employ normative concepts, but it does not take a stand on what kind of mental states these judgments are.

---

[12] Jon Tresan (2006) argues that internalism should be understood as a *de dicto* claim according to which we cannot call a mental state a normative judgment unless it was accompanied by motivation (i.e., the state can be found in other possible worlds without motivation but in those worlds the words 'normative judgment' do not apply to it). He argues that this thesis is compatible with externalist *de re* views about the nature and the content of normative judgments. This is why I formulate the considered views as *de re* claims about the relevant psychological processes.

[13] Here I follow Schroeder (2008, §5.1). The relevant state can be a compositional state that consists of many different individual states (such as a state of motivation and other inert states). For an example, see the hybrid first-order expressivist views discussed below.

The Type A views claim that the psychological state that constitutes a positive normative judgment concerning an action must itself *consist at least in part of* the state of being motivated to do that action. On these views, if you can sincerely use the sentence 'I should tell the truth' to make an assertion, the state you must thereby be in itself contains at least some motivation to tell the truth. These views are thus committed to unconditional internalism – that, necessarily, whenever you judge that you should φ, you have at least some motivation to φ. This modal connection is guaranteed because any positive normative judgment concerning an action will itself contain some motivation to do that action.

There are two main versions of the Type A views. The first kind of versions accept the Humean picture of human psychology according to which all mental states are either belief-like cognitive states or desire-like conative states.[14] The purpose of the belief-like states is to represent the world correctly: they have the mind-to-word direction of fit. In contrast, the purpose of the desire-like states is to motivate the agent to act in order to make the world fit how the agent wants it to be; they have the world-to-mind direction of fit. These states also have different functional roles. Belief-like states are causally sensitive to thoughts about evidence and motivationally inert, whereas desire-like states are insensitive to evidence (I do not lose my desire to go to California even when I know I am not there now) and yet they are able to push us to act (together with means-end beliefs).

According to the first kind of Type A views, normative judgments are at least in part desire-like states with a certain specific content. Following Toppinen (2015, 151), we can call these views 'first-order expressivist' views.[15] They claim that, for example, making the normative judgment of some action, φ, that you should do that action *consists at least in part* of a desire-like attitude toward φing. Thus, according to the basic first-order expressivist view, my judgment that I should tell the truth just is my desire to tell the truth. According to more complex views, even if that desire is one part of my judgment, my judgment also contains other mental states within it. One *pure* complex first-order expressivist view might claim, for example, that the previous judgment consists of both my desire to tell the truth and an additional desire not to lose that desire (Blackburn 1998, 66–70). Similarly, one *hybrid* first-order view might claim that my judgment consists of both my desire to tell the truth and my belief that I would always have this desire towards all instances of telling the truth.

The second kind of Type A views reject one central assumption of the previous Humean picture: the claim that no psychological state can have both mind-to-world and world-to-mind directions of fit. When this assumption is rejected, one can argue that, even if normative judgments, *as unitary single states*, aim at representing the normative reality correctly, they also contain motivation to act accordingly. These states that would have both belief- and desire-like qualities are known as 'besires' (Altham 1986).[16]

---

[14] See Smith (1994, 7–9).

[15] These views (different forms of emotivism, prescriptivism, expressivism and non-cognitivism) have a long history in metaethics (see Schroeder 2010, chs. 2–4).

[16] This view is sometimes attributed to John McDowell (see his 1978, 18–22) and Smith (1994, 121–5)).

2.2 Type B Views

In order to distinguish themselves from the Type A views, the Type B views claim that normative judgments and the motivation to act accordingly are two wholly distinct mental states: the latter state simply cannot be even a part of the former. Take the sentence 'I should tell the truth' again. The defenders of Type B views think that the psychological state in virtue of which you count as someone who can sincerely utter that sentence cannot be even in part your desire to tell the truth – it must be a different mental state altogether.[17]

This formulation is still neutral about the nature of normative judgments. The cognitivist Type B views take them to be paradigmatic beliefs whereas the non-cognitivist Type B views take them to be some kind of desire-like states. This non-cognitivist option is available so long as the desire-like state in question is not itself the desire to do the relevant action itself and does not even include that desire as its constituent.

Let me illustrate the non-cognitivist views with two examples, starting from 'second-order expressivism' (Toppinen 2015, 151). It claims that, even if my normative judgment that I should tell the truth is distinct from my desire to tell the truth, this judgment still consists of a desire-like attitude towards something else that is suitably connected to truth-telling. One view might, for example, claim that my judgment that I should tell the truth consists of a plan or a desire to desire to tell the truth.[18]

Similarly, according to more sophisticated hybrid second-order expressivism, a normative judgment consists roughly of a pro-attitude towards actions that have certain specific properties and a belief that the considered action has those properties (Ridge 2014, ch. 4). Thus, my judgment that I should tell the truth might consist of a pro-attitude towards maximising general happiness and a belief that truth-telling has that property. Here the normative judgment itself would be an aggregate of a belief and a desire-like state whereas the motivation to do the relevant action would be a different desire-like state.

Type B views then argue that, in order to explain how normative judgments shape our motivations, we should think that normative judgments have a direct causal power to produce motivational states in us *due to the kind of states they are*. On this view, a part of the essence of normative concepts is that they can produce motivation in whoever uses those concepts to make normative judgments. In this case, we would not need to refer to any other factors to explain how normative judgments motivate. Type B views thus also entail that, if an agent utters a normative sentence but is not motivated to act accordingly, her utterance does not express a genuine normative judgment. If the agent had made such a judgment, she would be motivated given the causal powers of her judgment.

---

[17] Someone could object to my distinction between Type A and B views by describing a view that does not seem to fall naturally to either category. On this view, normative judgments are motivationally inert beliefs, but in order for these beliefs to count as a genuine normative judgment, the subject must also have a separate, accompanying motivational state. Would this be a Type A or Type B view? The problem with this objection is that the way I defined normative judgments above makes the described view incoherent. I defined normative judgments in terms of psychological sincerity-conditions and so a normative judgment cannot be a mere belief if the belief does not count as a normative judgment without the accompanying motivation state that is taken to be a part of the sincerity-condition.

[18] See Toppinen's (2015, 151) interpretation of Gibbard (2003, 142–3).

When these views are formulated in this way, they too entail unconditional internalism: that, necessarily, whenever you judge that you should φ, you will have some motivation to φ. This is a problem because there are well-known counterexamples to unconditional internalism, including amoralists, listless and depressed agents, and evil people.[19] Take the following case from Mele (2003, 111):

> Consider an unfortunate person – someone who is neither amoral nor wicked – who is suffering from clinical depression because of the recent tragic deaths of her husband and children in a plane crash. Seemingly, we can imagine that she retains some of her beliefs that she is … required to do certain things… while being utterly devoid of motivation to act accordingly…. She has aided her ailing uncle for years, believing herself to be … required to do so. Perhaps she continues to believe this but now is utterly unmotivated to assist him.

Type B views, as described, would entail that, because the power to produce motivation belongs to the essence of normative judgments, the relevant state that fails to produce motivation in the unfortunate agent cannot be a normative judgment. Yet, more plausibly, her normative judgments have not changed due to her depression.

We should therefore re-formulate these views so that they would be committed only to conditional internalism: the thesis that, necessarily, if you judge that you should φ, either you have some motivation to φ or you fail to satisfy a certain condition C.[20] The previous counter-examples can then be avoided because the agents in them fail to satisfy the relevant condition C and so their lack of motivation need not entail that they are not making genuine normative judgments.

As a result, we get a view according to which (i) normative judgments and the relevant motivations must be wholly distinct mental states with no common parts and (ii) normative judgments have the power to produce motivation *at least when certain conditions C are satisfied*. This formulation, however, makes it difficult to distinguish Type B views from the Type C views (to be introduced in more detail below in §2.3).

The latter views claim that, in order to explain why normative judgments are usually accompanied by motivation, we need to posit a third mental state that helps the inert normative judgments to shape our motivations. However, the previous description of the Type B views threatens to make such views Type C views too because being in the third state could be claimed to constitute the relevant conditions in which normative judgments have the power to produce motivation.

In order draw a meaningful distinction, we must therefore add a further clause to the definition of the Type B views. I will stipulate that the conditions C in which, according to these views, normative judgments have the power to shape motivations must consist of some *general* qualities or dispositions of agents. By general I mean here that these qualities or dispositions cannot govern merely the connection between an agent's normative judgments and her desires in the

[19] For references to discussions of these cases, see Miller (2008, 235–236 fns. 8–10).

[20] For defenders of different versions of this claim, see Björklund et al (2012, 126–8).

agent's psychological make-up. Rather, they must regulate her psychological states more generally and be describable without making an ineliminable reference to the agent's normative judgments.

Traditional forms of conditional internalism take the conditions in which normative judgments have the power to produce motivation to consist of psychological normalcy and/or practical rationality.[21] These qualities are then claimed to consist of general dispositions towards coherent and unified sets of beliefs and desires.[22] Psychologically normal and practically rational agents tend to get rid of inconsistencies and acquire new beliefs and desires that support each other. Because these dispositions govern the agents' psychological make-ups generally and they can be described without referring to their normative judgments, the views that rely on them are Type B views.

Consider the first non-cognitivist Type B view introduced above. It claimed that normative judgments consist of plans to have certain first-order desires. This would give normative judgments the power to produce motivation in agents who are disposed towards coherence. These dispositions explain why an agent's judgment that she should tell the truth (i.e., her plan to desire to tell the truth) has the power to produce motivation to tell the truth in her. After all, given the agent's general plan to desire to tell the truth, having a desire to tell the truth is more coherent.[23]

There are also cognitivist Type B views. According to Michael Smith (1994, §5.9), the content of normative judgments is about what our fully rational versions would want us to do in our actual circumstances. Thus, on his view, your judgment that you should tell the truth is a belief that our fully rational versions would want us to tell the truth in your situation. Smith then argues that this belief would cohere with the desire to tell the truth whereas lacking that desire would be incoherent (ibid., 177). This is why, on Smith's view, insofar as you are practically rational and disposed towards coherence, your normative judgments, as beliefs, have the power to produce motivation in you.

Where Smith's view is a version of reductive analytic naturalism, T.M. Scanlon has outlined a corresponding non-reductivist, non-naturalist Type B view. Scanlon (2014, 54–5) grants that fully rational agents intend to do what they judge they have conclusive reasons to do. His explanation of this correlation has the same structure as Smith's. Both explain the connection between normative judgments and motivation by first specifying the content of normative beliefs. After this, they rely on rational agents' disposition towards coherence to explain how the judgments with the specified content produces motivation. However, instead of taking normative judgments to be about what our idealized versions would want us to do, Scanlon (2014, 57) claims that normative judgments are about a distinct non-natural, *sui generis* reasons-relations. He then argues

---

[21] For psychological normalcy, see, for example, Blackburn (1998, 59–68) and Gibbard (2003, 154). One worry one might have about these views is that only some of the ways in which you can be psychologically abnormal can interfere with the connection between normative judgments and motivation. See, for example, Jeppsson (2021). For practical rationality, see, for example, Smith (1994, ch. 3), Korsgaard (1996), and Wedgwood (2006, §1.3).

[22] For descriptions of the coherence aspect of psychological normalcy and practical rationality, see, for example, Blackburn (1998, 52–8), Smith (2004), and Scanlon (2007).

[23] Similarly, the second hybrid expressivist view claimed that a normative judgment is a desire to do actions that have a certain property P and a belief that the relevant action has that property. Here too, a rational agent disposed towards coherence will form a desire to do the action in question because it is supported by the previous two mental states.

that it is more consistent to have the motivations that match those judgments, given their unique subject matter.[24]

2.3 Type C Views

Finally, Type C views claim that, to explain how normative judgments shape our motivations, we need to rely either on a third mental state, usually some desire-like state, or a local disposition that governs the connection between normative judgments and motivation.[25] This section introduces three main versions of these views.[26]

Let me begin from a view, which Michael Smith (1994, 71–6) first attributed to externalists and which was then adopted by some actual externalists (Lillehammer 1997; Shafer-Landau 1998, 357 and 2003, 159). On this view, your judgment that you should tell the truth leads to motivation to do so only if you have a distinct *de dicto* desire to do whatever you should do. Let's assume that you have this desire – you desire to do what you should do, under that description and whatever those actions are. If you then also believe that telling the truth is one of the things you should do, your desire to do what you should do will produce in you a desire to tell the truth. Having that desire will, after all, be instrumental to satisfying your more basic desire. Yet, on this view, if you lack the relevant *de dicto* desire, your normative belief will not produce any motivation in you.

Here the relevant judgment and the *de dicto* desire will produce the new desire only if we assume that you are psychologically normal and practically rational. Such agents, due to their disposition towards coherence, tend to form the instrumental desires needed for satisfying their final desires. This view also requires that we have many *de dicto* desires that match all the normative concepts that we employ (a *de dicto* desire to do what one ought to do, a *de dicto* desire to do what is good, and so on). Otherwise not all normative judgments would lead to new motivation to act accordingly.

James Dreier (2000, 629–38) introduced a second version of Type C views. It suggests that we have a second-order desire to desire to do what we judge we should do. If you then have this desire and you judge that you should tell the truth, you will form a desire to tell the truth, again insofar as you are disposed towards coherence. After all, having the desires you desire to have is more coherent than not having them. And, here too, in order to explain how different normative concepts motivate, we need to assume that we have many higher-order desires that correspond to the different normative concepts we employ.

The third view claims that, instead of any specific desires, we have substantial local dispositions to have desires that match our normative judgments (Copp 1997, 50; Dreier 2000, 623–9). This is to claim that we desire to do what we judge we should because we happen to have a disposition

---

[24] One advantage of the non-cognitivist and Smith's Type B views is that they can explain of what the relevant consistency consists. There is a worry that Scanlon merely states that the relevant combination of a normative judgment and intention is more coherent without explaining why (Dreier 2015, 161–6).

[25] The defenders of the resulting views take normative judgments to be ordinary beliefs about either natural (see, for example, Brink (1989)) or non-natural properties (see, for example, Shafer-Landau (2003)).

[26] There is one form of externalism that does not fit my taxonomy. Some externalists argue that people tend to act according to their normative judgments because they have pre-existing desires to do certain things, which they then come to judge to be worth doing *post hoc* (see, for example, Svavarsdottir (1999, 198–9) and Shafer-Landau (1998, 356 and 2003, 158–60)). I do not discuss these views because they offer no explanation of how new normative judgments can *shape* our motivations (as acknowledged by Shafer-Landau (1998, 355–6)).

to form desires to do the things that we judge we should. Here this disposition is not understood as a general disposition towards coherence but rather as a disposition the inputs of which must be judgments employing specific normative concepts and the outputs of which are desires to act accordingly.

3. An Evolutionary Argument against the Type C Views

§1 already introduced a reasonable assumption according to which our ability to use normative concepts to shape our motivations is an *adaptation*. This ability arguably enabled us to plan and co-operate in much more efficient ways and that improved our chances to survive as a species.[27] In the terminology of evolutionary biology, the psychological process that is responsible for *how* the employment of normative concepts in normative judgments shapes our motivational states could then be called a *proximate mechanism*.[28] We should then think that the proximate mechanism that is causally responsible for an evolved adaptation must have also evolved. This allows us to investigate whether the general perspective of evolutionary biology on how proximate mechanisms evolve could shed light on which particular mechanism is responsible for the way in which our normative judgments shape our motivations (assuming that the latter ability is an adaptation).[29]

Evolutionary biologists call the question of which proximate mechanism is causally responsible for a given adaptation a *design problem*. We can hence understand the three views outlined in §2 as competing solutions to the same design problem as one of these views must describe the proximate mechanism that provided us with the relevant trait, which we are assuming is an adaptation.

When there are multiple solutions to a design problem, evolutionary biologists rely on certain general principles to make predictions that can then often be empirically verified. Three principles have been found to lead to accurate predictions: the principles of *availability, reliability*, and *efficiency* (Sober 1999, 142–3; Sober and Wilson 1998, 304–6).[30] Let us consider an example (ibid.).

---

[27] In addition, it could be argued that normative concepts that can shape our motivations effectively also enable us to have more accurate self-knowledge concerning our desires, intentions, plans and other practical attitudes (see Suikkanen 2018). Such self-knowledge can also further enable us to be efficient planners and co-operators.

[28] See, for example, Mayr (1963), Scott-Phillips et al (2011), and Sterelny (2013).

[29] My argument is inspired by Elliott Sober's corresponding argument against psychological egoism (Sober 1994 and 1999; Sober and Wilson 1998, ch. 10), which is best understood as not establishing the truth of psychological altruism but rather as merely pointing to new, previously neglected evidence for the view (Schulz 2011, 253). My argument is intended in the same modest spirit, but it is still significant for two reasons. Firstly, given the current stand-off in the internalism versus externalism debate, anything that provides even a modest increase in the probabilities matters. Secondly, the argument offers us a way to predict proximate mechanisms governing dynamics we do not yet understand after which we should be able to check if that adaptationist analysis really describes the phylogenetic history of a trait by doing the regular empirical tests of development, gene variants, homologies, cross-cultural comparisons, and so on. For objections to Sober's argument, see Stich (2007). Stich's objections target only the elements of Sober's argument that relate to the discussion of egoism and not the general structure of the argument or the principles it relies on. For responses, see Schulz (2011). For a defense of the idea that we should rely on evolutionary selective explanations of mental mechanisms, see Sterelny (1993).

[30] Note, though, that given the randomness of natural selection there are cases in which even these principles do not lead to correct predictions (James 2011, 127). It's just that they more often than not do so.

Certain marine bacteria must avoid oxygen to survive. This particular design problem could be solved in different ways. The *direct* strategy would be an oxygen detector that warns the bacteria of the presence of oxygen. An *indirect* strategy would be a detector that is sensitive to some other environmental variable, like depth, that is correlated with the presence of oxygen. Some organisms thus use magnetosomes to help them to use the earth's magnetic field to direct them downward to deeper water with less oxygen. Finally, *pluralistic* strategies rely on many different detectors at the same time.

Which of these strategies is natural selection then most likely to produce? *Availability* states that natural selection can only act on an actual range of variation. Thus, even if either one of the previous detectors were nice to have, natural selection cannot cause that detector to evolve unless some bacteria first have it as an ancestral variant due to a mutation.

Evolutionary biologists then ask which mechanism would give the relevant bacteria the greatest chance of survival by indicating *reliably* the presence of oxygen in their environment. This cannot be determined *a priori*: there is no antecedent reason for why either a direct or an indirect strategy would be more reliable. We *can*, however, describe circumstances in which the two strategies would differ in their reliability. There is no antecedent reason to assume that an oxygen detector would be more reliable in detecting oxygen than a depth detector in detecting depth. Therefore, when there is a perfect correlation between depth and the presence of oxygen, both strategies are equally reliable. However, whenever the presence of oxygen is not perfectly correlated with depth, the direct strategy is more reliable. This is the so-called 'Direct/Indirect Asymmetry Principle' (Sober and Wilson 1998, 306).

To this we can also add the 'Two-Is-Better-than-One Principle' according to which having many distinct reliable (though fallible) detectors, direct and indirect, is more reliable than having just one detector so long as these detectors do not interfere with each other (ibid., 306–7). This is because it is less likely that two distinct mechanisms fail simultaneously.

The final consideration relevant for predicting which proximate mechanism evolves is *efficiency*. Sometimes even if a reliable mechanism is an ancestral variant, natural selection will not sustain it because building and sustaining that mechanism requires more energy. After all, efficiency matters for survival just as much as reliability.

Let us then return to which proximate mechanism could be causally responsible for how our normative judgments shape our motivations. Evolutionary biology now enables us to predict which psychological mechanism is likely to have evolved in the lineage leading to us.[31]

This section compares the Type B process and the Type C process. What kind of concepts were our ancestors inclined to employ in planning and co-operation contexts? We first need to assume that there were some actual human populations whose members relied in these contexts on concepts that had, in their own right, a direct causal power to shape motivation (at least in agents disposed towards coherence) as captured by the Type B views. We furthermore need to assume that there were also other human populations whose members relied, in similar contexts, on

[31] Even if my argument for the two state views is empirical, it still fits the idea that the truth of the two state views is conceptual. A part of the nature of normative concepts may well be that they have the power to produce motivation even if there is an evolutionary explanation of how we came to employ such concepts in deliberation (Kauppinen 2008, 8 fn. 17).

concepts that were unable to shape motivation without the help of the third states as described by the Type C views. When the members of these second kind of populations then used their concepts in deliberation, their motivations conformed to their judgments only if they also had some other additional desires or local dispositions.

Let us then use the *availability, reliability*, and *efficiency* principles to predict which one of the previous communities would have been more likely to survive the natural selection process.[32] Firstly, there is no reason to think that either one of the previous mechanisms would not have been available. Both require the same Humean belief/desire psychology and both enable agents to use their respective concepts to shape their motivations. The only difference is that the Type B agents are employing concepts that can shape motivation with the help of the general disposition towards coherence whereas the Type C agents' concepts shape motivation only if the agents have certain additional desires or local dispositions. If one of these mechanisms did not evolve, this was thus probably not because it was not available as an ancestral variant.[33]

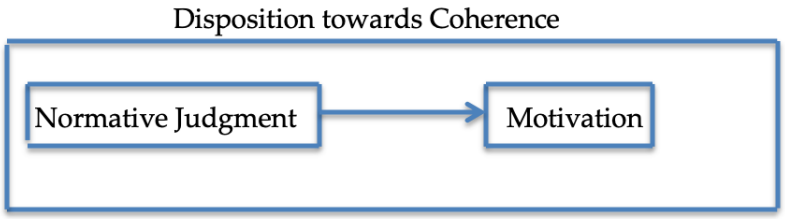
Similarly, *efficiency* cannot decide between the two mechanisms because it does not cost any more calories to build and maintain either one of them. Both Type B and C agents have the same Humean belief/desire psychology and what requires energy is building the hardware that implements it. Which beliefs and desires the organisms with that hardware then come to have does not make an energetic difference (Sober 1999, 146).

This means that, if evolutionary biology supports either view, this must be based on *reliability*. Here I want to argue that this principle predicts that we evolved to have a Type B rather than a Type C system. Let us consider how reliable the two mechanisms are. Suppose that I judge that I should tell the truth when my neighbour asks whether I scraped her car when parking my own car. The figure below shows how the Type B and C mechanisms work.

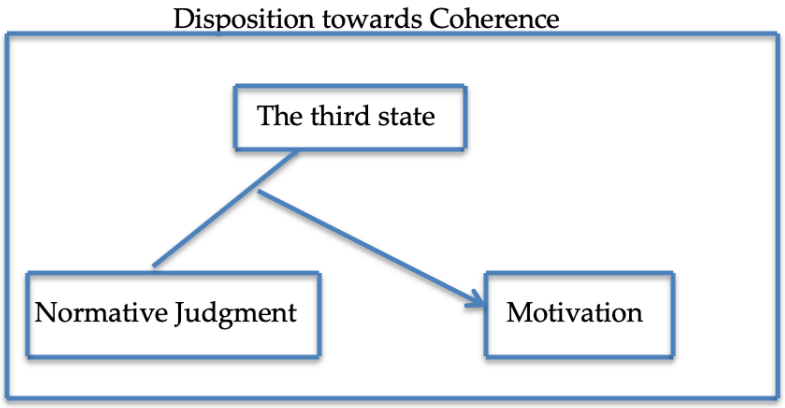
[32] Here I rely on the controversial idea of group selection. For defences, see Sober and Wilson (1998, chs. 2–5). For how the argument can be reformulated so as to rely only on individual selection, see Sterelny (2000, 277–81).

[33] For an analogical argument, see Sober (1999, 146).

Type B:



Type C:



Thus, if I am a Type B agent, I am using a concept such that my disposition towards coherence suffices to give employing that concept in a judgment the power to shape my motivations. In contrast, if I am a Type C agent, I must have a certain third state: a *de dicto* desire to do what I should do, a second-order desire to desire what I judge I should do, or a local disposition to form desires that match my judgments. The content of my normative judgment must then connect to the content of the third state in the right way. When this happens and I am disposed towards coherence, my normative judgment and the third state produce the desire tell the truth in me.

This makes the Type B mechanism a more *direct* solution to the design problem. This in itself is not a reason to prefer the Type B views as *a priori* we have no reason to assume that direct strategies are more reliable than indirect ones. However, the Direct/Indirect Asymmetry Principle applies again: We *can* describe circumstances in which the indirect Type C mechanisms are less reliable.

We are looking for a psychological mechanism that enables an agent's normative judgments to shape her motivations reliably. If there is such a mechanism, then, when an agent makes a normative judgment, she almost always comes to have the motivation to act accordingly. Furthermore, even if the motivation she comes to have need not always be overriding motivation, it must be often enough, or the agent will not be an effective planner and co-operator.

This means that the Type C views describe a reliable mechanism only if the third state they posit has a sufficiently high degree of *strength*.[34] These views thus need to claim that we generally have *a sufficiently strong de dicto* desire to do what we should do, or *a sufficiently strong* desire to desire what we judge we should do, or a *sufficiently strong* disposition to desire what we judge we should do. Otherwise, the psychological mechanism they describe would not be reliable often enough.

We can see that this is the case when we consider agents who have, for different reasons, strong antecedent desires not to do what they judge they should do. In these agents, the third state can bring about motivation that is sufficiently strong for planning and co-operation purposes only if the third state has sufficient strength *to impose* the new overriding desires on the agents against the resistance coming from their other desires. Thus, in the previous example, there are many reasons why I do not want to tell the truth to my neighbour: doing so would cost money and make my neighbour angry. If I then only have a weak desire to desire what I judge I should do, this desire (and my belief that I should tell the truth) cannot cause me to have a sufficiently strong desire to tell the truth given how much my other desires will resist having that first-order desire. This is why, in order to create a reliable enough connection between normative judgments and motivation generally, the third state has to be a sufficiently strong desire or a disposition.

The problem is that achieving this level of strength in the third state would require 'tricky engineering' – it would be difficult for natural selection to implement (Sober and Wilson 1998, 315). The standard ways in which the third state could be coded into our genes and be produced by them tend to lead to many individuals getting only a weak third state (or perhaps not having it at all). After all, there are very few, if any, substantial desires that are biologically hard-wired to *all* human agents in a very strong form. This means that the standard ways for implementing the third state in human psychology would make the connection between normative judgments and motivation unreliable in many. Furthermore, the biological implementation of the third state would also have to be such that its considerable strength would remain constant over our lifetimes. Yet, even our most basic biologically hard-wired desires and dispositions do not seem to be like that but rather their strength seems to vary in different contexts. I may be biologically hard-wired to desire sex, food, drink, and shelter but often these desires are not very strong.

Therefore, whenever the circumstances are such that some individuals have the third state only in a weaker form, the motivations of these individuals will be shaped less reliably by their normative judgments. This makes the populations consisting of Type C individuals less adapted insofar as more reliable connections between normative judgments and motivation enable more efficient planning and co-operation. Furthermore, avoiding this problem with sufficiently strong and stable third states in all members of a population would be a tricky engineering task even for natural selection.

Type B views, in contrast, avoid the previous problem. The direct mechanisms through which normative judgments shape motivation according to them are reliable in ordinary rational agents who are disposed towards coherence. Producing these mechanisms also does not require the kind of tricky engineering that would be needed for producing the third states in a sufficiently strong

[34] For an analogical argument, see again Sober and Wilson (1998, 315–6).

form in most human beings. This is one reason for why the reliability principle supports the prediction that we evolved to have a Type B mechanism.

We can also consider the role of coherence. In both Type B and C frameworks, our disposition towards coherence does important work. Type B views claim that this disposition produces the relevant motivation to act according to a normative judgment because the combination of those two mental states is more coherent. Similarly, two of the Type C views suggest that the relevant motivation is produced by the disposition towards coherence as having the relevant motivation coheres better with the normative judgment and either the relevant *de dicto* desire or the relevant higher-order desire.

You might then think that the disposition towards coherence and the relevant normative judgments shape the motivations of the Type B and C agents equally reliably. There is, however, one situation in which the disposition towards coherence will function more reliably in the Type B agents.

Consider a situation in which an agent holds two mental states such that the adoption of a certain third state would make the agent more coherent. Here the disposition towards coherence can produce the third state only if the other two states are connected to one another in the right way. Imagine that I believe that my business has been on a downward trend for three months. Based on general statistical evidence, I also believe that businesses that are on a downward trend for three months usually fail. If I am disposed towards coherence, I should then believe that it is likely that my own business will fail.

However, in this type of cases, our disposition towards coherence often fails to form the relevant third belief in us. This is because we tend to *compartmentalise* our beliefs about our own success: we tend *to isolate them inferentially* from general statistical information. There is a simple explanation of why we do this: we desire to succeed so badly that we often do not care about the likelihoods. And, sometimes this serves a purpose by making us try harder.

Return then to the Type C views. It turns out that our disposition towards coherence will produce the relevant motivation based on our normative judgment and the third state (a *de dicto* or a higher-order desire) only if these two states are connected in the right way and not compartmentalised and inferentially isolated. And, here too, there exists various psychological mechanisms, including biases and brute emotions, that would make us isolate our normative judgments from our *de dicto* and higher-order desires. For example, acting in a way that complies with our normative judgments is often demanding. This means that there can be circumstances in which the disposition towards coherence will fail to do its work properly in the Type C mechanisms.

Admittedly, Type B views too rely on that same disposition and there are also circumstances in which the disposition towards coherence can fail to shape the motivations of the Type B agents. However, note that there will be *fewer* of these situations as according to these views the relevant coherence relation obtains directly between the normative judgments and the relevant motivations. Because of this, in this framework, there cannot be situations in which an agent's normative judgment will fail to shape her motivations because her *other* mental states are compartmentalised. This is why there is also a second kind of circumstances in which the Type B

mechanism is more reliable. For this reason too, the reliability principle predicts that natural selection has produced a Type B mechanism in us.

Finally, recall the Two-is-Better-than-One Principle, which too is supported by the reliability principle (Sober and Wilson 1998, 307). It claims that having many different reliable mechanisms is even more reliable whenever these mechanisms do not interfere with one another. This is why natural selection tends to go for in-built redundancy whenever possible.

In this context, the Two-is-Better-than-One principle supports the prediction that we evolved to have both the Type B and the Type C mechanisms. The suggestion thus is that we evolved to use the kind of concepts in deliberation that have the power to shape our motivations directly insofar as we are disposed towards coherence. However, this proposal adds that we also evolved to have the relevant *de dicto* desires, higher-order desires, and/or local dispositions too. After all, if we evolved to have both mechanisms, then, even if one fails, the other will still enable our normative judgments to shape our motivations.

From the perspective of evolutionary biology, the previous proposal seems plausible, and the Type B theorists can perfectly well accept it. They have no reason to oppose the idea that, even if employing normative concepts can usually directly shape our motivations, there can also be other indirect psychological mechanisms at work at the same time. In contrast, it is more difficult for the Type C theorists to acknowledge that, even without the help of the third state, the normative judgments can themselves shape our motivations. This is why the Two-is-Better-than-One Principle too supports the Type B views.[35]

4. Against the Type A Views

You might then, however, think that availability, reliability, and efficiency support the Type A views even more. Given that these views are based on similar Humean Desire/Belief psychologies, it is unlikely that the psychological mechanisms described by them would have been any less available or efficient than the Type B mechanisms. Yet, surely the Type A mechanisms are even more reliable than the Type B ones. If a normative judgment to do a certain action consisted at least in part of a desire to do that very action, there would be a perfect correlation between normative judgments and the corresponding motivations. This is why the principles of evolutionary psychology seem to support the Type A views the most. This final section suggests that, despite this, we should not think that we evolved to have a Type A system.

§2.1 already explained the two types of Type A views: first-order expressivist and besire views. I will first use the availability principle to rule out the besire views. That principle states that natural selection can only act on an actual range of variation. The problem with the besire view is that it is unlikely that such states ever existed or even could have existed.

[35] Some philosophers might at this point object to positing two simultaneous mechanisms on the grounds of 'parsimony' – the thought that explanations that assume the existence of fewer entities tend to be true. However, here we are not positing the existing of many separate physical systems. Rather the thought is that there is just one 'hardware', our brain, that has evolved to carry out two independent psychological processes simultaneously in the same Humean Desire/Belief framework. In this sense, the view is not positing more of different kinds of entities. Also, as noted, there is some empirical support for the idea that natural selection tends to prefer built in redundancy. That's why, for example, we have two kidneys.

The existence of besires, as single unitary states with both directions of fit, would require that there could be beliefs, states that represent how the world is, that you could not be in unless you also simultaneously desired to do certain actions. We should not, however, think that there are such states because it *is* always possible to have any belief whatsoever without also at the same time being in any particular desire-like state (Smith 1994, 119). The Humeans are right to insist that beliefs and desires can at least in principle be pulled apart modally as distinct existences.

One crucial piece of evidence for this is that there are different forms of practical irrationality, cases of weakness of will, tiredness, depression, and the like, that can break the connection between normative judgments and motivation without changing the judgments (Stocker 1979, 744). The flaw of the besire views is that either (i) they have to deny the existence of these forms of practical irrationality or (ii) they have to claim that the agents who suffer from them no longer see the world in the same way as they did before. Yet, because these responses are not appealing, we should agree with the Humeans that there just cannot be besires and so we could not have evolved to be in these states.

This leaves us with the first-order expressivist views. According to them, my judgment that I should tell the truth *consists at least in part* of a desire to tell the truth. These views face a dilemma: they are either empirically flawed or more appropriately understood as Type B views.

It is widely agreed that the most basic formulations of these views are empirically false. The main evidence against them consists of the well-known cases like the unfortunate widow described by Mele (§2.2 above). As Michael Smith's (1998, 161) fictional character Cog puts it:

> After all, it is a commonplace that when (say) someone suffers from a deep depression then they may have no desire at all to do what they judge to be desirable. They see all the good to be done, but have no inclination to pursue it. It would be quite incredible to suppose that they temporarily fail to understand that they are making an evaluative judgement when they judge something desirable or worth achieving! Indeed, one of the more depressing aspects of depression is the fact that the value of the things that leave you unmoved is especially vivid to you.

Given that there are cases in which we make normative judgments and have no desire to act accordingly, we should conclude that the simple formulations of the first-order expressivist views are not tenable. We did not evolve to have the psychological mechanisms described by these positions because those mechanisms were not available for natural selection.

There is, however, an expressivist response to the previous objection, which draws a distinction between *desires* and *motivation*.[36] The idea is that we should understand desires as dispositional states that normally push us to states of being motivated even if, sometimes, they may fail to do so. This suggestion helps the first-order expressivists to claim that the normative judgment that I should tell the truth consists of my desire to tell the truth after all. They can argue that, in the alleged counterexamples where the relevant agents lack motivation, the agents continue to make genuine normative judgments because they continue to have the same dispositional desires. It's

[36] See, for example, Blackburn (1998, 61), Gibbard (2003, 154), and Toppinen (2015, §8.1).

just that in these abnormal circumstances these dispositions fail to produce any actual motivation in the agent who has the disposition.

This view may be correct, but it is not a Type A view. The difference between the Type A and the Type B views is whether the state of being motivated to act according to a normative judgment is (i) an element of the judgment or (ii) a distinct state. The defenders of the previous response must reject (i) and accept (ii) instead. After all, they claim that a normative judgment consists of a dispositional desire to act in a certain way and yet acknowledge that you can be in this dispositional desire-state without also being motivated to act in the relevant way. And, like the defenders of the Type B views, the defenders of this view claim that the normative judgments lead to being motivated only when the agent is psychologically normal and practically rational (Toppinen 2015, 157). This is why the Type A first-order expressivists are able to avoid the relevant counterexamples to their view only by formulating a position that is a Type B view instead.

5. Conclusion

I first introduced the Type A, Type B, and Type C views that offer competing explanations of how our normative judgments shape our motivations. This offered us a new and, in my mind, more illuminating way of classifying different views of how normative judgments and motivation are related. I furthermore briefly commented on how the psychological mechanisms described by these views are based on the different accounts of normative judgments in metaethics, and how they relate to the more traditional ways of formulating the debate in terms of motivational internalism and externalism.

In the rest of the paper, I then argued for the Type B views with a wholly new kind of an argument in this debate by relying on the methods of evolutionary biology. §3 claimed that evolutionary biology provides new support for the idea that we evolved to have a Type B rather than a Type C mechanism. This is because having a psychological mechanism that enables normative judgments to shape our motivations can firstly be assumed to be an adaptation. This allows us then to use the principles of evolutionary biology to predict which proximate mechanism evolved to be causally responsible for that adaptation. I suggested that the principle of *reliability* supports the Type B mechanisms because there are several ways in which the Type C mechanisms would in many circumstances be less reliable than the Type B mechanisms.

Finally, §4 argued against the Type A views. The availability principle rules out the besire views, whereas the first-order expressivist views are either empirically false or better understood as Type B views. This is why, of the three alternatives, we have most reason to think that the way in which our normative judgments shape our motivations is correctly described by a version of the Type B theory. Our normative concepts are likely to be such that when we employ them in normative judgments they have the power to produce motivation in us at least insofar as we satisfy certain very general psychological conditions that do not merely govern the relationship between our normative judgments and motivation, such as that we are disposed towards coherence. It is then a further question which specific version of the Type B mechanisms we evolved to have, but perhaps this question too can be investigated further in the future by relying on similar methods of evolutionary psychology in a more fine-grained fashion.

References

Altman, J.E.J. 1986. The legacy of emotivism. In *Fact, science and morality: essays on A.J. Ayer's* Language, Truth and Logic, ed. Graham MacDonald and Crispin Wright, 275–88. Oxford: Blackwell.

Björklund, Fredrik, G. Björnsson, J. Eriksson, R. Francén Olinder and C. Strandberg. 2012. Recent work on motivational internalism. *Analysis* 72: 124–37.

Björnsson, Gunnar. 2002. How emotivism survives immoralists, irrationality, and depression. *Southern Journal of Philosophy* 40: 327–44.

Blackburn, Simon. 1998. *Ruling passions: a theory of practical reasoning*. Oxford: Oxford University Press.

Brink, David. 1989. *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.

Copp, David. 1997. Belief, reason, and motivation: Michael Smith's "The Moral Problem". *Ethics* 108: 33–54.

Cuneo, Terence. 1999. An externalist solution to the "Moral Problem". *Philosophy and Phenomenological Research* 59: 359–80.

Dreier, James. 2000. Dispositions and fetishes: externalist models of moral motivation. *Philososophy and Phenomenological Research* 61: 619–38.

Dreier, James. 2015. Another world: the metaethics and metametaethics of reasons fundamentalism. In *Passions & projections: themes from the philosophy of Simon Blackburn*, ed. Robert Johnson and Michael Smith, 155–71. Oxford: Oxford University Press.

Gibbard, Allan. 1990. *Wise choices, apt feelings*. Cambridge, MA: Harvard University Press.

Gibbard, Allan. 2003. *Thinking how to live*. Cambridge, MA: Harvard University Press.

Hauser, Marc. 2006. *Moral minds*. New York: HarperCollins.

James, Scott. 2011. *An introduction to evolutionary ethics*. Oxford: Blackwell-Wiley.

Jeppsson, Sofia. 2021. *Philosophy, Psychiatry, and Psychology* 28: 233–49.

Joyce, Richard. 2006. *The evolution of morality*. Cambridge, MA: Harvard University Press.

Kauppinen, Antti. 2007. The rise and fall of experimental philosophy. *Philosophical Explorations* 10: 95–118.

Kauppinen, Antti. 2008. Moral internalism and the brain. *Social Theory and Practice* 34: 1–24.

Kitcher, Philip. 2006. Biology and ethics. In *Oxford handbook of ethical theory*, ed. David Copp, 163–85. Oxford: Oxford University Press.

Korsgaard, Christine. 1996. Skepticism about practical reason. *Journal of Philosophy* 83: 5–25.

Lillehammer, Hallvard. 1997. Smith on moral fetishism. *Analysis* 57: 187–95.

Mayr, Ernst. 1963. *Animal species and evolution*. Cambridge, MA: Harvard University Press.

McDowell, John. 1978. Are moral requirements hypothetical imperatives? *Proceedings of the Aristotelian Society*, suppl. 52: 13–29.

Mele, Alfred. 2003. *Motivation and agency*. New York: Oxford University Press.

Miller, Christian. 2008. Motivational internalism. *Philosophical Studies* 139: 233–55.

Ridge, Michael. 2014. *Impassioned belief*. Oxford: Oxford University Press.

Scanlon, Thomas. 2007. Structural irrationality. In *Common Minds: Essays in Honor of Philip Pettit*, ed. Geoffrey Brennan, Robert Goodin, Frank Jackson and Michael Smith, 84–103. Oxford: Oxford University Press.

Scanlon, Thomas. 2014. *Being realistic about reasons*. Oxford: Oxford University Press.

Schroeder, Mark. 2008. Expression for expressivists. *Philosophy and Phenomenological Research* 76: 86–116.

Schroeder, Mark. 2010. *Noncognitivism in ethics*. London: Routledge.

Schroeter, Francois. 2005. Normative concepts and motivation. *Philosophers' Imprint* 5: 1–23.

Schulz, Armin. 2011. Sober & Wilson's evolutionary arguments for psychological altruism: a reassessment. *Biology and Philosophy* 26: 251–60.

Scott-Phillips, Thomas, T. Dickins and S. West. 2011. Evolutionary theory and the ultimate-proximate distinction in the human behavioural sciences. *Perspectives on Psychological Science* 6: 38–47.

Shafer-Landau, Russ. 1998. Moral judgment and moral motivation. *Philosophical Querterly* 48: 353–8.

Shafer-Landau, Russ. 2003. *Moral realism – a defence*. Oxford: Oxford University Press.

Smith, Michael. 1994. *The moral problem*. Oxford: Blackwell.

Smith, Michael. 1998. Ethics and the a priori: a modern parable. *Philosophical Studies* 9: 149–74.

Smith, Michael. 2004. Humean rationality. In *The Oxford Handbook of Rationality*, ed. Alfred Mele and Piers Rawling, 75–92. Oxford: Oxford University Press.

Sober, Elliott. 1994. Did evolution make us psychological egoists. In his *From a biological point of view – essays in biological philosophy*, 8–27. Cambridge: Cambridge University Press.

Sober, Elliott. 1999. Psychological egoism. In *The Blackwell guide to ethical theory*, ed. Hugh LaFollette, 129–48. Oxford: Blackwell.

Sober, Elliott and David Sloan Wilson. 1998. *Unto others – the evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Sterelny, Kim. 1993. Evolutionary explanations of human behavior. *Australasian Journal of Philosophy* 70: 156–73.

Sterelny, Kim. 2000. Looking after number one? *Biology and Philosophy* 15: 275–89.
Sterelny, Kim. 2003. Cooperation in a Complex World: The Role of Proximate Factors in Ultimate Explanations. *Biolological Theory* 7: 358–67.
Stevenson, Charles. 1937. The emotive meaning of ethical terms. *Mind* 47: 14–31.
Stich, Stephen. 2007. Evolution, altruism and cognitive architecture: a critique of Sober and Wilson's argument for psychological altruism. *Biology and Philosophy* 22: 267–81.
Stocker, Michael. 1979. Desiring the bad: an essay in moral psychology. *Journal of Philosophy* 76: 738–53.
Suikkanen, Jussi. 2018. Judgment internalism: an argument from self-knowledge. *Ethical Theory and Moral Practice* 21: 489–503.
Svavarsdottir, Sigrun. 1999. Moral cognitivism and motivation. *Philosophical Review* 108: 161–219.
Tresan, Jon. 2006. De dicto internalist cognitivism. *Noûs* 40: 143–65.
Toppinen, Teemu. 2015. Pure expressivism and motivational internalism. In *Motivational internalism*, ed. Gunnar Björnsson, Caj Strandberg, Ragnar Francén Olinder, John Eriksson and Fredrik Björklund, 150–66. Oxford: Oxford University Press.
Wedgwood, Ralph. 2007. *The nature of normativity*. Oxford, Oxford University Press.