INTERNALISTS BEWARE – WE MIGHT ALL BE AMORALISTS!

Gunnar Björnsson
Ragnar Francén Olinder

ABSTRACT

Standard motivational internalism is the claim that by a priori or conceptual necessity*,* a psychological state is a moral opinion only if it is suitably related to moral motivation. Many philosophers, the authors of this paper included, have assumed that this claim is supported by intuitions to the effect that amoralists – people not suitably related to such motivation – lack moral opinions proper. In this paper we argue that this assumption is mistaken, seeming plausible only because defenders of standard internalism have failed to consider the possibility that our own *actual* moral practice as a whole is one where moral opinions fail to motivate in the relevant way. To show this, we present a cynical hypothesis according to which the tendency for people to act in accordance with their moral opinions ultimately stems from a desire to appear moral. This hypothesis is most likely false, but we argue, on both intuitive and methodological grounds, that it is conceptually possible that it correctly describes our actual moral opinions. If correct, this refutes standard motivational internalism. Further, we propose an explanation of why many have seemingly internalist intuitions. Such intuitions, we argue, stem from the fact that standard amoralist cases allow (or even suggest) that we apprehend the putative moral opinions of amoralists as radically different from how we understand actual paradigmatic moral opinions. Given this, it is reasonable to understand them as not being moral opinions proper. However, since these intuitions rest on substantial a posteriori assumptions about actual moral opinions, they provide no substantial a priori constraints on theories of moral judgment.

Keywords: Motivational internalism, amoralists, moral motivation, intuitions

1. INTRODUCTION

Motivational internalism is often construed as the claim that *by a priori or conceptual necessity,* moral opinions are accompanied by motivation to act accordingly. As such, it is often taken to provide an important a priori constraint on theories of moral opinions. It is also commonly supported by intuitions to the effect that people who lack all motivation to act in accordance with their putative moral opinions do not really have those opinions: someone who cares not a whit whether he hurts others does not really think that hurting others is wrong. Not all philosophers feel the pull of such intuitions, but we, the authors of this paper, belong to those who do. Until considering the argument of this paper, we also, like most proponents of internalism, presumed that such intuitions are evidence for the conceptual claim.

By undermining this presumption, the argument presented here is intended to break a notorious stalemate in moral psychology. Externalists have denied that motivation is required for moral opinions and have presented what they have thought are intuitively possible cases of 'amoralism', i.e. moral opinions without corresponding motivation [e.g. Brink 1989; Svavarsdóttir 1999]. In return, internalists have relied on intuitions to the effect that in such cases of lacking motivation, moral opinions are also missing, at least if the motivation is lacking even under normal circumstances or in a whole community [e.g. Dreier 1990; Smith 1994; Tresan 2009b]. At this point, it has seemed to many that appeals to intuitions about amoralist cases cannot take this debate further [Francén 2010; Smith 1994; Svavarsdóttir 1999].

In this paper, however, we shall argue that they can, once we focus on the right sort of amoralist case. Doing so demonstrates that there is no a priori or conceptual requirement that moral opinions be accompanied by motivation and that the sort of intuition that we have ourselves taken to support conceptual internalism actually provides no such support. Closer scrutiny reveals that this sort of intuition stems, not from an appreciation of *conceptual* constraints, but from a sense that the accompaniment of the relevant kind of moral motivation is a central feature of *actual* moral opinions.

To show this, the sort of amoralist case we outline is a cynical hypothesis according to which *no actual moral opinion* is related to motivation in the way generally assumed by internalists, but instead affects our actions only through a desire to appear moral to ourselves and others. First, although the hypothesis is most likely false, it strikes us as at least conceptually possible, suggesting that as far as our concept of moral opinions go, genuine moral motivation is not conceptually required for moral opinions. Second, for the event that our intuitions about this case would not be representative for people with internalist inclinations, we argue that there are strong methodological reasons to take the cynical hypothesis to be at least conceptually possible. If there is a necessary connection between moral opinions and moral motivation, then, the necessity in question can at most be of an a posteriori variety, and cannot serve as an a priori constraint on metaethical theorising.

The cynical hypothesis is presented in section 3. In section 4, we argue that it is conceptually possible that the hypothesis correctly describes *our actual* moral opinions. In section 5, we diagnose the mistake behind conceptual internalism, arguing that it parallels the mistake Kripke ascribes to those who propose certain analytical truths about natural kinds. In section 6, finally, we briefly discuss the implications for other variants of motivational internalism. First, however, we prepare the ground by characterizing the target motivational internalism in more detail.

2. GENUINE CONCEPTUAL MOTIVATIONAL INTERNALISM

The target of our argument is *genuine*, *conceptual* motivational internalism. A form of motivational internalism is *conceptual* if it takes the connection between moral opinions and motivation to hold by a priori or conceptual necessity. Most philosophers who discuss the subject have conceptual internalism in mind. This is evident both from their concern with the *conceivability* of 'amoralists' – people having the opinion that φ-ing is wrong, say, but still totally lacking motivation not to φ [Brink 1989: ch. 3] – and from the fact that internalism is treated as an a priori constraint on theories of moral judgement [Smith 1994].

A form of motivational internalism is *genuine* if it not only requires motivation to act in accordance with one's moral opinions, but also requires that the motivation have an ultimately moral source. By contrast, consider:

> *Shallow internalism*: Our concept of what it is for X to take φ-ing to be morally wrong requires that X be motivated not to φ.

Since shallow internalism is silent about the source of the required motivation, it allows that someone thinks that various acts are wrong without ever being ultimately motivated by moral concerns to avoid such acts. All it takes is that she is motivated by some other source, such as selfish fear of censure and punishment.

Although this is seldom made explicit, we take it that philosophers who discuss internalism have had something more restrictive than shallow internalism in mind. Some explicitly require that, as the name 'internalism' suggests, the relevant motivation should be intrinsic to the moral judgment itself, as in familiar forms of both non-cognitivism and some anti-Humean forms of cognitivism (see also [Zangwill 2003; 2008] for an externalist who stresses this interpretation of internalism). Others thinks that it must stem from desires emerging directly from moral beliefs and reasoning in the absence of weakness of will and the like [Smith 1994], from processes required for the formation of moral beliefs [Radcliffe 2006][1], or from motivational attitudes that determine the truth-conditions of moral beliefs [Dreier 1990; Prinz 2006]. Admittedly, these sources of motivation form a motley crew, but what they have in common, and what makes them relevant for our argument, is that they contrast with purely egoistic motivation to appear moral. Any requirement that moral opinions are accompanied not only by motivation to act in accordance with the moral opinion, but by motivation ultimately

[1] Radcliffe argues that for Hume, motivation is guaranteed by a process of judging essential to moral beliefs, rather than from their intrinsic nature.

stemming from sources more closely related to the moral opinion than such a purely egoistic motivation, is an example of *genuine* internalism.[2]

We shall call the combination of conceptual and genuine internalism, 'standard internalism'. Standard internalism comes in different strengths along multiple dimensions, but we take our argument to undermine even the weakest. For clarity, we shall mention a few varieties, starting with:

> *Unconditional individual standard internalism*: By conceptual necessity, anyone who thinks that her φ-ing in C would be morally wrong is morally motivated not to φ in C.

Few today accept unconditional individual internalism. It seems that we can imagine people who have moral opinions without being motivated accordingly, at least if they are under the influence of depression, listlessness, acquired cynicism or some other special condition explaining the absence of motivation [Stocker 1979]. However, many still have the intuition that if motivation is missing absent such special conditions, the moral opinion must also be missing. They opt for some version of:

> *Conditional individual standard internalism*: By conceptual necessity, anyone who thinks that her φ-ing in C would be morally wrong is morally motivated not to φ in C *absent special conditions* [e.g. Smith 1994; Gibbard 1993: 318-319; van Roojen 2010].[3]

[2] Our argument also targets requirements weak enough to be satisfied if moral opinions are necessarily accompanied by motivation stemming from *final* desires to be moral or act morally. Some internalists will want to rule out the last option, e.g. due to the motivation being fetishist [Smith 1994], but others do not [Tresan 2006].

[3] To preserve the idea that internalism is a substantial constraint on how moral opinions are related to motivation to act accordingly, the class of special conditions must neither be so wide as to exclude most cases where moral opinions do give rise to motivation to act accordingly, nor defined in a way that ensures motivation even in conjunction with paradigmatically externalist accounts of moral opinions.

Others think that we can imagine a person who has moral opinions but is unmotivated even in absence of special psychological explanations as long as that person is part of a community with a moral practice where moral opinions in general motivate. What we cannot imagine is a whole community whose moral opinions fail to motivate. They opt for:

> *Communal (and conditional) standard internalism*: By conceptual necessity, anyone who thinks that her φ-ing in C would be morally wrong is morally motivated not to φ in C (absent special conditions), *or is a member of a community where most people are thus motivated* [Dreier 1990; Tresan 2009b].

In this way, internalists have retreated to weaker forms of internalism to accommodate scenarios with unmotivated people who intuitively have moral opinions. Our argument will suggest that these retreats are insufficient, as long as internalism is construed as both *conceptual* and *genuine*.

The force of the argument is also independent of how the connection between moral opinions and motivation is conceptually guaranteed. Traditionally, internalists have accounted for the necessary connection between moral opinions and motivation in terms of the mental states that constitute moral opinions. Motivation is guaranteed because moral opinions are constituted by desires [Ayer 1936: ch. 6; Gibbard 1990], besires [Dancy 1993; McDowell 1978], or beliefs with a special content [Smith 1994] or character [Dreier 1990], and thus by their very nature guaranteed to motivate (absent special conditions). But standard internalism is not necessarily wedded to such constitutional claims. According to one recent proposal by Jon Tresan, moral opinions are constituted by beliefs with contents that do not guarantee motivation, even under normal conditions. Instead, our *concepts of moral opinions* (concepts of thinking that something is, say, morally wrong, or morally obligatory) are such that beliefs with these contents *count* as moral opinions proper only when accompanied by relevant motivation [Tresan 2006; 2009a].[4] However, as long as such a non-constitutional thesis is a form of genuine internalism, requiring genuinely

[4] Related ideas are also hinted at in [Svavarsdóttir 1999: fn. 33; Jackson 1998: 161; Radcliffe 2006]

moral motivation, even this weak form of internalism is a target for our argument.[5] In fact, *every* version of standard internalism is challenged by the following new kind of amoralist scenario.

3. THE CYNICAL HYPOTHESIS

In this section, we present a cynical hypothesis about human moral motivation. Its core contention is that people have no specifically moral motivation to avoid wrongdoing. Their moral opinions, or what they take to be moral reasons and considerations, motivate neither on their own nor in combination with any specific disposition or desire to act morally. When people act in accordance with their moral opinions, then, this is always because they happen to think that acting thus is in their own interest.[6]

In later sections, we argue, first, that the cynical hypothesis represents an a priori, conceptual possibility and, second, that this undermines all forms of standard internalism. To make the first claim plausible, we need to give the hypothesis some detail.

We start by adopting and adapting a familiar evolutionary explanation of moral thinking: According to the cynical hypothesis, the human capacity to think in terms of what is morally right and wrong has developed from a need to track action types that promote or counteract cooperation and peaceful coexistence.[7] To help establish cooperative contexts, humans have also acquired a dislike for the wrongdoing of others, often driving us to blame, threaten, punish or ostracise actual and potential wrongdoers. However, contrary to a more familiar optimistic view, the evolutionary pressure has not instilled in us a direct motivation to act morally ourselves.

[5] Tresan's own view is a case in point, since it requires that there be moral attitudes, in contrast to egoistic desires, in a community for people there to have moral beliefs [2009b: 186].

[6] The cynical hypothesis is similar to the view that Plato's Glaucon motivates by the tale of the ring of Gyges: people have an interest in upholding their reputation as virtuous, but not in morality as such.

[7] Whether the *content* of moral opinions concern such effects on cooperation is a further issue, going beyond our specification of the cynical hypothesis.

Rather, what we have acquired, as a result of human dislike and punishing of immoral behaviour, is a desire *not to be seen* as morally bad and thus to *present ourselves* as acting on good moral reasons. Since we do this best if we believe the presentations ourselves, we have also acquired a very strong tendency to think that we do act for moral reasons.

According to this picture, people are driven by self-interest and limited altruistic concerns, mostly for their own children. In situations where we appreciate that moral considerations apply to our own actions, we will typically (but unwittingly) seek to do what is best for ourselves, where the social cost of appearing immoral is one consideration among many. How we handle such considerations depends on the circumstances. On the one hand, if we take moral and self-interested reasons to coincide, we might falsely present ourselves (to ourselves and others) as having acted partly because of the moral reasons. On the other hand, if there seems to be a conflict – if φ-ing seems self-beneficial but morally problematic – we tend to do one of two things, neither of which involves acting from moral reasons: *adjust the picture* or *adjust the action*.

In *adjusting the picture*, one subconsciously uses one of two different strategies to maintain the picture of oneself as moral while φ-ing. The first strategy is to stand by one's opinions that φ-ing is wrong, but to present oneself as *not* φ-ing. When we successfully apply this strategy, we do not in fact act in accordance with our moral opinions, but we appear to, both to others and to ourselves. In one famous experiment seemingly showing this sort of mechanism at work, Daniel Batson and colleagues [1997] had subjects assign two tasks, one rewarding and one boring, to themselves and another (actually fictitious) participant. Subjects were given the opportunity to use a fair coin in deciding, on their own, without observers, whom to give the positive task. Of those who did not use the coin, 90% assigned the positive task to themselves, but so did 90% of those who reported that they flipped the coin, indicating that many in the latter group chose to falsely present themselves as using the fair method.[8]

[8] Coin flippers also reported the way in which they assigned the task as *more moral* than those who openly chose to assign themselves to the positive task: 7.30 rather than 4.00 on a 1 to 9 scale [Batson 1997: 1341].

The second strategy is to find ways of not seeing φ-ing as wrong. One way to accomplish this is to make ad hoc reinterpretations of the reasons leading to the initial judgement. Since most situations are complex, there is often some moral reason for performing a given action that can be stressed at the expense of self-interested reasons. Another way is to create ad hoc hypotheses about relevant empirical facts. For example,

> those who desired to maintain slavery tried to justify its existence in various ways. There were the claims about the natural inferiority of enslaved races that are still made to this day. In the southern United States, there was an attempt to justify slavery as a form of paternalism. Involuntary labor was transformed into legitimate return for the protection and direction of masters. It was claimed that slaves acquiesced to their fate, that they willingly became part of the "family" of which the white master was the father [Wong 1984: 54].

When we apply the second strategy, we act in accordance with our moral opinions, but only because these opinions rationalize self-interested desires.

Sometimes adjusting the picture is psychologically unfeasible because there is no plausible rationalization at hand or because the agent has a feeling of being watched. When this happens and when the social costs of appearing immoral outweigh the benefits from φ-ing, we tend to choose not to φ, but to act in a way that will be easier to present as morally acceptable: we *adjust our actions* instead of how we represent ourselves and the situation. This strategy might lead us to act in accordance with our moral opinions, but only because of our desire not to appear immoral. In line with this, Batson et al. [1999] discovered that when people in the task assignment experiments were facing a mirror, forcing them to see themselves from the outside, only 50% percent of those who tossed the coin choose the positive task for themselves, and only 60% of those who did not toss a coin.

According to the cynical hypothesis, strategies or mechanisms like these preserve the impression that our motivation to act in accordance with our moral opinions have an ultimately moral source. They explain why we feel bad about doing something that we think is wrong even when we *seem* to face no risk of being socially exposed or punished. In typical cases, perception of such risks has in fact been triggered, often by thoughts about potential judges, not seldom thoughts about God.

It is of course true that some people seem to be moral heroes, taking exceptional measures to act in accordance with their moral convictions, even at the risk of severe punishment and without much to gain for themselves. According to the cynical hypothesis there are various explanations of such appearances. One explanation is that some people have exceptionally strong 'inner mirrors' or a strong sense of being watched by God, making them distressed in face of (perhaps illusory) potential condemnations and encouraged by potential approval and rewards. Another is that people sometimes underestimate or repress the risks involved for themselves, thus overestimating the relative gains from appearing moral. Some people even seek out risky situations and behaviour for the excitement involved, to feel alive. Yet another explanation is that, for ultimately non-moral reasons, many people are exceptionally committed to some specific causes – often because the commitment gives direction, stability and meaning to their lives – and for some people this happens to be a cause that they (and others) also take to be a moral cause. It is because of mechanisms of these sorts, often working in synergistic unison, that some people seem to be moral heroes even though they are not ultimately moved by moral concerns.

Much more could be said to fill out the cynical hypothesis, adding to its explanations of why it seems that humans are often driven by moral concerns. What we have here, however, should provide enough substance to let us consider whether it represents a conceptual possibility.

4. THE CONCEPTUAL POSSIBILITY OF THE CYNICAL HYPOTHESIS AND THE FAILURE OF STANDARD INTERNALISM

The cynical hypothesis is probably false. Although it might be able to explain our everyday experience of morality, we have no reason to believe it. For all we know, even people like the slave-owner are genuinely motivated to act morally, though often *more* motivated by other considerations, including the desire to at least appear moral. Moreover, the hypothesis seems evolutionarily and psychologically implausible: given the pressures working on the hypocrite, humans would probably quickly develop at least *some* degree of independent moral motivation of the limited sort suggested by common sense.

Nevertheless, we suggest that the hypothesis is *conceptually possible*. This claim can be divided into two parts, both concerned with our putative moral opinions, i.e. with those actual states of mind that we typically identify as moral opinions. The first is the claim that it is conceptually or a priori possible that putative moral opinions are subject to the massive hypocrisy and deceit postulated by the cynical hypothesis. This seems unassailable. Our pre-scientific views about this aspect of moral psychology could certainly be mistaken, just as folk psychology has been deeply mistaken elsewhere.

The second part is the claim that *if* the cynical hypothesis correctly describes the psychology of putative moral opinions, these opinions still count as moral opinions proper: people still do think that certain acts are morally right and others morally wrong. We find this claim intuitively very plausible. It is one thing to think that *some* putative moral opinions are not the real deal, but quite another to think that *all* the seemingly clear cases we encounter daily fail to be moral opinions – all the cases where people, including ourselves, seem to think that actions are morally wrong, or obligatory. The cynical hypothesis concerns the actual states of mind that we paradigmatically think of as moral opinions, and it allows that they have almost all the characteristics we normally ascribe to them. They are still categorical, based on familiar moral considerations (e.g. wellbeing, autonomy and respect for rights), often in competition with our prudential considerations, invoked to settle practical issues, and expressed to condemn behaviour near and far. Moreover, people are still affected by moral considerations, some more than others. What is different is just that moral opinions affect action *less directly* than most of us think. Given all this, it seems unreasonable to deny that these opinions are moral opinions.

Our first argument for the conceptual possibility of the cynical scenario is based on this intuition. Admittedly, this argument has a weakness: to have dialectical force, our intuition about this case must be representative for people with 'internalist' intuitions. We have no guarantee that it is representative, only the observation that our intuitions about standard cases of putative amoralists have seemed to support internalism.[9] But there are also strong *methodological* reasons to accept the conceptual possibility of the cynical hypothesis.

To see this, consider two concepts that people nowadays think lack extension: *witch* and *phlogiston*. When people revised their beliefs about the powers of individuals previously identified as witches and their beliefs about the processes previously thought to involve phlogiston, it was concluded that neither witches nor phlogiston exists. But why was it a reasonable response to new beliefs about powers and processes? Why not conclude that witches and phlogiston are different from what one had previously thought? Crucially, in both cases, too much of what motivated our interest in witches and phlogiston was rejected by our revised beliefs. Our core interest in attributing witchhood was to attribute magical powers; since we concluded that no one has these magical powers, including the people previously identified as witches, we came to deny that there are witches. Correspondingly, our core interest in phlogiston depended on its role in explaining phenomena related to combustion, a role crucially relying on the assumption that combustion involved the release of phlogiston. When we learned that combustion does not in general depend on the release of any substance, but rather on the binding of oxygen, we concluded that there is no phlogiston.

[9] There are, of course, more general concerns regarding intuitions as evidence about conceptual possibilities. These concerns cannot be discussed here, but they are to some extent addressed by the methodological argument below. It should also be noted that our argument is meant to carry weight in a debate where both supporters and critics of standard internalism appeal to intuitions about amoralist scenarios.

The case of moral opinions and the cynical hypothesis is different. Most of us might believe that moral opinions genuinely motivate in the sense questioned by the cynical hypothesis, but our everyday interest to keep track of what people think is morally wrong, obligatory, and so forth hardly relies on this belief. Much more likely, it stems primarily from the characteristic social role of moral opinions. This role, in turn, stems from the fact that opinions of this sort tend to be based on certain kinds of reasons, have a categorical form, and be accompanied by motivation, emotion and action.[10] All *these* features are left intact by the cynical hypothesis. Given our everyday interests in moral opinions, then, we would still have reason to keep track of what people think is morally wrong or obligatory (though we might also have reason for both theoretical and practical adjustments).

The cynical hypothesis would also leave intact the interests driving empirical studies of moral opinions. Such studies take various forms. For example, psychologists study how people arrive at and justify their moral opinions, and how moral opinions are affected by and affect emotional reactions under various circumstances. Philosophers, anthropologists, psychologists and biologists alike have proposed hypotheses about the evolutionary, social and psychological origins of moral thinking, moral opinions and moral motivation. Metaethicists, finally, have paid considerable attention to how moral opinions are actually expressed in natural language in an effort to determine what sort of states these opinions are. What is significant about the cynical hypothesis is that it does nothing to call into question the legitimacy of these inquiries or the integrity of their subject matter. Again, as noted, it leaves intact most of the features commonly

[10] Given a cognitivist interpretation of morality, our interest in moral opinions and their social role will also plausibly depend on their cognitive content. (The cynical hypothesis might rule out certain forms of internalist cognitivism that depend on the existence of genuine moral motivation.)

Notice that a correct characterization of our interest in moral opinions might depend on our best theory of moral motivation. What is relevant here is how our interest is best understood under the assumption that the account presented in the cynical hypothesis is correct.

associated with moral opinions, including the fact that they tend to be accompanied by motivation, and it does nothing to undermine the fact, surely motivating most of this research, that moral opinions play a central role in human cognition and social life.

All told, then, it would be unreasonable to respond to the findings outlined in the cynical hypothesis by retiring or fictionalizing our concepts of various kinds of moral opinions in the way we have retired or fictionalized concepts of witches or phlogiston. The cynical hypothesis should therefore be treated as a conceptual possibility.

If our intuitive and methodological assessment of the cynical hypothesis is correct, then, the cynical hypothesis represents a scenario where a whole community of people who lack genuinely moral motivation even under normal conditions nevertheless have moral opinions. This directly undermines standard internalism, whether unconditional or conditional, individual or communal, constitutional or non-constitutional.

This negative claim is also our main conclusion. In the two final sections, we shall first give a diagnosis of why internalists have thought that amoralists are conceptually impossible, and second, partly building on this diagnosis, say something about the implications for non-standard forms of internalism.

5. THE DEVIANCE FALLACY

Many philosophers, including the present authors, have felt the urge to deny the possibility of various amoralist scenarios, and to accept some form of standard internalism. Given the conceptual possibility of the cynical hypothesis, this calls for an explanation. Why have people felt this urge, and why have they taken this to support standard internalism?

Start with the question of why people have found various amoralist scenarios impossible. The answer, we suggest, is that they have unconsciously relied on the commonsense assumption that *actual* moral opinions normally are followed by genuine moral motivation. When various amoralists and amoralist communities are considered against this background assumption, they will seem significantly *deviant*: after all, they lack a feature central to our ordinary explanations of the practical psychological and social role of moral opinions, the role that makes moral opinions particularly interesting. This, we think, is why it has seemed intuitively reasonable to deny that the 'foreign' opinions are moral opinions proper. This would straightforwardly explain why we do not have corresponding 'internalist' intuitions when considering the cynical hypothesis that our *actual paradigmatic cases of moral opinions* do not provide genuinely moral motivation. In considering this hypothesis, we have to bracket our ordinary understanding of moral psychology, thus blocking the comparison.

In light of this, the reason why many of us have been tempted to accept some form of standard internalism seems to be that we have only considered amoralist scenarios allowing for the comparison with our ordinary understanding of moral psychology: scenarios containing either individual amoralists in our actual community [Brink 1989; Stocker 1979] or merely counterfactually possible or culturally isolated societies with amoralist people [Dreier 1990; Lenman 1999; Tresan 2009b]. Failing to recognize that the intuitions about such scenarios depended on the unconscious background assumption, we committed 'the deviance fallacy', mistaking an intuitive judgement of *deviance from our model of paradigmatic cases* for a judgement of *conceptual impossibility*.

To understand the nature of our objection to standard internalism, and the related diagnosis just given, it is helpful to see that it parallels Saul Kripke's criticism of alleged analytic truths about natural kinds [Kripke 1980: e.g. 120-21]. Suppose that it is claimed that as a matter of conceptual necessity, tigers cannot be reptiles. In support of this claim we are asked to imagine that, in some distant jungle, animals looking just like ordinary tigers are discovered, which on closer inspection turn out to be peculiar looking reptiles. Upon this remarkable discovery, we would intuitively conclude that these animals are not really tigers, but tiger-looking reptiles. Is this not evidence that, by conceptual necessity, if something is a tiger, it is not a reptile? No. For suppose instead that an even more surprising discovery is made: that all those animals that we actually identify as tigers are in fact peculiar looking reptiles. Previous biologists have been deceived by some illusion to think that they are feline mammals. Would we then say that there turned out to be no tigers after all (only tiger looking reptiles)? Kripke's point (one of them) is that we would not. We would say that tigers turned out to be reptiles. Consequently, our concept of tigers does not preclude that tigers are reptiles.

According to Kripke, the reason that we would say that the reptiles in the first scenario are not tigers is that '…they are not of the same species as the species which we called 'the species of tigers" [1980: 120]. If we put this in a more general form, we can say that the reason is that they are in crucial ways different from (of a different species than) the animals we have actually identified as 'tigers'. Similarly, we have argued that the internalist denial that the opinions of non-motivated individuals in usual amoralist scenarios are moral opinions stems from the appreciation that these opinions are significantly different from the actual states that we have understood as 'moral opinions' (different in not being accompanied by moral motivation). But when we imagine that the animals we have actually identified as tigers turn out to be reptiles, and the opinions we have actually identified as moral opinions turn out to be non-motivating, it is more appealing to conclude that we have been wrong about the nature of tigers and moral opinions than to deny their existence.[11]

[11] Relatedly, we suggest that there is an intuitive difference between the cynical scenario *considered as actual*, and the scenario *considered as counterfactual* [see e.g. Chalmers 2004: 159]. In this paper, we have considered it as actual – as a way that the actual world might turn out to be – and denied that it would mean that our actual putative moral opinions are not real moral opinions. To consider the cynical scenario as counterfactual, on the other hand, is to hold one's view of our actual moral practice fixed and think of the possibility that people are never genuinely motivated by their putative moral opinions as a way the world could have been but is not. Given the assumption that our actual moral opinions are opinions that paradigmatically motivate in a genuinely moral fashion, it seems quite intuitive to hold that, if instead we

6. WHAT NOW?

The question of this paper is one of the most debated in contemporary metaethics: are there significant conceptual constraints on how moral opinions relate to moral motivation? The argument we have given for a certain negative answer indentifies a previously overlooked type of amoralist case, involving all *actual* putative cases of moral opinion. We have used this case to make three points. First, to the extent that our intuition about the case is typical of people with 'internalist' intuitions, previously convinced internalists should intuitively recognize the possibility of moral opinions without moral motivation when faced with this example. Second, there are strong methodological reasons to accept the conceptual possibility of this sort of case: it fails to undermine some of our core interests in keeping track of various kinds of moral opinions. Third, there is a straightforward explanation of why self-professed internalists, including ourselves, have the intuitions they do about ordinary amoralist scenarios in spite of the falsehood of standard internalism: unlike the cynical hypothesis, these scenarios fail to undermine the assumption, common to internalists, that most actual moral opinions are accompanied by genuine moral motivation.

If we are right about this, there is no real stalemate between (standard) internalist and externalist intuitions: standard internalism is false, and the intuitions in its favour can be explained away. Moreover, since standard internalism is false, any theory of moral opinions or moral motivation implying it is equally false, and any metaethical argument relying on it – such as versions of the Humean motivational argument for non-cognitivism – is unsound.

would have had opinions of another kind that we cared to act on only when, and because, there was some personal gain in it, these would not have been moral opinions.

At the same time, however, we should be clear about the limits of the argument, as it leaves untouched *non-standard* variants of internalism.[12]

First, we have not ruled out conceptual but *non-genuine* forms of motivational internalism. For instance, our concepts of moral opinions might still require that such states belong to systems of opinions that often affect motivation – this is still true on the cynical hypothesis. However, since there is no requirement that that the motivation be genuinely moral, this does very little to constrain moral psychology and theories of moral judgment.

Second, and more importantly, we have not excluded genuine but *non-conceptual* forms of internalism. For example, for all we have said here, a non-cognitivism like Allan Gibbard's in *Wise Choices, Apt Feelings* [1990] or a sentimentalist account like Jesse Prinz's in *The Emotional Construction of Morals* [2007] might be true about actual moral opinions, providing the best overall account of moral motivation, reasoning, disagreement and metaphysics. Moreover, if we find out that some such theory is indeed true about our actual moral opinions, we might very well (in accordance with the diagnosis in section 5) conclude that possible or actual states of mind that lack the connection to motivation postulated by the theory are too different to qualify as moral opinions proper [Björnsson 2002]. The situation would be analogous to how many think about natural kinds: on the assumption that actual water is H_2O and the assumption that actual tigers are mammals, nothing counts as water unless it is H_2O and nothing counts as a tiger unless it is a mammal. If a 'metaphysical' necessity is one grounded in the (perhaps empirically discoverable) nature of that which it concerns, we would thus find ourselves accepting:

> *Metaphysical internalism*: By metaphysical necessity, anyone who thinks that her φ-ing in C is morally wrong is relevantly related to moral motivation not to φ in C.

[12] It should also be noted that our argument only concerns *moral* internalism. It is a further question whether it can be extended into an argument against (conceptual and genuine forms) of internalism concerning e.g., judgments about *what we must do all things-considered.*

So although it has been a mistake to take 'internalist' intuitions about common amoralist cases to justify conceptual internalism, such intuitions might well point in the direction of a metaphysical variant.

Shifting the debate about internalism from conceptual to metaphysical versions would have two significant and related consequences. The first is that versions of internalism could no longer be defended solely on basis of intuitions about cases. Instead, they would need to be grounded in empirically adequate accounts of moral opinions. The second is that genuine internalism could no longer be taken as an *a priori* constraint on metaethical theorizing. For example, if we conclude that moral opinions are necessarily intrinsically motivating, as Hume seems to have suggested, this would be the upshot of a theory of moral opinions rather than its starting point. But this, we now think, is as it should be.[13]

University of Gothenburg

Umeå University

[13] We are grateful to audiences at Stockholm University (especially Jens Johansson) and UC Riverside, to members of the MMER project at the University of Gothenburg, and the referees for the Australasian Journal of Philosophy for their comments and suggestions. This research was supported by the Swedish Research Council (grant number 2009–1517).

REFERENCES:

Ayer, A. J. 1936. *Language, Truth, and Logic*. London: Gollanc.

Batson, C. Daniel, Kobrynowicz, Diane, Dinnerstein, Jessica L., Kampf, Hannah C and Wilson, Aangela D. 1997. In a Very Different Voice: Unmasking Moral Hypocrisy. *Journal of Personality and Social Psychology* 72/6: 1335–48.

Batson, C. Daniel, Thompson, Elisabeth R., Seuferling, Greg, Whitney, Heather. and Strongman, Jon A. 1999. Moral Hypocrisy: Appearing Moral to Oneself Without Being So. *Journal of Personality and Social Psychology* 77/3: 525–37.

Björnsson, Gunnar 2002. How Emotivism Survives Immoralists, Irrationality, and Depression. *The Southern Journal of Philosophy* 40/3: 327–44.

Brink, David O. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Chalmers, David J. 2004. Epistemic Two-Dimensional Semantics. *Philosophical Studies* 118/1–2: 153–226.

Dancy, Jonathan 1993. *Moral reasons*. Oxford: Blackwell.

Dreier, James 1990. Internalism and Speaker Relativism. *Ethics* 101/1: 6–26.

Francén, Ragnar 2010. Moral motivation pluralism. *Journal of Ethics* 14/2: 117–48.

Gibbard, Allan 1990. *Wise choices, apt feelings: a theory of normative judgement*. Oxford: Clarendon Press.

Gibbard, Allan 1993 Reply to Sinnott-Armstrong. *Philosophical Studies* 69/2–3: 315–27.

Jackson, Frank 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

Kripke, Saul A. 1980. *Naming and necessity*. Cambridge: Harvard University Press.

Lenman, James 1999. The Externalist and the Amoralist. *Philosophia: Philosophical Quarterly of Israel* 27/3–4: 441–57.

McDowell, John 1978. Are Moral Judgments Hypothetical Imperatives? *Aristotelian Society Supplementary Volumes* 52: 13–29.

Prinz, Jesse 2006. The Emotional Basis of Moral Judgments. *Philosophical Explorations* 9/1: 19–43.

Prinz, Jesse 2007. *The emotional construction of morals*. Oxford: Oxford University Press.

Radcliffe, Elizabeth 2006. Moral internalism and moral cognitivism in Hume's metaethics. *Synthese* 152/3: 353–70.

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

Stocker, Michael 1979. Desiring the Bad: an Essay in Moral Psychology. *Journal of Philosophy* 76/12: 738–53.

Svavarsdóttir, Sigrun 1999. Moral Cognitivism and Motivation. *Philosophical Review* 108/2: 161–219.

Tresan, Jon 2006. De Dicto Internalist Cognitivism. *Noûs* 40/1: 143–65.

Tresan, Jon 2009a. Metaethical Internalism: Another Neglected Distinction. *Journal of Ethics* 13/1: 51–72.

Tresan, Jon 2009b. The Challenge of Communal Internalism. *The Journal of Value Inquiry* 43/2: 179–99.

van Roojen, Mark 2010. Moral Rationalism and Rational Amoralism. *Ethics* 120/3: 495–525.

Wong, David B. 1984. *Moral Relativity*. Berkeley; London: University of California Press.

Zangwill, Nick 2003. Externalist Moral Motivation. *American Philosophical Quarterly* 40/2: 143–54.

Zangwill, Nick 2008. The indifference argument. *Philosophical Studies* 138/1: 91–124