

# Believing Falsely Makes it So<sup>1</sup>

## 1 Introduction

In this paper I defend a cognitivist account of ethics. What is distinctive about this account is how I deal with the issue of moral internalism—the idea that there is some kind of necessary connexion between moral belief and moral motivation. Internalism has some plausibility, but it gets us into trouble when combined with cognitivism and another view that seems to be right—the Humean doctrine that belief and desire are distinct existences. If cognitivism and Humeanism<sup>2</sup> are both true, then we must have moral beliefs about the world that can be true or false, and they must be connected to desires by some sort of necessary or rational link. Showing that a belief can have all these features is not an easy task<sup>3</sup>.

My solution is to see what motivates internalism, and to see if there is something in this motivational base that can be retained, without retaining the whole of internalism. I then develop an account of ethics which retains this, but does not come to grief with the inconsistent triad of Humeanism, cognitivism and internalism, since the analysis—a cognitivist and Humean one—while it respects the intuitions that underlie internalism, is not in fact internalistic.

The case of internalism in ethics is one of a range of cases where it seems that there is a metaphysically problematic claim that is central to an important folk concept. Free will, qualia and personal identity are all topics in which there is mention of properties that seem *prima facie* to be crucial to the there being any free will, qualia or personal identity, and yet the world may not contain these properties. One solution is to simply deny that, on reflection, the problematic property is required to vindicate realism in the discourse—this is what compatibilists do in the case of free will. Another is to embrace an error

---

<sup>1</sup> Thanks to Daniel Cohen, Frank Jackson, Justine Kingsbury, Kristie Miller, Denis Robinson and Caroline West for valuable comments.

<sup>2</sup> Henceforth 'Humeanism' will always refer to the doctrine that beliefs and desires play different roles, rather than to Humean doctrines in any other area of philosophy.

<sup>3</sup> Michael Smith's *The Moral Problem* Oxford: Blackwell, 1994 is perhaps the *locus classicus* in which this task is undertaken.

theory for the discourse, and a third is to argue that the conceptual truths in question are of a conditional form, where the problematic property counts as essential if it is actually instantiated, but not otherwise.<sup>4</sup>

This paper offers another way of dealing with these cases: elevating *beliefs* about the metaphysically problematic properties to be among the truth conditions of the discourse. So I will argue that while internalism may well be false, ethical realism is nonetheless true. There are things we ought do: though part of what makes this so is that many of us believe in internalism.

### 1.1 Internalism

Internalism is the doctrine that there is a necessary connexion<sup>5</sup> between motivation and moral judgement. Internalism comes in many forms, but the kind that I take to be most persuasive and will discuss in this paper is something like Brink's weak hybrid internalism about motivation: a principle which tells us that it is a conceptual truth about morality that an agent who judges that she ought  $\emptyset$  will, insofar as she is rational, be *prima facie* motivated to  $\emptyset$ <sup>6</sup>.

Motivations for believing internalism vary. For some, a kind of rationalism is a motivational basis for internalism. But for others some form of ethical rationalism is the price for saving an intuition that has more to do with setting the subject matter of ethics. The thought is that without a connexion between motivation and moral belief, then any cognitivist account of ethics is open to

---

<sup>4</sup> See my 'Qualia and Analytical Conditionals' forthcoming in *The Journal of Philosophy*.

<sup>5</sup> 'Necessary connexion' covers a multitude of possibilities: some think the connexion is necessary because there is identity between a kind of belief and a kind of desire. Others that there is an anti-Humean necessary connexion between judgement and motivation, at least in rational agents. Perhaps the most common gloss, though, and the one that will be assumed here, is that there is supposed to be a necessary connexion between judging that one ought  $\emptyset$  and being ideally motivated to  $\emptyset$  in rational agents just because it follows from some kind of analytic truth about the meaning of 'ought' or of 'rational'.

<sup>6</sup> Brink, D. O. *Moral Realism and the Foundations of Ethics*, *Cambridge Studies in Philosophy*. Cambridge ; New York: Cambridge University Press, 1989, pp 37-41.

some version of the open question argument. Given some account of what objective state of the world is good (and therefore one should act to bring about), it is always possible to ask, it seems, whether it is *really* good.<sup>7</sup> This intuition, I think, disappears if there is some kind of necessary connexion between moral motivation and making the judgement that an objective state obtains. If we can say that a right action is one which, when you judge that it has the property of being right, then (perhaps ideally) you are motivated (perhaps in a particular way) to perform it, then that seems to be answer enough to the open question argument. This is because what leaves a gap for the open question to squeeze in, is the idea that until we feel some kind of (even ideal) motivational force behind someone's judgement that some action deserves to be called 'right', it can still seem opaque whether that property is correctly considered a moral one. That's why I described it as subject setting: the subject of ethics is about certain actions which can be described, but until some kind of connexion is drawn between the actions and motivation, we can't be sure that they are actions that deserve to be described as right or wrong.

## 1.2 The inconsistent triad

So cognitivism, Humeanism and internalism jointly pose a problem. If Humeanism and cognitivism are true, then internalism is in trouble: how can judgements about truth makers that aren't desires be intrinsically motivating? If internalism and Humeanism are true, then only if the function of discourse is not fact stating—i.e. some kind of non-cognitivism is true—will consistency be restored. If internalism and cognitivism are true, then there will be no problem if Humeanism is not, for the cognitive judgements may themselves be a kind of desire and hence motivating.

Let us now focus on internalism. If there is a necessary connexion between cognitive judgements that certain actions are right, and motivation, then how is that connexion maintained? The judgement is a kind of belief, but the motivation is a kind of desire.

One approach to the inconsistent triad would be to abandon strict internalism and rely on contingent psychological facts about our motivational

---

<sup>7</sup> Moore, G. E. *Principia Ethica*. Cambridge: Cambridge University Press, 1922.

profiles to account for internalist intuitions. It might be that we just are creatures that often desire to perform actions we call 'right'. But that doesn't give us what internalism seems to demand—something like an answer to why we *should* be so motivated. It just tells us that we are in fact so motivated, and that seems to leave open the 'is it really right?' question, returning in the guise of 'should we really desire to perform the actions that we call 'right'?' So contingent connexions between desire and belief don't seem to account for the internalist intuitions.

Alternatively, returning to strict internalism, we could hope that we can show that some beliefs are in fact desires: we could deny the Humean thesis that beliefs and desires are distinct existences. This would give us a necessary connexion between belief and desire of the tightest kind—identity—but attempts to make this plausible have been notably unsuccessful, and there are persuasive arguments that it is logically impossible.<sup>8</sup>

Alternatively we could appeal to some kind of subjectivism: simple subjectivism simply gives the content of the judgement as a claim about motivation, thus guaranteeing that motivation to  $\emptyset$  will exist when someone correctly judges that they ought  $\emptyset$ . Views like Michael Smith's<sup>9</sup> can be thought of as sophistications of subjectivism, which get the necessary connexion via analytic means. Finally we could of course deny cognitivism. Any theory of ethics which denies that moral claims are truth apt but are mere, for example, expressions of desire, will have a kind of connexion with desire that is very close, but that is, at least for me, a move of last resort.

So we seem to be in trouble if we think each of cognitivism, Humeanism and internalism is true. Trying to show how these features can be reconciled is what Smith famously calls the Moral Problem<sup>10</sup>, and which he tries to solve in his book of the same name. The triad is not of course inconsistent, since after all Smith offers a solution, and he does not mean by that making the inconsistent consistent! One approach would be to argue that Smith's aim is to make what seems inconsistent prove to be, on reflection, consistent. But there is something

---

<sup>8</sup> Lewis, D. 'Desire as Belief.' *Mind* 97 (1988): 323-332., Lewis, D. 'Desire as Belief II.' *Mind* 105 (1996): 303-313.

<sup>9</sup> Smith, M. *The Moral Problem*. Oxford: Blackwell, 1994.

<sup>10</sup> *ibid*

more than mere *prima facie* tension here. Perhaps it is this: there is an inconsistent *tetrad* if we add 'denial of substantive reason about desire revision' to the list. Internalism, cognitivism, Humeanism and the *denial* of substantive reason about desire revision<sup>11</sup> are jointly inconsistent. I of course embrace this last principle; whereas Smith accepts substantive reason about desire revision (or at least rejects Humeanism about normative reason, which is near enough equivalent).

My strategy here is to almost solve the moral problem. The problem won't be solved, because what I reconcile won't be cognitivism, Humeanism and internalism. Instead I will be reconciling a denial of substantive reason about desire revision with cognitivism, Humeanism and something that does the work that makes internalism plausible, without actually being internalism.

## 2 Tracking naturalism and Smith's convergence

At this point I want to introduce two general approaches to cognitivist ethical theories: an approach I call *tracking naturalism*, and Michael Smith's brand of rationalist cognitivism. I discuss these doctrines the better to locate the view I will present, which will sit between them.

Tracking naturalism, roughly, is a version of cognitivist ethics that settles the moral properties by determining which properties our use of ethical terms in fact track. The idea is to locate ethical terms in an implicit folk theory, and the moral properties are the properties in the world that realize the theory, and are thus being tracked by the ethical language.<sup>12</sup> Usually the theory is not taken to include only what is explicitly believed or regarded as platitudinous, but rather is constituted by the complete sets of behaviours and interactions with the world and with moral concepts. And usually it is idealized: it depends on how the theory would develop under improved information. It can come in various versions. There are versions that identify the right with what we would call

---

<sup>11</sup> This is close to what is sometimes called Humeanism about normative reason.

<sup>12</sup> Examples of views that share some of the features of what I call tracking naturalism include: Jackson, F. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press, 1998., and Jackson, F, and P. Pettit. 'Moral Functionalism and Moral Motivation.' *Philosophical Quarterly* 45 (1995): 20-40.

'right' ideally. Tracking naturalism also comes in versions in which 'right' labels actions that we may not be motivated towards,, but rather would be so motivated on reflection or equilibration (stripped of the rationalist thought that there are rational constraints on how our desires would evolve under such equilibration).<sup>13</sup> Such a doctrine is cognitivist—claims that in possession of more information we will have certain beliefs or desires are clearly truth apt. And it is Humean, for the belief that in conditions of more information we will have these beliefs or desires is plainly a pure belief (even if it is a pure belief about a desire). The problem is in the internalist constraint. What necessary motivational pull is there in judging that, in states of improved information, we will have certain beliefs or desires? In the case of beliefs, what we will even ideally believe has no connexion to desires (except to instrumental desires where the future beliefs tell us what would be a good instrument to achieve our basic desires). In the case of desires, what we would desire in the possession of more information is equally unmotivating, at least *prima facie*<sup>14</sup>. For this merely tells us that in other circumstances we would desire differently, and for a Humean there is no reason to change our current desires just because we would desire differently if we were in the possession of more information (once again, excluding the instrumental case). It might just be that more information will causally affect our desire profile.

Michael Smith's solution adds to tracking naturalism some constraints of reason. The idea is that there is a substantive account of rational change of desire to be had, by analogy with accounts of rational change of belief. What we cognitively track is not just what we would desire in possession of more information or after arguing, but what we would desire after *rational desire change*.<sup>15</sup> And just as we might now be motivated to believe something if we were told that it's what we would rationally come to believe in the light of evidence (and believed this), we might now be motivated to desire something if

---

<sup>13</sup> Jackson, F. *op cit*

<sup>14</sup> *ibid*

<sup>15</sup> This is a simplification for expository convenience. The actual view is that what we ought do is what, on rational convergence, we would desire ourselves to desire in our current circumstances.

we knew that we would rationally come to desire it after argument, equilibration, better information and so on.

This certainly satisfies the internalist constraint. But the price is the substantive account of rational desire change that must be provided. There is nothing on offer that promises anything like the success of Bayesianism, say, for belief change. And in addition there is a worry about a certain circularity. Absent a convincing independent account of desire change, all that we are left with is the doctrine that what we ought do is what we would, on rational convergence, all come to desire, and that rational desire change is that process by which, if agents follow it, they will all come to desire to perform right actions.

So what would be ideal for the cognitivist would be to find something we can add to tracking naturalism which will satisfy enough of the intuitions that ground internalism, without invoking something as strong but implausible as substantive reason about desire.

### **3 What underlies the internalist intuition?**

Why do we think that there is some kind of connexion between moral concepts and motivation? Here is a proposal: it's because we suppose that a world in which although people use the words 'right', 'wrong' and 'ought' in a counterfactually robust way of the same descriptive states (including actions) that we do, but in which there is not even the slightest motivational tug towards what they call 'right', that these agents do not share our concepts of right, wrong, and so on. I'm supposing here that these agents not only are not motivated, but they also don't have any second order desires for motivation, are not made uneasy by not being motivated and crucially do not believe in any connexion between motivation and beliefs about what is right. Let us call such a world a Tracking Naturalist Equivalent world.

Quite how such agents could succeed in tracking these properties is of course mysterious. The problem is that probably we track certain naturalistic moral properties by some kind of internal simulation that tells us what we would desire after equilibration. In addition, most of us believe that we should be motivated by what would motivate us in those circumstances. The latter is by stipulation not true on the TNE world, so perhaps their desire profiles are somewhat different from ours. So how can they be tracking the same

properties as we are? The mechanics of this don't matter: perhaps they have an introspectively inaccessible module which predicts what we would desire, or perhaps there is enough in common between the states that we would ideally desire that they can have independent access to the naturalistic properties. The key thought we will explore later is that they can be tracking the same properties without sharing the same concepts.

Agents in such a world have some kind of useless dongle in their psychology: the capacity to classify states of the world in a certain way that is for them quite arbitrary though repeatable. Or perhaps the terms for them are phenomenal property terms: suppose that actions called 'right' are detected because their sensory apparatus generates a strange phenomenal colour cast when they are contemplated. Then it seems that 'right' is for them a phenomenal term, not a moral term. This is why pure tracking naturalism will not fly: according to tracking naturalism these agents have our moral concepts, but this is surely a mistake. Agents in the TNE world do not understand the point of ethical categorization; it does not play the social role that ethical categorization plays for us, and their answer to the open question argument either reveals that there is no answer 'dunno, I'm just disposed to use the term 'right' here' or else reveals that the concept is really not ethical at all 'Oh - that active deserves to be called 'right' because it—or the contemplation of it—makes me experience a distinctive blue halo'.

What is missing from their moral concept? They do not understand what moral concepts are for: they do not understand the role they play in individual and social deliberation. Nor is there any connexion between these roles and their judgements<sup>16</sup>.

What exactly are the roles moral concepts play in *us*? This is a difficult issue in the substantive analysis of our moral concepts, about which I have more to

---

<sup>16</sup> Elsewhere I give two dimensional analyses of concepts where one dimension is given in terms of the social role. This is developed in the case of personal identity in Braddon-Mitchell, D and West, C 'Temporal Phase Pluralism' *Philosophy and Phenomenological Research* 2001 pp 62, 1-25. See also the remarks of Michael Dummett in Dummett, M 'Truth' in *Truth and Other Enigmas* Duckworth, London 1978 pp 1-24.

say elsewhere.<sup>17</sup> For our current purposes we need only an illustrative placeholder; but I choose one of signal importance. Moral concepts are thought to be necessarily connected to motivation. It is the internalist intuition that is missing. So what if we were to add such a requirement to our tracking naturalism: suppose we had a two part theory which says that *A* is right iff

- (a) *A* has the properties that the tracking naturalist uncovers
- (b) Judging that *A* is right is necessarily connected to motivation

This however, leaves us open to the error theory. For very likely nothing is right, so construed. Clause (b) may very likely fail, since it far from clear that there are any judgements that *prima facie* count as moral and which are necessarily connected to motivation. Those, like me, who doubt there are any such judgements do not take it that moral nihilism is the inevitable result. So can we explain the failure of these agents to share our moral concepts without full-blown internalism?

Here is another attempt—one in the direction I urge.

(Belief about motivation tracking naturalism: BAM TN')

*A* is right iff

- (a') *A* has the properties that the tracking naturalist uncovers.
- (b') Judging that those properties are exemplified by *A* is believed to be necessarily connected to motivation.

Now we are saved from the clear and present danger of the error theory. For even if there is no necessary connexion between moral judgement and moral motivation, actions still get to be right provided that they are *believed* to possess the relevant connexions, and of course provided that the tracking naturalist conditions are also met. Thus the view is straightforwardly cognitivist. Of course it is not strictly internalist, for there is no requirement that there actually be a necessary connexion between moral motivation and moral properties. For these beliefs could in fact be false. There are two things to notice, though. The first is that on BAM TN' something *like* weak internalism is true: one cannot judge that an action is right without *judging* that there is a necessary connexion with motivation. If this means that one is in Brink's sense not 'completely unmoved' then BAM TN' is indeed a kind of weak internalism about moral judgement. But even if it is not a kind of internalism, it does explain in a

---

<sup>17</sup> See Braddon-Mitchell, D. *Concepts and Conceptual Change*, MS in prep.

relatively principled way the connexion between actual motivation and moral judgement. If one believes that there is a necessary connexion between judging that *A* is right, and being motivated to *A*, and one does judge that *A* is right, it's hardly surprising that one would feel uneasy or indeed irrational about a lack of motivation towards *A*. Of course the connexion between a belief in necessary connexions between judgement and motivation on the one hand, and actual motivation or unease at lack of it on the other, is contingent. But it would not *seem* contingent, and so explains the overall phenomenology. This is how a doctrine, without embracing internalism, can explain the data that internalism explains.

### 3.1 The Theorist

There is an obvious objection here. Suppose that as a moral theorist I come to believe in the incoherence of internalism. I read all the literature, I conclude that internalism can be vindicated only if either the belief as desire hypothesis is correct, or there is a substantive theory of rational desire change. I am persuaded that neither of these is true. So I am persuaded that there are no necessary connexions between moral judgement and motivation. It is thus not possible for me to reasonably judge that any state *A* both meets the TN condition, and to hold that judging this is connected to motivation in a necessary way.

If moral concepts were understood in an idiolectic way—for each individual, truth conditions of 'right' in their mouth are that the (more social) TN condition obtains and *they* judge that the necessary connexion holds, then the theorist, on concluding their line of reasoning, suddenly fails to judge that anything is right. They will of course have to grant, since the theorist I am supposing, qua theorist, believes in the BAM TN' analysis, that 'right' in other people's mouths determinately picks out something. But the discovery of the metaphysical and conceptual truths instantly debar the theorist from making moral judgements.

The idiolectic version is then relativistic. Not because of the TN component—though plausible versions of it are I think relativistic,<sup>18</sup> but

---

<sup>18</sup> One important difference between the naturalistic stories about moral properties that Smith and Jackson tell is that for Smith relativism is ruled out *a*

because of the second component. The extremely relativistic conclusion here is, I think, implausible because of the strange effect that the metaphysical discovery has on the theorist. What it directs our attention to is the question of *who* should have the beliefs about the truth of internalism.

One obvious possibility is to require complete convergence. The most straightforward version of this would be:

BAM TN''

A is right iff

(a'') A has the properties that the tracking naturalist uncovers.

(b'') Everyone now judges that the internalist relation holds, or everyone would hold ideally that it held.

This takes us into the dark embrace of nihilism all too quickly. For everyone does not now believe in these necessary connexions. So there are no actions which both meet the TN condition, and of which it is believed by everyone that if it is judged to fall under <right> there is a connexion with motivation that is necessary. So nothing is right. And even if it were the case that everyone did hold the beliefs in (b''), the first theorist to be persuaded (in a deep way— not merely verbally) by an externalist polemic would cancel out the rightness of everything.

Perhaps we can allow that convergence at the meta-level may fix this: actions are right only if we would converge on the judgement about the connexion between ethical judgement and motivation. This is not going to persuade the likes of me, however, because after all I take the judgement to be *false*. It is hardly likely that improved argument etc will make us converge on a falsehood.

Here is what I take to be the right version. We require the tracking naturalist story, plus the requirement that we most of us believe that there is a kind of necessary connexion.

BAM TN'''

A is right iff

---

*priori* by requiring complete convergence as a necessary condition for the existence of the good (for him if we converge on different points this is the vindication of the error theory) whereas there is no such *a priori* constraint for Jackson.

(a'') *A* has the properties that the tracking naturalist uncovers

(b'') Judging that those properties are exemplified by *A* is generally and with strong behavioural and social consequences believed to be necessarily connected to motivation.

On this reading, the theorist can come to see the falsity of the view that there is a connexion, whilst still judging that actions are right, and be saying something true. The theorist will merely come to see that there is something response dependent about the idea of the right. She will come to judge that there is a crucial connexion between its being true that some actions are right and some false metaphysical beliefs that are widespread. I at least have stopped finding this counter-intuitive. It satisfies the intuition that underlies internalism, but it also satisfies the conviction that many of us have that there is a darkness in the heart of meta-ethics—that if we look closely at moral concepts we'll find that they depend on metaphysical error. I know many philosophers who have avoided ethics not because they think it is unimportant, but because they fear that on reflection they would become error theorists. But on my account there is a diagnosis of the metaphysical error we will find, and an explanation of why it does not lead to the error theory.

The theorist will not herself cause any kind of moral collapse. But notice that if the theorist were persuasive enough—if vast numbers of people were indeed persuaded of the falsity of the connexion between judgement and motivation, and that persuasion was more than merely verbal, then it really would be the case that the world would be stripped of its moral properties.

Notice that there appears to be something vague about phrases in (b'') such as 'most of us' and 'generally'. Absent an appeal to epistemic vagueness according to which there might be some objective cut off point which is forever hidden from enquiry,<sup>19</sup> I agree that making it precise could only seem arbitrary: how could the difference between 71.1% of the population having the beliefs mentioned in (b'') and 71.11% having them, ever plausibly make the difference between there being right actions in the world, and its being a moral wasteland? But vagueness, it has been objected, might seem worse. For then the persuasive nihilist who gradually removes the beliefs of (b'') from the

---

<sup>19</sup> Williamson, T. *Vagueness, The Problems of Philosophy : Their Past and Present*. London ; New York: Routledge, 1994.

population will eventually create a state of affairs where there is no fact of the matter as to whether there is rightness in the world.

My response to this is just to endorse this as a desirable consequence of the view. The view is, after all, a kind of response dependent realism. Among the things that have to be true of an action for it to have the property of rightness are some epistemic relations of a complex socially mediated kind. And these are exactly the sort of things which, while they have clear cases, have vague borders. Just as there can be borderline cases of there being language in the world if the only communicating species are fairly simple in their methods, there can be borderline cases of morality. Surely ethics can be expected to be at least as hazy at the edges as linguistics.

#### **4 An Objection**

Suppose that in fact, as many of us suspect, there are no necessary connexions between judgement and desire. On my account there will still be moral properties, so long as such connexions are believed to hold. But suppose all of us become theorists whose eyes are opened to the metaphysical truths and we lose our belief in these connexions between desire and judgement. Suppose also that this theoretical belief is psychologically isolated; it does not change how we interact with each other. It is one which we all become prepared to assert in philosophy examinations and is a commonplace in talk about morality, but our habit of changing our desires when we see that something possesses the naturalistic properties, and our expectations about other persons' desires all remain as they are.

It may seem that my account allows a change in theoretical belief to strip the world of its moral nature all too readily. For something is right iff it has the properties that we track, and (most of us) have the relevant beliefs about judgement and desire. Since our metaphysical discovery we none of us have these beliefs. Thus there is nothing of which the second clause of the account is true, and so no actions are right. But surely, runs the objection, something as superficial as these theoretical beliefs shifting cannot erase moral nature from the universe.

I agree, but the keen reader will notice that I will somewhat tiresomely slip in '(tacit)' or '(implicit)' before 'belief' in much of this paper, and often talk of a merely verbal change in belief. This is because what I have in mind is a

functional picture of belief according to which beliefs supervene on the whole range of dispositions that agents have to interact with the world, not only the relatively superficial disposition to assert sincerely or some such<sup>20</sup>. An agent whose connexions between desire and judgement are (contingently) just like ours, and whose counterfactual judgements about their desires in other circumstances are the same, and who engages in practices of reasoned debate about what they should do, is someone who tacitly believes that there is a rational connexion between desire and judgement, even if there is no such connexion and they explicitly avow that there is no such connexion. Thus the merely verbal theoretical change is not enough to make it the case that there is a defeat of clause (b''') and thus that nihilism is true. What it would take to make nihilism true would be for us to lose even the tacit belief in the connexion, which would require us to give up the practice of reasoned deliberation about ethics, and cease to have any pull towards the actions that we judge to be right. And that change does sound like a change to a world where nihilism is true.

Of course some will be sceptical that losing tacit belief in such connexions will result in any change to our practices whatsoever. Externalists, for example, claim already to believe that there is no such connexion, and that most people do not believe there is such a connexion. For them, belief in such connexions is not widespread, and it is false. It is a piece of false doctrine held exclusively by some theorists—in particular by philosophers who are internalists about motivation.

If they are right, then of course what I say is wrong. My view is not supposed to persuade anyone who thinks that there is a prior, convincing externalist picture of ethics. My view is a kind of externalism for those who believe that there is something importantly right about the internalist intuition.

---

<sup>20</sup> This same fact explains how the analysis I offer escapes the so-called paradox of analysis. The idea is that if a theory can explain the pattern of use of an agent, even if there is no explicit, consciously accessible representation of the theory available to the agent, it can be attributed as a tacit belief. One job of analysis is to make explicit these tacit beliefs. So what looks like theory building by something like the hypothetico-deductive method is in fact finding out what the implicit analysis must be via inference to the best explanation (see my *Folk Theories of the Third Kind*, forthcoming).

So it is a starting point of this paper that there is no straightforward externalist story about ethics which is both realist and doesn't unrecognisably alter our understanding of rightness. The task of arguing for the inadequacy of simple externalism has been done elsewhere. So for our purposes, our reactions of unease at our failure of motivation when forming moral judgments is best explained by tacit belief in some kind of necessary connexion. I look forward to new alternative explanations.

#### **4.1 A Further Objection**

This account makes it a truth condition for a state of affair's being right that certain beliefs are held about a rational or necessary connexion between judgement and desire. But I also believe that probably there are no such connexions. And if there are no necessary or rational connexions, then plausibly there are necessarily no such connexions. So the required belief is in a necessarily false proposition. But on some accounts there is only one necessarily false proposition, on others there is no such proposition. In general there is a problem with belief in things that are necessarily false.

But everyone needs a solution to this. Perhaps it is metasemantic, maybe it's belief that 'There is a necessary connexion....' expresses a truth. But whatever the solution that you prefer to how I could have believed falsely that Fermat's last theorem was not a theorem, it can be deployed here.

### **5 The tracking naturalist equivalent world**

Let us consider more carefully the tracking naturalist equivalent world of section 3. This is a world where no-one in fact has any motivational connexion to those actions they label 'right' (or more realistically, their rankings of actions in terms of what they call 'right'). There are two important questions to ask. First, we need to know what to say about the concept  $\langle \text{right}_{\text{TNE}} \rangle$  that these agents possess. Second, we need to know whether any actions are in fact right in this world.

#### **5.1 What is the concept $\langle \text{right} \rangle$ in the TNE world?**

On fairly minimal assumptions about common knowledge, most agents on the TNE world will know that most of them do not have such motivational connexions. They know there is in fact no connexion between judgement of goodness and motivation. Since  $\neg p \supset \neg Lp$  then (once more assuming some

rationality) they know that there is no necessary connexion. It is reasonable to assume that in addition they know it is common knowledge that there is no necessary connexion, and thus that most people do not believe that there is a necessary connexion between judging that an action is right and being motivated towards it. But by hypothesis they do believe that there are right (or comparatively right) actions. So for them the concept  $\langle \text{right}_{\text{tne}} \rangle$  cannot include as a necessary condition for  $\text{rightness}_{\text{tne}}$  obtaining:

b''') Judging that those properties are exemplified by  $A$  is generally and with strong behavioural and social consequences believed to be necessarily connected to motivation.

In which case, it follows that their concept is not the same as ours. This is the desired result. It is the result that a world full of trackers of naturalistically the same actions, but without the beliefs that most of us have about moral motivation, contains agents who are conceptually deficient with respect to our concept  $\langle \text{right} \rangle$ .

## 5.2 What actions are right in the TNE world?

So much for their concept; but are any actions *in fact* right in such a world? This question depends of course on whether we mean 'right' in the sense it has in their mouths, or the sense it has in our mouths. Plenty of things fall under 'right' in their sense—indeed naturalistically similar actions to those we judge to be right—but since this is a different concept from ours this is unexciting.

What about the sense which 'right' has in our mouths? Should we judge that actions are right or wrong in the TNE world, and which actions should we judge to be right or wrong?

This depends on how we regard such a world. If we consider the world as counterfactual, we should regard it as a world full of rightness and wrongness, but where the locals don't possess the concepts  $\langle \text{right} \rangle$  and  $\langle \text{wrong} \rangle$ , whilst still being applying the same words—'right' and 'wrong'—to the same actions. This would be on the assumption that the beliefs that most of *us* have about the connexion between moral judgement and motivation are the right making properties, even for other worlds.

Suppose we regard the world as *actual*, in the sense of this that comes from two dimensional semantics.<sup>21</sup> In that case we are settling the issue in the same way as we would if we were to consider the world as a way we might discover the actual world to be. So suppose we discover the actual world to be one where none of us has ever been in fact (contingently) motivated by our judgements of what actions fall under <right>. Would this be the discovery that, implicitly, we never really had the *requirement* that our judgements be so motivating? In other words, would it be the discovery that the analysis I am advocating is wrong? In that case, it would turn out that there are right actions, for all that would be required would be clause (a); for on this concept of right, pure tracking naturalism is correct. Or would our conviction remain that such beliefs are required, and thus the discovery be one that the error theory is true (or at least becomes true when the word spreads, much as in the case where the theorist persuades the population at large of the motivational irrelevance of ethics) and thus nothing is right? There are important issues to be settled here, but not ones that must be settled for our current purposes.<sup>22</sup>

### 5.3 The actual world theorist and the TNE theorist

It is useful at this point to draw a contrast between the individual theorist of section 3.1 and a theorist from the TNE world. It may seem as though a theorist whose has come to believe that there are no necessary connexions between moral judgement and motivation is rather like someone in the TNE world.

---

<sup>21</sup> See for example Chalmers, D. 'The foundations of two dimensional semantics' forthcoming.

<sup>22</sup> It might be, for example, that a more baroque version of the theory could be true. Moral realism might not require even that anyone actually believe that there are necessary connexions between certain judgements and motivation, but it require only that enough people believe (perhaps falsely) that *other people* believe (perhaps falsely) that there are necessary connexions between certain judgements and motivation. Such strange indirect beliefs often explain social phenomena, such as when no economist believes that a certain economic indicator is significant, but enough economists believe that other economists (falsely) believe that the indicator is significant for it to have an effect on the economy.

They both label the same actions 'good' but neither possesses the motivations and the beliefs about them that we have.

But there is a crucial difference. Suppose that we bring a TNE agent into the actual world. Suppose further that a theorist in the actual world and the TNE agent both point at some actions (perhaps lobbying for many bike paths) and label them right, and do so on the basis of the same detection mechanism. Do they share the same concept? They both judge the same actions to be right and wrong, their detection mechanisms are similar or the same, and neither believes in any necessary connexions between judgement and motivation.

However, while neither of them is motivated towards right action, the theorist only believes actions to be right if she holds that most others believe that there is a rational requirement to be motivated by their judgements that actions are right. She shares our concept and thus her judgements depend on her background assumption that most other agents have the (by her lights) false belief that there is a necessary connexion of the right kind. For her, if it were the case that all actual agents lost that belief, then no actions would be right. The TNE agent would be surprised at this. He would still regard the lobbyings as right, since for him satisfying (a'') alone is sufficient for an action's deserving to be called 'right'. Agents' beliefs about the connexions between motivation and desire are irrelevant.

## 6 Some issues about relativism

What if there is a lack of unity in moral motivation? In the above case these differences are purely theoretical in a way which abstracts from the likelihood of any real moral significance. This is because we agree that the agents with no (even tacit, behaviourally recognizable) beliefs about the connexion between judgement and desire have an entirely different concept from us. So the issue is just the issue of whether *our* concept is one according to which what is rightmaking in that world is, *inter alia*, the beliefs about motivation of the agents in that world, or our beliefs.

In the actual world, though, there is always the possibility that there will be differences not in (b''') of the account, but in (a'''). In other words it strikes many of us as likely that there may be differences in the tracking naturalist component: in the natural properties that we do in fact track. Those who think that the convergence on what we track is mediated by rational changes in

desire can hope that caveat may mean that complete convergence is inevitable. I doubt even that: processes of improvement can often end up with multiple convergences at local maxima. We need only think of optimising models of natural selection that still result in populations stranded at local maxima, because at the local maximum there is no small change that is an improvement, even though there are logically possible but nomologically vanishingly unlikely, large changes that would be improvements. Similarly in the case of rational change of desire; there is nothing in the idea of an objective account of rational change that guarantees that rational changes from different initial positions will converge on the same place, even if the process of rational improvement is the same. Much less can we suppose that we will have this kind of convergence if we cannot appeal to substantive rationality.

If two populations converge on different points, and one of them lacks the beliefs required by clause (b''') - i.e. they do not have and never have had any (tacit) beliefs about the connexions between desire and judgement, then on my account this is not a case of relativism. Even though the populations may say different things: one says 'A is right' and the other says 'A is wrong' and each speaks truly, this is not problematic relativism, because each means something straightforwardly different by their term 'right'. For this is an even stronger difference than in the inter-world case we discussed earlier; they differ on *both* clauses not just the second.

But what if they differ in the first clause, but do not differ in the second? In other words, what if two populations converge on tracking different naturalistic properties, but their practices show that they believe that there is a necessary connexion between desire and judgement about that property?

This is more unsettling, because it seems that the second clause has more of a subject-settling role. The second population has *a* concept of the right in virtue of those beliefs about the connexions between desire and judgement. This is why it is disturbing that they can disagree with the first population and yet neither of them speaks falsely. On the other hand they do not have the *same* concept of the right, since the properties tracked by clause (a''') are different. This is what explains how it is possible for the disagreement to be faultless. Do the two populations mean the same thing by 'right'? Well if 'meaning' is taken to be a referential property, no they don't. But if meaning is taken to be

something more abstract—something like *A*-intension<sup>23</sup> in two-dimensional semantics—then perhaps they do. It is the job of another paper to show how two dimensional semantics can be used to explain how clause (b'') is subject settling and fixes the *A*-intension, and clause (a'') the *C*-intension. Here it suffices that some of the puzzle of ethical relativism is removed when we see that there can be two conceptions, similar in virtue of a crucial clause that makes them concepts in the same family. This similarity makes it puzzling if populations faultlessly differ—but at the same time the difference between conceptions explains why this faultless disagreement is possible.

## 7 Strange truth conditions and the flight to error

In a number of areas of philosophy I have been accused of a procedure where I vindicate realism in an area via strange truth conditions. I am not alone in this; compatibilism as a doctrine about free will, for example, is a case where realism about a domain is vindicated via odd truth conditions<sup>24</sup>.

A very natural response might be that if somewhat surprising metaphysical facts turn out to obtain, then rather than allocate odd truth conditions, we should instead see that the error theory is true, or perhaps give up the idea that the discourse is truth-apt.

Returning to the domain of ethics: I claim that careful analysis of the use of moral concepts shows that we judge actions are right when the two clauses of the analysis hold. But the second clause (b'') requires that most people have certain beliefs that are in fact likely to be false. So we get the implausible outcome that a truth condition for an action being right is that people believe falsely that there is a necessary connexion between judging it to fall under the concept <right> and desiring to perform it! Perhaps this is so absurd, runs the objection, that we should say that even if this is the best deserver of the term 'right' then it is still not deserver enough. We should thus resort to the error theory, and deny that any action is right. Perhaps this would proceed by

---

<sup>23</sup> This terminology is that introduced by Jackson *op cit* pp 48-9. Some reader may be familiar with the terms used in Chalmers *op cit* : 1-intension and 2-intension.

<sup>24</sup> Braddon-Mitchell, D. 'Lossy Laws.' *Noûs* 35, no. 2 (2001): 260-277 and 'Qualia and Analytical Conditionals' forthcoming in *The Journal of Philosophy*.

upgrading the second clause to the claim not that internalism is *believed* to be true, but rather that it is true. On this internalistic conception of what is necessary for an action to be action, the assumption that there are no actions which have a property the recognition of which is necessarily connected to motivation, is tantamount to the assumption that no actions are right. Alternatively, the objector might proceed, we could give up on the idea that the discourse is fact stating.

A very simple response would be to say that if it is possible to show that the analysis I give is what really does, however tacitly or implicitly, govern moral judgement then it does give the meaning of moral terms. Thus if there are things that have the properties that are found in the analysis then the error theory is false.

But this is far too simple. For I have said little about what *governing* a moral judgement amounts to. For we can of course find out there are things *causing* our judgements, and that the things that cause our judgements sometimes exist (as they must!) and still discover that the moral states of affairs we judged to obtain in fact do not obtain. This is usually because the actual causes of our judgements are too different from what we took it that they were meant to be. A simple causal theory of governing would inherit the problems of simple covariational theories of content. Just as in those cases misrepresentation is impossible because whatever causes a representation is the state of affairs that is represented,<sup>25</sup> in this case whatever causes the moral judgements will be the naturalistic analysis of the moral states. Thus the error theory cannot be vindicated just so long as *something* is causing the judgements.

In fact what often happens is that when we find out what is causing our judgements, we conclude that it isn't good enough to be a truth maker for our judgement. When we find that our judgements that phenomena are ghostly are not caused by spirits of the dead we do not conclude that the causes, whatever they are, are ghostly.

In this case perhaps it depends on a consciously articulated theory of the ghostly whose analytic status is so central, that on discovery that the actual causes do not realize this theory, elimination of the ghostly follows. It is not

---

<sup>25</sup> I mean here the literature that descends from Dretske, F. *Knowledge and the Flow of Information*. Cambridge, MA: Bradford Books, MIT Press, 1981.

always that simple; there are cases where it would have been very difficult to predict in advance which parts of the theory are in the analytic inner circle that determines matters of elimination.<sup>26</sup>

The short version of the story that I will adopt but not defend here<sup>27</sup> goes this way: the crucial thing is the reaction to the discovery about what it is that typically causes the judgement. What needs to be defended is the claim that, if the account of the conditions under which we form moral beliefs is right, and it were known to be right, that we would in fact not adopt an error theory or regard our discourse as non-cognitivist. So in a strange way the task of vindicating this cognitivist account is a species of philosophical prediction: predicting that the account, if right and known to be right, would not make us all error theorists.

It is important that what matters is not each individual's prediction about what their reaction would be to the discovery that the properties they are tracking lack the nature that they imagine they have. Strong theists sometimes claim that for them the divine command theory of ethics is analytic, and thus that the discovery that this is a godless universe would be, *ipso facto*, the discovery that the error theory is true of ethics. However those who lose their faith rarely become nihilists, and the continuity of their moral practice and discourse is often great enough to make us think that they are not adopting completely new concepts when they continue to hold, for example, that murder is wrong. There is much to be said about the kind of conceptual priority possessed by a supernatural theory of ethics. I am inclined to think that if there were platonic entities that somehow were connected to desire, or there were divine commands, we really would regard other possible worlds that lacked these properties as morally empty.<sup>28</sup> But for our current purposes we just note that the surrogate properties seem to do the trick, even if people do not have

---

<sup>26</sup> Braddon-Mitchell, D. 'The Subsumption of Reference, (forthcoming)., Stich, S. and M. Bishop. 'The Flight to Reference.' *Philosophy of Science* 65. no. 1 (1999): 33-49.

<sup>27</sup> It is defended elsewhere in Braddon-Mitchell, D 'The Subsumption of Reference, and 'Qualia and Analytical Conditionals'.

<sup>28</sup> This is exactly analogous to my treatment of qualia in 'Qualia and Analytical Conditionals'.

much of a theory of what they are. So it is not at all clear to me that if someone were to come to believe that the present analysis is one which is in fact governing moral judgement, that they would embrace nihilism. Few of us now expect the truth conditions of ethical claims to be simple, or entirely independent of human concerns and beliefs.

## 8 Conclusion

What has been defended is a kind of cognitivist ethical theory that is Humean and which answers to much of what motivates internalism. It is straightforwardly cognitivist, for the judgement that (a'') and (b'') hold is a straightforward judgement of fact. It is Humean in that it accepts the distinctness of beliefs and desires. But it is not internalist, although it has features that resemble internalism. Recall that in section 3 I noted that one of the proto accounts of BAM TN was almost internalistic, because while it was compatible with there being no necessary connexions between actions being right and motivation, nevertheless one could not judge an action is right without believing that there is a rational requirement to be motivated. The final version is even less internalistic. There is no connexion between judging that an action is good and either being motivated or even believing that one would be rational to be motivated. But there is a weaker kind of connexion: there is a connexion between judgments about the right, and beliefs about most people's beliefs about necessary connexions between judgements and motivation. It is this indirect link that explains the intuitions that ground internalism, and its very indirectness avoids the metaphysical confusion and difficulty of full blown internalism.

There is, though, one peculiar feature of my view. If one person comes to explicitly believe my analysis (coupled with the view that there are in fact no necessary connexions between desire and judgement) or even if it becomes orthodox among philosophers, then this does not affect the presence of rightness in the world. For there will still be plenty of people holding the beliefs mentioned in (b''). If the entire community, however, comes to explicitly believe the analysis—and to accept that the connexions do not hold—then no one will have the beliefs mentioned in (b'') and, as we saw before, no actions will be right. The error theory will become true.

I embrace this conclusion. I do not think it likely that it will happen—especially as I noted above that mere theoretical opinion will not be enough to count as lacking belief in a connexion between desire and judgement. That would require a great change in the human form of life. In the end, though, I think it's right that the moral properties could be snuffed out by a universal change of belief of this sort. The point is exactly that while there are ethical properties, what they depend on in part are metaphysically mistaken beliefs. Some actions are right, but it is believing falsely that makes it so.

## References

- Braddon-Mitchell, D. *Concepts and Conceptual Change*, MS in prep.
- Braddon-Mitchell, D. 'Lossy Laws.' *Noûs* 35, no. 2 (2001): 260-277.
- Braddon-Mitchell, D. 'Qualia and Analytical Conditionals.' (forthcoming in *The Journal of Philosophy*)
- Braddon-Mitchell, D. 'The Subsumption of Reference'. (forthcoming)
- Braddon-Mitchell, D. and C. West, 'Temporal Phase Pluralism.' *Philosophy and Phenomenological Research* 62 (2001): 1-25.
- Brink, D. O. *Moral Realism and the Foundations of Ethics*, Cambridge Studies in Philosophy. Cambridge ; New York: Cambridge University Press, 1989.
- Chalmers, D. 'The foundations of two dimensional semantics'. (forthcoming).
- Dretske, F. *Knowledge and the Flow of Information*. Cambridge, MA: Bradford Books, MIT Press, 1981.
- Dummett, M 'Truth' in *Truth and Other Enigmas* Duckworth, London 1978 pp 1-24
- Moore, G. E. *Principia Ethica*. Cambridge: Cambridge University Press, 1922.
- Jackson, F. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press, 1998.
- Jackson, F. and P. Pettit. 'Moral Functionalism and Moral Motivation.' *Philosophical Quarterly* 45 (1995): 20-40.
- Lewis, D. 'Desire as Belief.' *Mind* 97 (1988): 323-332.
- Lewis, D. 'Desire as Belief II.' *Mind* 105 (1996): 303-313.
- Smith, M. *The Moral Problem*. Oxford: Blackwell, 1994.
- Stich, S. and M. Bishop. 'The Flight to Reference.' *Philosophy of Science* 65. no. 1 (1999): 33-49.

Williamson, T. *Vagueness, The Problems of Philosophy: Their Past and Present*.  
London ; New York: Routledge, 1994.