Moral Rationalism and Rational Amoralism[1]

(DRAFT of 2/2009)

Metaethical Rationalism can be roughly characterized as the idea that the requirements of ethics are requirements of practical reason. The idea is attractive, in part because it can explain the plausibility of certain versions of motivational internalism about moral judgements. Since rationalism entails that right action is a species of rational action, it appears rational people must be motivated to do what is right, something many internalists believe.

But rationalism's attractions are often not well-enough appreciated because the very feature that makes it attractive also generates a *prima facie* objection. Rationalism seems to require that those who refuse to acknowledge correct moral demands are therefore irrational. Yet such people don't always seem irrational to us. The recent metaethics literature is rife with examples of apparently rational people unmoved by some or all moral requirements, and we can think of more. Some people act on incorrect moral views and appear rational when they do. People sufficiently removed from ourselves in time, place, and culture often have a divergent conception of what morality requires. If we are right about what morality requires, then they are wrong. Yet it seems unfair to accuse them of irrationality as opposed to some other sort of mistake; nothing in their experience prepared them to see things in the way morality requires. Still other apparently rational individuals
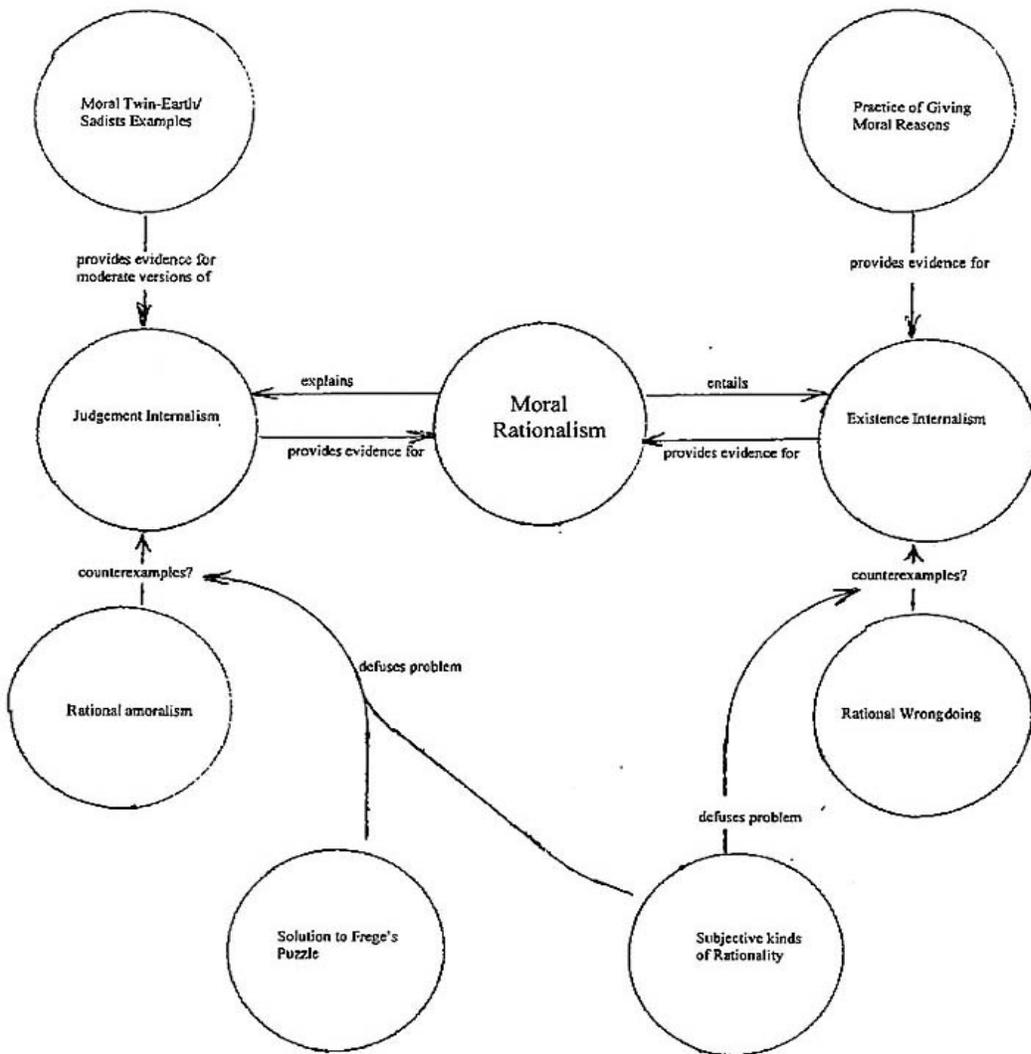
remain unmoved by what they believe to be right.[2] If this means they are not moved by what they believe is most reasonable to do, the lack of motivation would appear to count as a species of irrationality. It is partly on the basis of such examples that many theorists conclude that the requirements of ethics cannot be the requirements of practical reason.

In this paper I will defend rationalism against these worries. But I hope to do more than that. To begin with I intend to show how the rationality of people like those described above is compatible with two plausible versions of internalism. Secondly I will show how properly formulated rationalism serves to *explain* these plausible internalist theses and the plausible cases of rational amorality and immorality which they allow. The result will be that plausible internalism and well-formulated rationalism are mutually supporting theories.

Two sets of ideas are critical to my argument. One turns on the recognition that reasoning is a temporal process and that what it is rational for a person to think or do can depend on features of her history, circumstances, or information. I argue this holds for both *a posteriori* and *a priori* reasoning. I embed the relevant points within a framework distinguishing various senses or kinds of rationality, each of which can be defined relative to distinct features of the agent's history, psychology, and epistemic or practical situation. When fully worked out, this package helps us to explain how rational people can have the wrong moral views.

A second set of ideas interacts with those just described to handle a different sort of counter-example – that of rational persons who may or may not have the right moral views but who are unmoved by their moral beliefs. Here I invoke considerations familiar from the literature about Frege's puzzle and related issues in the philosophy of mind and language to show that rationalism does not rule out such examples. And I argue that the resulting view is still powerful enough to defend a moderate internalist thesis connecting morals and motives, one which has real bite but does not render the counter-examples impossible.

Together these ideas allow defenders of rationalism to defend two plausible and moderate versions of internalism about moral judgements, one connecting the *truth* of moral judgements with rational motivation in certain conditions, and the other connecting *belief* in a moral judgement with motivation. And they allow rationalists to argue in favor of their overall position as the best explanation of the sorts of internalism that are most plausible. Since the argument is complex, the following diagram may help to understand the relations of the various claims:

Moral Twin-Earth/
Sadists Examples

Practice of Giving
Moral Reasons

provides evidence for
moderate versions of

provides evidence for

explains

entails

Judgement Internalism

Moral
Rationalism

Existence Internalism

provides evidence for

provides evidence for

counterexamples?

counterexamples?

Rational amoralism

defuses problem

Rational Wrongdoing

Solution to Frege's
Puzzle

defuses problem

Subjective kinds
of Rationality

4

A Terminological Note

Since I am going to use the word 'rational' and its relatives quite a bit, I should clarify how I am using the term. I intend to use 'rational' as a term for a very general normative property, roughly the property an action, intention, or belief has if it makes sense. An action is rational if it makes sense to do it. A belief is rational if it makes sense to accept it. An intention is rational if it makes sense to adopt it. To show that a belief, desire, or action is rational is to justify holding the belief or desire or doing the action. To show that a belief, desire, or action is rational is to demonstrate that we have normative reasons for it.[3]

What Sorts of Internalism Need Explaining?

Two different sorts of internalism are favored by arguments independent of any particular overall metaethical theory. (1) It is plausible that having an obligation to do something is necessarily a reason to do it, or to put it another way, that true moral judgements give us reasons to act in the ways they commend. And (2) it is plausible that there is a necessary connection between believing something right and being motivated to do it. The first sort of internalism, existence internalism, connects *true claims* (whether believed or not) with *reasons* for action. The second sort, judgement internalism, connects moral *beliefs* with *motivation* on the part of those who accept them. (Darwall 1983)

Both sorts of internalism are controversial, so I will say a few words in

5

defense of each and of the particular versions of each employed in my argument. Existence internalism connecting moral truths with reasons is supported by the role moral arguments play in the justification of actions. When someone asks for a reason to do something, it is appropriate and not obtuse to explain that the action in question is morally right and to offer an explanation of why it is right. No further answer to the why question would normally appear to be needed.[4] Critics of internalism are correct in pointing out that some agents may be unsatisfied with this answer. A person can doubt that she has a reason to do the action in question even in the face of such an explanation. But this does not by itself show that there is no such reason. What it shows is that not everyone accepts internalism. If the doubters are rational one might, however, think that this result would show that existence internalism is false. After all, how could rational people ignore reasons they have? My account of rationality and the way it is tied to morality is partly aimed at diffusing this worry. Once the account is in place and further nuanced to accommodate Frege puzzles, we will have reason to think such doubts compatible with the claim that moral judgements are essentially connected to reasons. This is a promissary note – I need most of the rest of the argument in the paper to make good on this claim. If the objection can be defeated as promised, the *prima facie* case rooted in our use of moral arguments to justify and give reasons for action is sufficient to support the relatively strong existence internalism which rationalism entails.

What defense do I offer of judgement internalism connecting moral judgements and motives, and what sort of necessary connection do I defend? My favored formulation is one explicitly defended by Jamie Dreier, and one which gathers support from thought experiments discussed by Hare, Dreier himself, and Horgan and Timmons.[5] Roughly stated, the favored formulation claims that a rational person who believes an available action right will normally be motivated to do that action.[6] Whenever we find a rational person who sincerely expresses a judgement that an action is right and who yet remains unmoved, it will be a case in which that person is abnormal in some way and one where other normal people appropriately related to that person would be moved by such a judgement. Following Dreier, I will call this 'moderate internalism' or 'moderate judgement internalism'. Thought experiments which support this version of internalism all suggest that our willingness to translate a foreign term with a moral term of our own depends upon the use of that term by normal members of a community to express action-guiding judgements.

Let me illustrate by discussing some examples from the literature. Note that moderate internalism is consistent with the possibility of Brink's amoralist, a person who seems to sincerely avow that some action or other is right and yet claims to have no motivation whatsoever to do the action in question, even when she is in a position to easily do so. Suppose we ask ourselves why we are inclined to take the amoralist's

avowals to express the belief that the action in question is right. It is not just because her term, 'right' is the same as our term 'right' which we use to express the thought that an action is right. If we think of the amoralist in isolation, uttering the same sentence and showing no motivation to do what she calls 'right' there is no reason to attribute a thought about rightness to her. When we think of the amoralist as expressing thoughts about rightness, we imagine her as someone like Uriah Heep or Thrasymachus[7], as a member of a speech community also using the same term to refer to rightness. We attribute thoughts about rightness to the amoralist because she is part of a speech community that uses her term to predicate rightness of actions. This then raises a question about why we are confident *her community* is using the term to predicate rightness. The best answer is that community members use the term in such a way as to guide action – normally these people treat the thought that the action is right as sufficient for showing that it makes sense to do the action in question. Simply stated, they treat the believed rightness of an action as sufficient to rationalize a motive to do that action. The upshot is that we can imagine people who believe an action right while remaining unmoved, but only against a background in which this is not the normal case. Moderate internalism, thus, is vindicated by careful consideration of Brink's purported counter-example.

Outline of How Rationalism Explains Existence Internalism and Moderate Judgement Internalism

Existence internalism postulates a necessary connection between having a moral obligation and having a reason. By reducing facts of morality to a (possibly improper) subset of facts about rationality, rationalism entails this sort of internalism. For example, rationalism might say that to have an all things considered moral obligation/reason to φ, is just to have an all things considered reason (perhaps with the right sort of ground) to φ. For it to be right to φ is for it to be rationally required to φ (again perhaps on certain grounds).[8] And so on. The analysis will entail that true moral judgements imply that an appropriately situated agent has reason to do what the judgements commend, hence it entails existence internalism. This much is simple. The complication comes in explaining how rational people can be unmoved by what they have moral reason to do – a task I will move to shortly.

It is not as simple to show that judgement internalism falls out of rationalism. Judgement internalism entails that even false moral judgements are necessarily connected with motivation in those who believe them. Brink's amoralist is normally considered a problem for judgement internalism, but I have already explained why a moderate version of judgement internalism is immune to the objection. In light of that explanation what we need to explain is the moderate internalist claim: Necessarily, rational agents will normally be motivated by the moral judgements they accept.

Rationalism suggests a very tight necessary connection between sincere

9

moral judgement in rational people and motivation. It explains this connection in much the same way that we might explain why rational persons will do what they believe they have most reason to do. For, according to a rationalist, to a first approximation the belief that φ-ing is right for one is equivalent to the belief that one has most reason to φ. And it looks like it is a requirement of rationality that one be motivated to do what one believes one has most reason to do. The problem for rationalism, is thus explaining the weaker version of the theory suggested by moderate internalism. If there is a requirement of rationality to do what you believe you have reason to do, doesn't rationalism make those who are unmoved by their sincere moral judgements irrational? As I'll explain below, there are reasons to think that even rational people can be unmotivated by what they regard as true moral judgements. This is we will need to include Dreier's 'normally' in the correct statement of internalism, even if we have already limited ourselves to quantifying over only rational people.

My explanation of the details – that is how abnormal but rational failure to be moved is consistent with rationalism – requires both a sketch of a theory of rationality and some mode of theoretical response to Frege's puzzle. Here I just note that the problems for squaring rationalism with each kind of internalism are similar. Each involves explaining how the postulated necessary connection between moral judgements and motives can be rationally disregarded consistent with moral

10

requirements remaining rational requirements.

<p style="text-align:center">**Additional Resources for the Theory**</p>

<p style="text-align:center">**Human Rationality**[9]</p>

The kind of rationality that a moral rationalist should work with is rationality for human beings. And because humans are limited in various ways, the appropriate conception of rationality should take those limits into account. We have limits on how much we can know, on what kinds of investigation we can pursue, on what we can do, and on what we can perceive. Partly as a result of such limitations, human rationality is something both more and less demanding than having full information. It is more demanding because a person with full information is not thereby rational. And it is less demanding because one can be more or less rational despite lacking information. To be sure, rationality requires true beliefs about a substantial number of things, but few of these are directly and universally specifiable by their content. Which true beliefs a rational person should have depends in large part on the evidence to which that person has been exposed. Rationality requires of me that I believe I am wearing a vest, but only given that the vest is here within eyesight or touch and that I have looked or felt. In this way, the requirements of rationality are conditioned by the history of the person to whom they might apply. Or to put the point another way, the requirements of rationality are often conditional on having had certain experiences or having gone through certain thought processes.

<p style="text-align:center">11</p>

Perhaps not every rationally required belief is like this. Perhaps some are just straightforwardly required by rationality; Descartes' *cogito* may be an instance. I will not quarrel over such examples, although I will insist that they are less common than one might suppose. There is a temptation to think of all *a priori* knowledge as rationally required without regard to the knowing agent's circumstances or history. Perhaps this is because of a tendency to think that the conceptual nature of *a priori* knowledge means that grasping the concepts needed to express them is sufficient to justify a person in believing their truth.

This is misleading. Even if *a priori* knowledge is conceptual knowledge, knowing the conceptual truth can often involve a great deal more than understanding the concepts involved. Take some of our best candidates for such knowledge, say knowledge of arithmetic or logical truths. It would be absurd to fault a person's rationality for lacking any moderately complicated bit of mathematical or logical knowledge, even where she understands all of the concepts involved. The reason is that knowledge of many such propositions requires proof – a process of justification of these propositions starting with better established claimss. If I am rationally required to believe one of the less immediately obvious propositions *a priori*, that requirement rests on my having gone through an appropriate process of reasoning. The *a priority* of a conclusion thus does not exempt the rationality of believing it from dependence on historical facts concerning the person who is required to accept

it.[10] Empirical beliefs are even more obviously dependent on features of the agent's history. The rationality of empirical beliefs depends both on the empirical evidence one has actually met and the reasoning one has done based on that evidence.

This brings us to another kind of requirement and limitation to human rationality. Rationality concerns not only what to believe in what circumstances, but also what efforts we should make to collect evidence and what reasoning we should do. Hence there are rational requirements to pursue evidence and also to reason our way through to various conclusions in appropriate circumstances. For us humans there are limits to our abilities in these regards, and rationality for creatures such as us reflects that. We are not required to accept every consequence of everything we believe, nor to collect all the evidence relevant to our possible beliefs.

So far I think this should be uncontroversial. But the discussion so far has left out a large part of rationality, that part regulating desires, intentions, actions, and the interrelation of belief and desire. The details of this part of rationality, call it practical rationality, are controversial but the claim that rationality is concerned with practical matters should not arouse controversy. There are obviously rational requirements relating means to ends. Humeans will believe these sorts of norms to be all there is to practical rationality. Non-Humeans will supplement norms such as this with norms governing the choice of ends themselves. Whatever the exact content of such principles, they are important, and importantly different than

requirements governing only belief, in that they are essentially practical. Failure of rationality in this respect is not reducible to failure of rationality with respect to belief (Korsgaard 1986).

Practical rationality also respects human limits. It includes principles that have application just because there is only so much one person can do. We each have many more ends than we can bring about in one life time. Thus it cannot be a requirement of rationality that we do whatever is necessary to bring about any given end. Given limited time, there will be conflicts. We need to choose between them. There will be controversy over what norms should govern such choices, but this does not undermine the idea that there must be some such norms. Even if there are many different correct ways to choose, some choices will be irrational and some choices will be more rational than others.

Human limitations have another role to play here, analogous to the role they played in determining the rationality of belief. Since we have limited time to think through our options, and since the best way to trade off will not always be clear, the rationality of choosing one way versus another will be partly a function of the opportunities a person has to think things through. Flipping a coin to determine a trivial matter where time is tight makes sense; relying on a coin toss where the matter at hand is important and time is ample does not. Furthermore, the rationality of a practical decision is in part also a function of the actual process of thinking it

through. While it might be rational to embark on a given course of action if one has not thought through its consequences, it may no longer be rational once one anticipates certain bad effects.

The picture I am sketching depicts rationality as involving multiple kinds of requirements, the application of which depends on factors regulated by other requirements. Each of these is subject to human limitations, generating still further principles for dealing with such limitations. The upshot is that what even full rationality requires of a person will depend on a variety of factors, including the situation the person is in and the opportunities that situation gives the person, both for investigation and for action, as well as the actual history of deliberation that the person has engaged in. Furthermore there are a variety of ways that people can depart from full rationality, even holding fixed such background conditions. These departures too can ground reasons. Thus we can idealize along each of the relevant dimensions when we consider what rationality requires of a person. We can idealize people's epistemic positions to a greater or lesser extent while holding fixed that they are fully rational, and thereby vary what we think a fully rational person would know or do. We can hold fixed what the person believes and desires, or the person's evidence and opportunity to deliberate, and rank courses of action for rationality.[11] We can hold fixed even an irrational feature of the agent's psychology and rank different courses of action for rationality given that fixed psychological feature.

These rankings will have to take into account a variety of factors – all of the factors I have been characterizing as distinctive of human rationality. The relative rationality of a belief, desire, intention, action or choice will depend both on the circumstances a person is in and on the course of events and actions that led up to that situation. One important upshot is that many of the reasons people have would not exist for creatures who were more ideal than us in various respects. We have reason to seek additional information because we don't know certain things. We have reason to choose an outcome with a high expected benefit because we don't know and don't have time to find out which option will actually have the highest actual benefit. We have reason to avoid temptation because we know that we can be successfully tempted. And so on.

The idea here is related to a certain standard proposal for drawing a contrast between what a person should rationally do objectively speaking vs. what they should do subjectively speaking. Roughly put, the standard proposal is that an option is objectively rational if, given the actual situation, it would make sense for an agent to choose it. It is subjectively rational if given what the agent believes it makes sense for her to choose it.[12] The basic idea here is fine, but I think it is presented much too simply. There is not just one determinate way to make this sort of contrast, since different subjective features of an agent can determine what it makes sense for that agent to do. Thus for each such feature, what makes sense for

the agent to do with the feature present differs from what it would make sense to do were it removed.   Each of these features might be used to generate something like the contrast between subjective and objective rationality.[13]

To illustrate let me discuss a couple of limitations. An agent may only have probabilistic information about the results of her actions. Given that limitation on her actions, it will be rational for her to choose the action with the greatest expectation of bringing about a desired result even if objectively speaking that action will not bring about the desired result. Here we can say that, objectively, she should not have done what she did, but subjectively speaking, she did do the right thing. Our correctly saying this depends on assuming that there are certain limitations on the agent's information gathering capacity, or perhaps at least that she reasonably believes there are. We might well be interested in what is rational relative not to full information, nor to the information that she actually has, but instead relative to information she could have had she done some investigation prior to deciding what to do.[14] This sort of rationality is also subjective in one good sense though it is not the same sense in which choosing the option with the highest probability of bringing about the good given what she knows. It is also objective in a good sense, in so far as it is relative to information not currently subjectively available to her.

Most of the examples above involve no failure of rationality, since a lack of information or time to think is not a rational failure. But some reason-grounding

17

limitations differ from these insofar as they constitutively involve irrationality. One example of this type has already been mentioned: I might be weak-willed and hence have a reason to avoid certain sorts of temptation. This means that an otherwise rational person with this sort of disposition will avoid temptation when it is within her control to do so, other things equal. In one good sense the person has a reason to avoid (say) stocking the refrigerator with ice cream. In another sense, there is no objective reason for her to do this, since were she fully rational it would be convenient if the fridge contained ice cream and a rational person would not be weak willed.[15]

Frege's Puzzle And Strategies to Capture It

Philosophers are fond of offering analyses that postulate identities between items that seem to be distinct. And they often argue for these identities by suggesting that the relevant identities, if true, would explain why competent cognizers act in various ways that would be explained if the two items were identical and they knew it.[16] But various examples presented by Frege suggest we must all recognize that facts about identities, including facts about the identity of properties, are not always cognitively available to people, not even to fully rational people. Even though Venus is the Morning Star, it appears one can know that one is looking at the Morning Star without Knowing that one is looking at Venus. This complicates the above sort of argumentative strategy for establishing a philosophical analysis. The

explanation of the relevant cognizers' behavior seems to require not only that the entities are identical but also that the cognizers are in a position to know that the things in question are one and the same.

Rationalism is or entails an identity thesis. It claims that one property of actions, rightness, is identical with another, that of being rational to do. And theorists often argue for it in roughly the way indicated above, by showing how its truth would explain the actions of competent cognizers employing moral concepts . Frege Puzzles complicate this strategy of argument for rationalism, just as they complicate similar arguments for other philosophical analyses. If such cognizers can rationally doubt the identity, and indeed they can, the explanation of what they do must be more complex.

Furthermore, it doesn't really matter whether or not the identity is knowable *a priori* or only *a posteriori*. One important upshot of the previous section of this paper is precisely that lack of knowledge of *a priori* matters is not always, or even often, a rational failing. Even if a fact can be known *a priori*, knowledge of it may still depend on having gone through the relevant process of reasoning to show that it is true.   Thus as someone who wishes to argue for rationalism using data about what competent speakers say and think about morality, it is incumbent on me to add the needed complexity to account for Frege's puzzle. I'll do this with a discussion of the two main approaches to the puzzle, Fregeanism and Millianism. I'll offer some

suggestions about how each approach should treat judgements about moral properties such as rightness and wrongness. Many of the general points will be familiar; the innovations lie in my explanations of how to connect it up with internalism.[17]

It should be uncontroversial that two terms can designate the very same thing, even while a competent speaker is unaware that they do so. Using the example of co-designating names, it is possible for a competent speaker to use two names, say 'Cicero' and 'Tully', and yet not know that they designate the same person. What is controversial, of course, is how to accommodate this fact within a theory of meaning for the relevant terms. Millians about names will want to treat the terms 'Cicero' and 'Tully' as having the same meaning or semantic value, which, along with some auxiliary assumptions about the contents of beliefs, will lead to the result that a competent speaker can without irrationality believe contradictory propositions. The general Millian picture is this. The sentences, 'Cicero was a Roman' and 'Tully was a Roman' express the very same content because each of their constituents have the same meanings and contribute the same semantic values to what is asserted by the whole. Thus, if propositions are whatever assertive utterances express, the two sentences express the same proposition. Furthermore, if what I say when I say "Cicero was a Roman," is just what I believe when I believe that Cicero was a Roman, and similarly for other such beliefs, this belief will have the same content as the belief that Tully was a Roman. Thus someone who believes the former but

20

also believes that Tully was not a Roman has contradictory beliefs.[18]

Fregeans on the other hand, will want to postulate senses or modes of presentation as constituents of the meanings of the relevant designating expressions so as to explain how a rational person can treat the sentences in question as differing in truth value. Roughly speaking, a sense is a way of picking out what the term designates and the same thing can often be picked out in multiple ways. If different terms are associated with different senses, the terms will contribute different constituents to what is expressed in using them and the corresponding beliefs will be different even though the objects designated by the terms may be the same. On this way of going, the beliefs of someone who accepts a judgement expressed using one term and disbelieves what is expressed by substituting a co-designating expression for that term need not be contradictory. This is because the propositional attitudes will be toward different propositional objects or involve different constituents provided by the sense corresponding to the terms. Even so, given that the terms designate the same object, it will be true that the two beliefs cannot in fact differ in truth value.

Each of these treatments can be extended from names to predicative expressions. Two predicates might designate the same property and yet competent speakers may be unaware of this. Some theorists will want to treat at least some of these expressions as analogous to the Millian conception of names. For them, the

relevant predicates will take the same property as their semantic value, so that these terms will contribute that property to the content of what is asserted in sentences using those expressions. And from there a similar line of reasoning to that outlined above will generate the conclusion that a person can believe contradictory things without irrationality so long as she is unaware that the predicates in question designate the same property. Other theorists will want to treat properties in the Fregean way, postulating senses as the semantic values of predicates even while these senses pick out the same property, though in different ways. On such a view two sense-mediated beliefs can have incompossible truth conditions and yet be rationally held.

The terms that concern us in metaethics normally function as predicates. We can say of a certain action that it "is right". We can say of an action that it "is rational". When we say the former we are predicating a property, rightness, of the action. When we say the latter we are predicating the property of being rational of the action. The rationalist proposal is that these are in fact the very same property, and that the term 'right' designates the same property as ' rational' or something very similar. Given this analysis, it is open to the rationalist to treat either or both predicates as the Millian treats names, that is as contributing just the property to what is expressed, or as the Fregean treats them, as contributing a mode of presentation of the property to the proposition expressed.

While my inclination is to think the Millian view correct, at least for the term 'right'[19], my main point relies only on accepting what the two views share in common - that competent speakers can be unaware that co-designative property terms pick out the same property. I think this possibility is open even where the fact that the two terms are co-designative is secured by *a priori* philosophical argument. For we can find examples fitting the pattern exemplified by the Cicero and Tully example even when what seem to be two properties are necessarily coextensive and even identical, and where this can be shown *a priori*. Just as one can be ignorant of an empirically confirmed identity because one has not made the necessary investigations, one can be ignorant of an *a priori* accessible identity if one has not gone through the relevant reasoning processes. Earlier in this paper I argued that ignorance of a fact need not by itself constitute irrationality. For uncovering the fact might depend on some process of discovery that one is not irrational for not having undertaken. This remains true when the matter to be discovered is the identity of seemingly distinct objects or properties, and even when the process is one that leads to *a priori* arguments for those identities.

To illustrate the point about *a priority*, let's go back to mathematics. One can be unaware of a mathematical identity because one has not found or constructed the relevant proof or done the relevant calculation. In such a circumstance a rational person may seem to believe of some arbitrarily chosen number *n*, which turns out to

23

be the 2001st smallest prime, that it is not a prime number, and hence that it is not the 2001st prime number.[20] That is, she will avow that the number as named by its ordinary Arabic numeral is (probably) not prime and that is normally sufficient to license our attributing the relevant belief. At the same time, she will no doubt avow that the 2001st prime is prime. Since Fregeans have postulated senses precisely to capture what people who seem to have contradictory beliefs about the same objects have in mind, it will be in keeping with the spirit of their position for them to agree that the two thoughts involve expressions with the same referent but different senses. And even if there is an *a priori* proof that each sense necessarily picks out the same object, Fregeans should allow that there need be no irrationality on the part of someone who has both of the beliefs in part because she has not gone through the relevant proof procedure. As noted earlier, failure of knowledge about complicated *a priori* matters is not irrationality and this holds even of senses that are *a priori* equivalent.

For Millians things are a bit more complicated because only one of the two expressions referring to this number, its Arabic numeral name, will likely be treated in the Millian fashion. A thought referring to the same number as the 2001st prime seems more plausibly to invoke a particular way of thinking of its object. But it seems reasonable for the Millian to think that the person believes the number in question is not prime, and also believes of that number that it is prime. In any case,

24

the Millian should regard this sort of case as involving no more irrationality than it would have been before the discovery of Pluto to believe the number of the planets not divisible by three, and yet at the same time to have thought that nine was divisible by three. In each case there was some fact, unknown to the thinker, which makes the contents of these beliefs problematic. And in each case the lack of knowledge of that fact is not due to a rational flaw. Failure to engage in arbitrary lines of *a priori* reasoning are no more failures of rationality than failures to research all the empirical facts.

That the conflicting beliefs involve no irrationality has consequences for the rational assessment of action as well. How it is rational to act depends on what one believes. The belief that a number is not prime may (in appropriate circumstances) rationalize trying to factor it or asking the nearest mathematician what it's factors are. The belief that a number is not prime would normally rationalize not doing any such thing. Where a person has beliefs of both sorts, either in the sense-mediated way the Fregean postulates, or in the way Millians favor, we would not be surprised to find a rational person attempting to find the factors for the number when picked out by an Arabic numeral, but refusing to do so when it is picked out via the description "the 2001st prime." Something analogous can be the case for the property that is designated by 'right' and also (if the analysis is correct) by 'rational'. This point will be important later on to explaining one way in which rational people can remain

unmotivated by their moral beliefs.

Given these considerations, advocates of rationalism can go on in either of two ways. If Fregeanism is correct, they can hold that a thinker or speaker may employ distinct senses or modes of presentation when thinking of rightness. These distinct modes can explain how a thinker can believe that something is the right thing to do, while doubting that it is the rational thing to do (or vice versa) even while the properties rightness and rationalness might be identical. Or, if Millianism is correct, rationalists should say that people can rationally believe two thoughts which are strictly speaking inconsistent. On the one hand they can believe that an action is right while on the other hand believing that it is not rational. Because the 'right' and 'rational' contribute the same semantic values to the thoughts expressed using the terms, the speakers will thereby be thinking a thought and its negation. But because competent speakers may not be in a position to know this, they may nonetheless rationally accept both claims. Neither of these two ways of proceeding will commit the theorist to saying such thinkers and speakers need be irrational.

Rationalism and the Two Kinds of Internalism in Light of These Complications

With these materials, the multiple relativized notions of rationality and either of the methods of accommodating the account of moral property terms to Frege's

Puzzle, we can begin our explanation of internalism and of rational amoralism.

Our first task is to clarify the rationalist thesis given the multiplicity of kinds of rationality. Which of these conceptual resources should a rationalist use when she reduces moral facts to facts about rationality? My claim is that she should use all or most all of them, and that when she does many of the putative problems with rationalism and with realist treatments of moral judgements evaporate.

A rationalist should say that morally right actions are those actions which a rational person would choose in a given circumstance. But a rationalist should not have to choose between identifying rightness with what a fully rational person with full information would do, and what a fully rational person with the agent's limited information would do. A rationalist can have it both ways, so long as she is clear about what she intends to say. And similarly for the other limitations that an agent might be under. She can identify one sort of objectively right action with what would be chosen by a fully rational person under conditions of full information, and she can identify different sorts subjectively right choices with what should be chosen in conditions that depart from the ideal. Since there are several ways that one's situation might be limited (information, time, etc.), this will mean there might be several subjectively right options corresponding to different limitations.

One might worry about the most objective notion of rationality as something that a rationalist is entitled to employ. There are two concerns. One is that a theory

27

which adds full information to the requirements of rationality to generate moral obligations is not really entitled to bill itself as a form of rationalism. For one can be rational and not have full information, as I have emphasized. The other concern has to do not with the label, but with the content of the rationalist analysis. If a rationalist reduces objective moral obligations to what a rational agent would do if she had full information, there is a good chance that the analysis will be circular if full information requires information about right and wrong. Those are just the concepts the rationalist was trying to analyze.

The first worry strikes me as merely terminological, and perhaps a bit unfair. No one ever thought that empirical information was irrelevant to which actions we are morally required to do. And that includes rationalists. Even a theorist like Kant, who to my mind under-appreciated the relevance of empirical information, would admit that the need for empirical information in determining which particular action is right or wrong. Even if it is a truth of reason that lying is always wrong, we won't know which utterances are lies unless we know contingent facts about the world that rationality by itself will not tell us. If Kant was not a rationalist, who was? More importantly, nothing of any philosophical interest can turn on a mere terminological objection like this. If we give up the term, the substantive issue of whether we can make sense of moral facts by reducing them to truths about what makes sense to do given certain sorts of information will remain. 'Rationalism' seems to me a good

term for the thesis that we can, but that is not a philosophical claim.

The worry about circularity can be handled by qualifying the rationalist analysis to avoid it. When we first explain that the right thing is what a rational person would do given certain information we can be careful not to include information about what is right and wrong in that information, or limit the analysis to empirical information. Or at least we can start that way and build up from there. We may need to proceed in stages because there may be second order truths about what is right to do given that some other thing is right or wrong to do. For example, it may be right to discourage others from doing wrong. Thus, if capital punishment is wrong, working to end it would be right. The rightness of the working to end it would depend on the wrongness of that practice. Since we think there is a norm of rationality requiring opposition to what is wrong, we think that a rational person who knows that capital punishment is wrong would oppose it. So long as the initial judgement that capital punishment is wrong does not require independent knowledge of the moral fact that it should be opposed, we introduce no circularity by allowing the rationality of further actions to depend on knowledge of moral status of capital punishment.

In any case, the philosophical point is that by relativizing the rationality of various actions and attitudes to various limits inherent in our situation, we end up with both a fully objective sense of rationality and with many differing subjective

senses of rationality. And we can use each of them. If a rationalist equates rightness with the property that is possessed by those things that are rational to choose, she can generate different notions of rightness, objective and subjective, mirroring each of the senses of rationality we might have an interest in.[21] One advantage to this approach is that it fits many of our actual judgements about what is right and what is wrong. Sometimes we say it is right to do some action where that claim can only make sense relative to some feature of the agent that is less than ideal. At other times we make judgements about rightness that can only be interpreted as a claim about the ideal. This approach allows both, and can allow features of the context to disambiguate which is meant in that context.[22] This is all to the good. But a second reason to like the proposal is what it enables us to say about some of the puzzling cases of immorality and amorality that began the paper.

Explaining Existence Internalism Consistent with Rational Wrongdoing

I've already explained that existence internalism is a fairly straightforward consequence of rationalism. If the truths of morality just are facts about what we are rationally required to do, then we will necessarily have a reason to do what is morally right. That's part and parcel of the rationalist project of reducing moral truths to truths of rationality. Having introduced multiple relativized notions of rationality and equated them with correspondingly relativized notions of rightness, the same line of reasoning vindicates existence internalism for each of these kinds

of rightness. If (for example) relative to an agent's evidence it is right for her to intervene in a dispute, then relative to that same evidence it is also rational for her to intervene in that dispute.

The fact that we can make such judgements relative to different features of the agent's situation can now be invoked to explain how it might be that a person might, without irrationality, do what is in fact wrong. The general idea is to account for various kinds of rational immorality by noting that judgements of irrationality are usually or often made relative to one of the subjective senses of rationality. People who do what is objectively wrong will not be counted as irrational in one good sense, so long as what they did made sense relative to the information that they have. Thus there is a sense in which those who do what is objectively wrong can still be rational, though in one of the subjective senses.

One sort of rational immorality which a rationalist should have no trouble admitting involves actions which are rational because the agent lacks certain empirical information which would, if available, have changed what made sense to do. Clearly such agents are not subjectively irrational since they are doing what makes sense given the evidence they have. But this result is compatible with the chosen action being irrational relative to fuller information that the agent might have possessed. And, by equating what is objectively morally right with what is objectively rational in light of full information, we can truly say of such cases that

the agent did something objectively morally wrong, but rational given what she knew.

This response might seem useless for other sorts of rational immorality. Can't we imagine a person with full non-moral empirical information who still does something morally wrong but is not irrational for doing so? Doesn't history provide us with just such examples? In answer to this we can extend the previous answer. Our access to even *a priori* information can depend on our actual history of reasoning through to such knowledge. If we have not gone through the relevant reasoning processes, we can lack that knowledge and yet be rational. Even so, such knowledge can be rationally privileged, insofar as someone who had gone through the right sort of reasoning would have that knowledge and because in certain circumstances, that process of reasoning would itself be rationally required. This idea can be used to explain several additional sorts of seeming rational immorality. One sort of case is simply where the reasoning leading to the conclusion that some action is right is complex and time consuming. One need not be irrational if one does not do an action that would take a one a long time to determine one should do.

This strategy of response may cover those who are in societies with abhorrent moral views if we think that the process of reasoning which would lead to the rejection of these views requires, among other things, consideration of alternative

ways of life which might not be obvious to people without appropriate acquaintance with other cultures. Even when there is an *a priori* argument in favor of a certain sort of hypothesis, it might take some experience before one is likely to consider the hypothesis and look for reasons for it. And if moral hypotheses can be justified by inference to the best explanation, even if *a priori* argument shows a competing moral hypothesis to be a better explanation, our consideration of that hypothesis might itself require imagination, luck, or experience. Those in situations which do not make the alternative hypotheses salient are not therefore irrational. If those in the societies in question hold their views for reasons such as this, a rationalist need not regard their immorality as irrational.

Explaining Judgement Internalism Consistent with the Amoralist

Judgement internalism is a substantive claim over and above the rationalist thesis, so its explanation will be more complicated than the explanation of existence internalism. I argued earlier that the particular connection to be explained is that necessarily normally a rational person who believes an action all things considered right will be motivated to intend or try to do it.

An overly simple but instructive way to see the beginnings of an explanation is to think about whether it could be rational to act in a way one thinks irrational. It seems not, and this provides the core of our explanation. On the rationalist analysis the belief that an action is *a priori* equivalent to the belief that the action is rational.

33

Or, in the words I used earlier to explain what I meant by 'rational', it is *a priori* equivalent to the belief that the action makes sense to do. To say it is the only right action and hence morally required entails that it is the only action that in those circumstances makes sense to do. So it looks like someone who acts contrary to her moral judgements must be irrational in just the way that someone who acts contrary to her judgements of rationality is irrational. Both Frege's puzzle and the relativization of rightness and rationality argued for above complicate this simple explanation. And that complication is necessary to explain how someone like Brink's amoralist or Huckleberry Finn is possible.

Factoring In Relativised Rightness

The belief that an action is rational to do might have the content that the action is objectively or ideally rational – that it is what a fully rational agent with no rational flaws, all relevant evidence, sufficient time to deliberate, etc. would do. Or it might have the content that the action is rational in one of the various subjective senses relativized to evidence, limited time to deliberate, one's actual rational limits, and so on. It is only irrational to act in a way that is irrational given one's actual condition; It is not irrational to act in a way that would be irrational if one were in some other set of circumstances. Thus the belief that an action is objectively right, in a sense not relative to the actual features of the agent's situation, need not have the tight link with motivation that the simple explanation above relies upon. Only the

more subjective senses must have such a tight connection.

This may seem puzzling if you think only about the rationality of belief. Normally if you think believing some claim is justified, you think that if your evidence were more ideal than it currently is you will remain justified. That is because you think your current evidence justifies the claim that the thing you believe is true and that if the claim is true ideal evidence relevant to its truth will show that it is. If one believes it is rational given one's evidence to hold some belief (say that quantum mechanics is true) one should also believe that one has good evidence to think that one would hold that belief given full evidence. Thus, one should only believe it rational relative to one's current evidence to accept the truth of quantum mechanics if one also believes it would be rational in a fully objective sense to do so.[23]

The rationality of actions and intentions does not sustain as tight a connection between what is rational in actual circumstances and what would be rational under ideal conditions. One sort of example involves the interaction of epistemic limitations with one's goals; another involves practical limitations of the agency. The sorts of epistemic cases I have in mind involve actions under uncertainty. If I don't have conclusive evidence for some claim it may not be rational to act as if the claim is true, even while I might think that if I had all the relevant evidence I should act that way. It may be rational to spread my investments around rather than put them

all in one place, even when I think that under conditions of ideal information I would

know that Firm X would be the best investment and that in those circumstances it

would be rational to invest all of my assets in Firm X. Under conditions of

uncertainty it may be better to maximize one's expected benefit or to avoid risk even

if I believe that another option in fact would make sense under conditions of fuller

information.

Practical non-epistemic limits on a person's agency may also underwrite

divergence between what would be rational in ideal circumstance and what is

rational in one's actual circumstances. For example, if an agent has a certain

disposition to be irrational it can be rational for the agent to avoid situations in which

that disposition will tend to manifest itself even though it would not be rational for

a fully rational agent, that is an agent without such a disposition, to act in the same

way. And this is something we can recognize from a first person perspective. Given

what I know about myself it is wise for me not to stock the freezer with ice cream,

though were I more rational and therefore more resolute it would actually be more

economical than not doing so and hence rational.[24]

These kinds of cases demonstrate that what an ideally rational or informed

agent would do, and what it is rational for an actual agent in actual circumstances to

do can be very different. This complication is in fact an advantage for a rationalist.

We can use it to accommodate certain exceptions to the idea that moral judgements

must motivate rational agents. It will allow us to claim that an agent can rationally choose to act in ways that a perfectly moral agent would not choose, and to show how this is compatible with reducing morality to rationality. This in turn will explain another way in which a person can sincerely admit some action right, while failing to be moved to do it. What that person can sincerely admit is that the action is fully objectively right. In other words, she admits that if she were completely rational she would in fact do it. But she might know that she has some sort of rational failing, such as weakness of will or a bad temper that she should take into account in deciding what to do in her actual circumstances. It may, for example, be right to apologize even when one has not really done anything wrong, in the sense that a person with ideal self-control and a good temper should rationally do so. However, if one has only limited self control, trying to apologize in such circumstances might only lead one to say something that makes things worse. Given certain self-knowledge, it is no evidence of one's insincerity that one does not do what one thinks it is right to do in this sense.

On the other hand, it would show insincerity or irrationality if one thinks that in one's actual situation, as one actually is, it makes most sense to apologize and yet does not do so. I have argued earlier that there is a relativized notion of rightness corresponding to this relativized notion of what makes sense to do, and that the two properties are identical. If a person claims it is right to apologize in this more

subjective relativized sense and yet is not moved to do apologize, we might rightly question either their sincerity or their rationality, or both.

Factoring In Frege's Puzzle

We might rightly question these things, but the person need not always be either irrational or insincere. The considerations I rehearsed earlier about Frege's Puzzle generate a further way that a person could remain unmoved, even if rationalism is true and even if they sincerely judged an action right for a person like them in their actual circumstances. While this complicates the kind of judgement internalism rationalism will underwrite, the complication will again be a be welcome. It will enable us to explain the sorts of amoralists highlighted by Brink, Stocker and Rosen. Since these sorts of amoralists were the primary motivation for accepting the normalized internalist claim embodied in moderate internalism, the complication will cause our rationalist explanation of judgement internalism to precisely match what needs to be explained.

If rationalism is correct, a rational person can have a thought that something is right and yet remain unmoved, but only provided that she is not in a position to recognize the identity of rightness with the property of being rationally required. If she does or should recognize the identity, then she also should rationally be motivated. While some amoralists may be irrational either because they recognize the identity but fail to act accordingly, or because they should but don't recognize the

38

identity, some may fail to recognize it through no rational fault of their own. Even if there is an *a priori* argument to secure the identity, a person who has not worked the argument out might well be ignorant of it and hence not be irrational, despite failing to recognize the connection between rightness and practical rationality. The phenomena highlighted by Frege's puzzle might thus serve to explain this sort of rational amoralism, provided we think those amoralists are ignorant of the identity postulated by rationalism.

In fact the most plausible of such examples do seem to be ignorant in just that way. Uriah Heap and Thrasymachus are most plausibly thought of as doubting the rationality of moral action. Thus they doubt the identity that would otherwise rationalize acting on the relevant moral beliefs.

Fregeans and Millians will describe the states of mind involved somewhat differently. Fregeans will say that the thought that an action is morally right and the thought that an action is practically rational will have different contents even if these turn out to predicate the very same property of the action in question. For, according to Fregeans, the contents of these thoughts will include the ways in which the property is picked out, and this way will be different when a person thinks of the property as moral rightness as opposed to the way involved in thinking of it as practical rationality. Millians will think that the two thoughts have the same contents but that the thinker is not in a position to know that they have the same content.[25]

39

But despite having different ways of characterizing the states contents of the states of mind involved, both accounts should agree that the identity of the property in question does not by itself make it irrational to doubt the rationality of moral action.

Why then think that we have any explanation of judgement internalism? Why doesn't the account allow no connection between sincere moral judgements by rational persons and motivation rather than the normal connection postulated by moderate internalism? Couldn't everyone be unaware of the identity in question and hence not moved by their judgements of what is right? No they could not, for if no one was so moved their judgements would not be judgements about what is right. The Fregean version of this story will suggest that a person can think a thought whose truth conditions require that the action thought right be rationally required while thinking of it in a way that depends (perhaps through deference) on other people taking it to be action guiding. The Millian can say that thinking a thought with this particular content depends on being in touch with others who take the rightness of an option to rationally permit or require one to act rightly. Thus on either approach when a thinker thinks an action right this depends on there being appropriately related to a normal case in which the rightness of the action motivates rational agents.

You might, however, want to know a few more details about the resulting internalist requirement. A full answer will require me to say more about what each

the Fregean Rationalist and the Millian Rationalist have to say. I'll start with an analogy that both sorts of rationalist will want to endorse before moving on to discuss each position separately:

An Analogy

Burge tells us that 'arthritis' refers to a distinctively painful disease of the joints that could not be had in a location that is not a joint. This seems to be purely a matter of the practices of the relevant experts regarding arthritis – medical doctors. They could have decided to apply the term 'arthritis' to similar pains occurring also in the thighs or shins, but they did not. When we find speakers who express their beliefs by saying, "I have arthritis in my thigh," we attribute to them the belief that they have arthritis in their thigh. At least we do if their community is one in which the relevant experts have defined the term in the way that ours have. Had these speakers instead been members of a community that had allowed the term to extend to pains not in the joints, we would have attributed a different thought to them, not about arthritis but about some other similar ailment that extends also to the thighs.[26]

I think there is a marked similarity between this example and the amoralist example as rationalist must conceive of it. In each case we are willing to attribute a thought the truth conditions of which would seem to entail that the speaker is expressing something ruled out by the correct analysis of the terms used to express the belief. And in each case we are willing to do so only against a background in

41

which the most competent and rational speakers would not avow those attitudes, and in which those other people are members of the first speaker's language community. If the amoralist were isolated from communities in which the term 'right' was used to commend we would not have attributed a thought about rightness to her, just as we would not have attributed a thought about arthritis to the medically ignorant patient in a community where doctors did not use the term 'arthritis' to pick out exclusively a disease of the joints.

Because of the role of experts in the arthritis example, the analogy is not perfect. Most people aren't medical doctors, so while they are "normal" in the sense that their beliefs are normative for the extension of arthritis, they may not be statistically normal. But there is no principled reason why for some terms the "experts" who determine the meanings of an expression could not be the normal people, or most rational people, or some other group which would make the content of a person's thought depend on the normal case. For example, it seems plausible to me that this is just how it is with color terms. The term 'red' picks out what it does because of how it is used by normal sighted people. Yet, color-blind people can use the term and refer to the same color, even though they can only do so in a speech community where those who determine what the word means are not color blind. I think that a person could believe that chartreuse was a shade of red, but that this could not be the normal case, even in the statistical sense. The referent of color

terms is determined by the practices of most people in the language community. Competent speakers can run afoul of those practices and yet still possess the relevant concepts and attitudes constituted by them, but only in virtue of the background conditions involving the normal speakers.

So here we have a certain sort of necessary connection between the attitudes of normal speakers in a community but of a sort that does not require that all members of that community share the attitudes. The explanation is that the designatum of a speaker's terms can depend on the practices of the community in which she is a member, and the content of her thoughts expressed using those terms can depend on the same facts about the same community. She may flout the norms of her community and yet harbor thoughts which are partly constituted by the very norms she flouts.

Millians and Fregeans can agree that the actions and practices of most normal speakers in treating rightness as sufficient for rationalizing and justifying an action make it the case that 'rightness' designates the same property as 'practically rationally permitted'. It is these actions by normal speakers in the community that make it the case that the attitude expressed by a speaker of that community by calling an action "right" predicates a property which a thing has only if it makes the most sense to do of the available options.

Fregean Responses

But Millians and Fregeans will disagree on the contents of the thoughts involved. Fregeans want the senses, reflecting the different ways in which speakers think of the referents of their terms, to both determine the referents of the relevant terms and to be constituents of the contents of the thoughts in question. Thus the natural move for them is to build deference to the normal speakers in the community into the contents of the thoughts in question.[27] They take themselves to be referring to what the normal speakers in the community are referring to, and the norms of these normal community members get to determine the referents of their thoughts because they are so-referenced in their contents. In this way a Fregean rationalist can explain the necessity of a normal connection between thinking a thing right and being motivated, while allowing that some rational individuals may not be so motivated.

Millian Responses

Millian rationalists have both an easier and a harder time. Their time is easier insofar as they need not suggest that any particular way of thinking of the property in question is necessary to entertaining thoughts about either rationality or morality. Thus they can explain a lack of motivation. But they then have a harder time explaining why the view generates an explanation of internalism, since it may look like it underwrites no necessary connection with motivation. To begin with the easier task, Millians will say that the two thoughts predicate precisely the same

44

property of the action in question and that the property itself just is the semantic value of the relevant terms. The property is part of the content in both thoughts of morality and thought of rationality. A person can have two thoughts with identical content without recognizing that this is so - witness thoughts of Hesperus and Phosphorus.[28] Thus entertaining a content which when viewed in the right way would rationalize a certain response need not rationalize that response if one does not recognize this content as one capable of doing so. Motivational responses are no different from other rational responses in this respect.

As far as the harder task of generating any substantial internalist commitment, Millian rationalists should first agree with the normal case internalism offered by moderate internalism. Insofar as we've offered principled reasons for accepting this upshot in the story of how these thoughts get their content, Millians have met their burden of explaining just the form of judgement internalism they needed to explain. By this I mean that it is part of the Millian framework that speakers can designate the same item with a term as others in their community in virtue of standing in the right sorts of relations to use of a term in their community without knowing everything there is to know about that term. So if rightness is a property that actions have when they make sense to do, a community member can use 'right' to designate that property if they interact with others in their community who use it to express that knowledge, even when they themselves lack it.

45

In more detail, Millians claim that people gain the ability to use a term to express a certain property insofar as they interact with others using that term to express appropriate knowledge of the referent and come to use the term because of that interaction. On the present rationalist view, 'rightness' refers to the property an action has when it is rational to do because people have used right to express their knowledge that certain things make sense to do. In the usual case they will have gained this knowledge by thinking about what makes sense to do in various circumstances. Needing a term to to express this thought, the predicate 'right' was pressed into service. Those who use it for this purpose will be generally guided by their judgements about rightness expressed using this term. It was by using the term for just this purpose that it got the referent that it has. Those in the community who don't use it with this action guiding purport depend on those who do insofar as they use the term to refer to this property.

We can say a bit about the way in which normal speakers must be moved by their moral judgements. The way I'm thinking of them, they are the ones who either tacitly or actually might be said to know the truth of the reduction. They are in a state of mind that would most naturally be expressed by asserting the sentence capturing the rationalist reduction.[29] For them it would be irrational to think that some action was morally right to do and to remain unmotivated by that thought unless it was equally rational to remain unmotivated by the corresponding judgement

46

of rationality because of its place on the scale from subjective to objective rationality.[30] So here we have a further necessary connection between having the thought that something is right and motivation. But even it is a defeasible connection, for reasons having to do with the distinction between objective and subjective rationality as explained above.

A we can add a further quasi-internalist requirement consistent with the Millian picture without abandoning moderate internalism. Millians need not deny that there are descriptive conditions on competent use of a given term, only that these conditions are sufficient to determine a referent. In the arthritis case it is plausible that a thinker needs to know at least that it is a disease or that doctors think that it is a disease or disorder. A person with no idea of even this much might well not be best interpreted as thinking about arthritis using by using that term. Similarly, insofar as the use of moral terms requires interaction with speakers who use these terms to make reason guided action guiding judgements it would not be farfetched to require even those who do not to know that others in their community take the rightness of an action to indicate reasons to do it. So, it is at least consistent for the Millian to require this knowledge – that there is some sort of connection with reasons – of all speakers who can think thoughts with this content.

Millian and Fregean rationalists can, thus, explain moderate internalism and this means they can accept many sorts of rational amoralism.

47

What Were They Thinking?

A reader might be forgiven for wanting an answer to the question of what (consistent with the analysis) rational thinkers could be thinking when they doubt their reasons to do what is right. So I should say a bit about this to make the view seem plausible from the inside, even though I suspect there is no general account of rational amoralist thoughts. Perhaps I can sketch the situation of a person who has such thoughts in such a way that those thoughts seem rational. One possibility is that a speaker might know rightness as a property which the members of his or her community took as reason-giving, without believing that his or her peers were in fact right for doing so. For example, in a society where people had a tendency to go in for reprehensible things, a person might also believe that the property in question supervened on reprehensible features of actions such that things with this property did not in fact have much to commend them. If everyone around them thinks support for slavery right and acts accordingly, the person might reasonably think these others don't know much about what is right. But that person might instead just as reasonably think their neighbors are right about the extension of rightness, but wrong to do what is right.

A person in that community can use the communal words, 'right' and 'wrong' to think the thoughts of Huckleberry Finn. Huck thought it was wrong to work to free slaves in general and Jim in particular. He believed himself generally

uneducated and those around him to be better judges on all sorts of matters, including matters of morality. He believed he was going to pay a high cost for helping Jim to flee. But thoughts about morality and even worries about his own long-term post-death self interest were insufficient to motivate him to act in accordance with his beliefs about the wrongness of helping slaves. It seems fair to say that he did not see them as providing sufficient reason to act in ways incompatible with his friendship with Jim.

Huck's thoughts here seem perfectly rational. It was not irrational for him to be ignorant of the fact that those around him were no experts about morality or metaphysics. And given his deference on these matters, it was also not irrational for him to accept his neighbor's views about the extension of moral obligation. At the same time, it is most reasonable to think that if morality requires returning slaves it does not make sense to do what morality requires. On this basis Huck could reasonably conclude that it was right to return Jim and that he had no sufficient reason to do so. Rationalism need not say otherwise. Millians and Fregeans will capture the exact content of Huck's thought that it makes sense not to do what morality requires in somewhat different ways, but both can accept this basic picture.

Summary

This leaves us in the following position. Rationalism, as I am construing it, identifies moral rightness with practical rationality or a restricted domain thereof –

roughly an action is morally right if it is practically rational to do (perhaps on certain grounds). Rationalism together with two auxiliary sets of ideas entails two moderate internalist theses, each of which is itself plausible.

First, we can couple rationalism with the recognition that what is subjectively rational for a person depends on historical and psychological contingencies. From this combination we get various requirements of subjective rationality in addition to a most objective set of rational requirements constituting what would be practically rational for a completely rational person with full information to do. These subjective requirements are such that it can be subjectively rational for an agent not to do what would be objectively rational. The agent may lack the information, time, or opportunity to do the reasoning which would figure out what an epistemically ideal and rational person would do. If the objective rightness of an action is equated with its objective rationality, it can be subjectively rational to fail to do what is objectively right because one lacks such information. This underwrites a moderate version of existence internalism: If something is objectively right, then a fully rational, fully informed person, who uses enough time to think things through, would do that action. This is a moderate thesis insofar as it does not rule out that a subjectively rational person might do the objectively wrong thing due to various contingencies of her situation as explained above.

Coupling rationalism together with an adequate recognition of the core

phenomena underlying Frege's puzzle also generates a moderate version of judgement internalism – normally a rational person who judges an action morally right will be motivated to do that action. Rationalism identifies rightness with what is practically rational for a person to do. But even if it is *a priori* that the two properties are identical, someone who has not had an opportunity to work it out is not irrational for lacking that knowledge. Thus there can be rational people who are not motivated. Still they will be unusual. For the use of the term 'rightness' to express the same property as 'practically rational' is dependent on a wider community which uses that term to think about what makes sense to do. Fregeans accommodate this by building deference of some sort into the sense of the relevant expressions. Millians can do it more directly by allowing community use to connect the term to the referent.

Bibliography

Bach, Kent. (1994) "Conversational Impliciture," *Mind & Language,* 9: 124-62.

Brink, David. (1986) "Externalist Moral Realism," *Southern Journal of Philosophy*, Supplement: 23-40.

_____ (1989) *Moral Realism and the Foundations of Ethics*, Cambridge; Cambridge University Press.

Burge, Tyler. (1979) "Individualism and the Mental," *Midwest Studies in Philosophy*, IV, ed. P. French, Minneapolis, University of Minnesota Press.

Darwall, Stephen. 1983) *Impartial Reason*, Ithaca; Cornell University Press.

Davidson, Donald. (1984) "Radical Interpretation," in *Inquiries into Truth & Interpretation*, Oxford; Oxford University Press: 125-140.

Dreier, James. (1990) "Internalism and Speaker Relativism," *Ethics*, 101, No. 1: 1-26.

Ewing, A. C.. (1953) *Ethics*, Macmillan.

Falk, W. D.. (1947-8) "Ought and Motivation," *Proceedings of the Aristotelian Society*, 48: 492-510, reprinted in Falk (1986).

_____ (1986) *Ought Reason and Morality*, Ithaca; Cornell University Press.

Gibbard, Allan. (1990) *Wise Choices, Apt Feelings*, Cambridge; Harvard University Press.

Grice, Paul. (1989) *Studies in the Way of Words*. Cambridge, MA; Harvard

University Press.

Hare, R. M. (1952) *The Language of Morals* (Oxford; Oxford University Press.

Harman, Gilbert. (1986) *Change in View*, Cambridge; M.I.T. Press.

Horgan, T. & Timmons, M. (1992) "Troubles For New Wave Moral Semantics: The "Open Question Argument' Revived," *Philosophical Papers* XXI: 153-175.

Johnson, Robert. (1997) "Reasons and Advice for the Practically Rational," *Philosophy and Phenomenological Research*, LVII, No. 3: 619-625.

_____ (1999) "Internal Reasons and the Conditional Fallacy", *The Philosophical Quarterly*, 49: 53-71.

Kalderon, Mark. (2004) "Open Questions and the Manifest Image," *Philosophy and Phenomenological Research*, 68: 251-289.

King, Jeffrey. (1995) "Structured Propositions and Complex Predicates," *Noûs*, 29: 516-535.

_____ (1998) "What Is a Philosophical Analysis?" *Philosophical Studies* **90:** 155–179.

_____ (2007) *The Nature and Structure of Content*, Oxford; Oxford University Press.

Korsgaard, Christine. (1986) "Skepticism About Practical Reason," *Journal of Philosophy*: 1-23.

Lewis, David. (1983) "Radical Interpretation," in *Philosophical Papers, Volume 1*,

Oxford; Oxford University Press:108-121.

Lewis, David. (1989) "Dispositional Theories of Value," *The Proceedings of the Aristotelian Society*, Supplementary Volume 63, 113-137.

Parfit, Derek. (1984) *Reasons and Persons*, Oxford; Oxford University Press.

Prichard, H. A. (1912) "Does Moral Philosophy Rest on a Mistake," *Mind*, xxi, no. 81, reprinted in Prichard (1949): 1-17.

Prichard, H. A. (1932) "Duty and Ignorance of Fact" The Henrietta Trust Lecture, reprinted in Prichard (1949): 18-39.

Prichard, H. A. (1949) *Moral Obligation*, Oxford; Oxford University Press.

Salmon, Nathan. (1991) *Frege's Puzzle*, Atascadero; Ridgeview.

Schroeder, Mark. (Forthcoming) "Having Reasons," *Philosophical Studies*.

Sidgewick, Henry. (1907) *The Methods of Ethics*, 7th edition. MacMillan and Company. (Reprinted by Dover in 1966).

Smith, Michael. (1994) *The Moral Problem*, Cambridge; Blackwell.

_____ (1995) "Internal Reason," *Philosophy and Phenomenological Research*, 55:109-31.

Soames, Scott. (2002) *Beyond Rigidity*, Oxford; Oxford University Press.

Stanley, Jason & Szabó, Zoltan. (2000) "On Quantifier domain restriction," *Mind & Language,* 15.

Stocker, Michael. (1979) "Desiring the Bad," *The Journal of Philosophy*: 738-753.

Svavarsdottir, S. (1999) "Moral Cognitivism and Motivation," *The Philosophical Review*, 108: 16-219 .

Thau, Michael. (2002) *Consciousness and Cognition*, Oxford; Oxford University Press.

Thomson, Judith Jarvis (1986) *Rights, Restitution, and Risk*, Cambridge; Harvard University Press.

Twain, Mark. (1885) *The Adventures of Huckleberry Finn*, New York; Charles Webster & Co.

van Roojen, Mark (1995) "Humean Motivation and Humean Rationality," *Philosophical Studies*.

_____ (2000) "Moral Rationalism: a Somewhat Less Idealized Account," *Philosophical Quarterly*, 199: 233-241.

_____ (2006) "Knowing Enough To Disagree: A New Response to the Moral Twin Earth Argument," in Russ Shafer-Landau (ed.) *Oxford Studies In Metaethics*, Volume 1 Oxford; Oxford University Press, 161-193.

Wedgwood, Ralph (2004) "The Metaethicists Mistake," *Philosophical Perspectives*, 18: 405-426.

_____ (2006) " ," in *Oxford Studies in Metaethics*, Volume 1, Oxford; Oxford University Press.

_____ (2008) *The Nature of Normativity*, Oxford; Oxford University Press.

NOTES

2. The most widely noted examples include David Brink's amoralist (1989, pp 46ff.), Michael Stocker's aging politician (1979), and Gideon Rosen's Sadists example from Dreier (1990).

3. What I'm saying here parallels what Allan Gibbard says as he introduces the notion of rationality that he will go on to analyze (1990 pp. 6-7).

4. Prichard (1912), Falk (1947-8).

5. See Hare (1952), Dreier (1990) and Horgan & Timmons (1992). I discuss the latter paper and its internalist upshot in greater depth than here in my "Knowing Enough to Disagree," (2006).

6. Dreier (1990, at p. 14) defends the principle that necessarily in normal contexts a person will have some motivation to promote what he believes to be good. If we make the relevant changes to talk about rightness rather than goodness, this principle is close to the internalist truth that I wish to explain. The main difference (aside from the substitution of rightness for goodness) is that my favored internalist datum only refers to rational people, a stipulation favored by Korsgaard (1986), and Smith (1994). Since rational people are a subset of people, the general claim entails the claim about rational people. Normalcy here is not just statistical . Rather the claim is that we can only interpret an unmotivated speaker as sincerely accepting an ethical claim against a background in which other people with whom she is in contact are generally motivated. Thus unmotivated believers in ethical claims must be abnormal

within the relevant context.

7. These examples are from Brink (1986, at 30). Dreier (1990) presents essentially the argument I'm making here in discussing Gideon Rosen's sadists.

8. I'm being intentionally vague here about the exact nature of the claimed identity because most of the arguments of this paper could be used by a supporter of any number of more specific rationalist theses. My own current view is that 'is rational' and 'is right' contribute the same property to the literal meaning of sentences embedding them. I recognize the simpler view has drawbacks and the reasons I favor it are to complex to discuss here.

9. This section of this paper draws heavily on my "Motivational Internalism: a Somewhat Less Idealized Account," (2000).

10. This sort of argument has been a theme in Gil Harman's work on rationality for some time. See, Harman (1986, chapter 2).

11. It may be worth emphasizing that we can rank departures from rationality in terms of seriousness, even when they involve breaches of different rational norms. For more on this see van Roojen, (1995, 2000).

12. Philosophers have long used the terms of art 'objective' and 'subjective' to mark distinctions of roughly this sort. Parfit (1984, 25) makes a distinction both within rationality and without it, one based on actual effects and the other expected effects. He distinguishes objective and subjective notions of 'ought'. 'right' 'good' and

'bad', but only use the word 'wrong' to indicate the subjective sense of what we should avoid aiming at. Ewing (1953, p. 63) distinguishes two senses of 'ought'. One corresponds roughly what one ought to do relative to the actual facts, on the assumption that one knew them. The other is what one should do based on an assessment of the likely consequences where likelihoods seem to be based on what is epistemically available to the person. Later (page 127) he distinguishes two senses of ought where one sense tracks what one believes one ought to do and the other is based on what one should think one ought to do had one assessed the facts and their import correctly. He also claims that there are correspond to these two senses of 'right', 'wrong' 'duty' and so on. Mark Schroeder has a nice discussion and defense of Ewing's sort of view in Schroeder (forthcoming). H. A. Prichard (1939) draws a line in a similar place in his discussion of duty, before opting for the subjective view. Sidgewick distinguishes "objective" and "subjective" senses of right and wrong where one refers to what one ought in fact to do given how things are, and the other refers to what one believes one should do. (Sidgewick 1907, Book III, Chapter 1, Section 3.) Not all of these formulations are equivalent.

13. I've been following the tradition by speaking of "objective and subjective senses of rationality." One might be skeptical of distinguishing such senses. (See, for example, Thomson 1986, p. 179.) The important point I need for my argument is not really one about the meanings of the words in question. What I need for my

argument is the idea that what makes sense for an agent to do depends on features of the person's situation in such a way that changes in their information, self-control, available time and so on will effect what it makes sense for them to do. And I need it to be the case that when people are thinking about what it is rational for an agent to do, they allow these sorts of features to effect their verdicts about what is rational. Whether that is because the word 'rational' genuinely has two meanings, or whether features of the context of utterance of evaluation allow us to use a term with one univocal meaning to convey different information in different contexts does not deeply effect the most important issues here. This is not to say that the appropriate account of the semantics that allows us to explain this is uninteresting; it is highly interesting, and I do think that a univocal account is better. But a number of different accounts would deliver what I need for the argument here.

14. You might think that this notion could be collapsed into the previously stipulated subjective notion by saying that one of the actions the agent should consider choosing is one in which the agent investigates further. Subjective rationality would then always be a matter of maximizing expected good, where one of the options that maximizes the good involves investigating further before choosing. But that's too simple. We can separate the question of what I should do given the evidence I actually have, from the question of what I should do given the evidence I should have had (had I only investigated) yet these may both differ from what I should have

61

chosen had I only had full information.

15. In light of this we could use the terms 'subjective' and 'objective' to draw the relevant contrasts in couple of different ways. We might reserve 'objectively rational,' and its kin for only those reasons an agent would have if fully ideal and stipulate that the remaining reasons which depend for their grounds on limitations of the agent all count as subjective reasons. We could then distinguish among these by noting that some, those less dependent on limitations of the agent, are less subjective and more objective than others. Or we could adopt a somewhat looser way of speaking, one which allows the context to indicate which contrast we are intending to draw. I don't think there is anything of substance that turns on how we choose to talk about this, so long as we are clear in what we are saying.

16. An impressive example is Michael Smith's *The Moral Problem*, which argues for a particular analysis of moral judgements by suggesting that it would make sense of rational changes in motivation when people change their moral beliefs. The similarity to the present argument should be striking (and the causal influence obvious). One reason I think that argument is unsuccessful is precisely for moving too quickly over the complications I'm trying to emphasize in this paper.

17. The general phenomenon that competent speakers and thinkers can be ignorant of identities is widely noted in discussions of the Open Question Argument. See, Lewis (1989, pp. 125 ff.), Smith (1994, p. 35-39) and Kalderon (2004). But

considerations of the sort highlighted by Frege's Puzzle are not usually noted in discussions of internalism. Wedgwood (2004, 2008) has proposed using Fregean senses to do the needed work. He postulates modes of presentation for moral properties and relations which are linked with various moral terms (2004, p. 417) and which are essentially action-guiding (pp. 418-420). He uses these modes to explain internalism. Wedgwood's proposal seems still to have problems with rational amoralism insofar as anyone using the relevant moral terms is assumed to be using them with their action-guiding sense. Finally my own (2006) develops a Millian account of reference determination for moral terms and offers some suggestions about how this will explain the sort of internalism postulated in this paper.

18. My exposition here benefits from having read Kalderon (2004), Soames (2002), Thau (2002), and Salmon (1991). I've also been helped tremendously by extended email and discussion with Mark Kalderon about the significance of Frege's puzzle for the analysis of moral terms.

19. I don't have room to argue for that in any detail here. My doubts about Fregean views of these property terms are just that its unlikely that there is any special way of thinking about the relevant properties associated with the words across speakers that differ systematically with the expressions used. Yet we seem justified in attributing the relevant beliefs based on sincere avowals using the terms, other things equal. To have either concept a speaker will have to recognize that the terms are

used to evaluate actions, but this is common to the two expressions and hence cannot constitute a sense that picks out one versus the other.   Still, the main argument here remains neutral between Fregean and Millian treatments.

20.It may be rational to believe this because it is far more likely that a randomly chosen number in would be not prime than prime, even if the primes and non-primes are equinumerous.

21. Someone may worry this generates too many different relativized notions of rightness and that this would make it hard to explain how we can say what we want to say and how we know what we are saying. In fact I think there is no more problem here than there is for the corresponding notions of rationality which are similarly relativized. And I think we should recognize that the claims I make about rationality are true, and also that in ordinary talk we have little trouble figuring out what claim is at stake in a given conversational context. If, in a context where time is short, I say it is rational for us to just arbitrarily choose an option, the salience of the shortness of time plus the fact that given more time we would be able to make an optimal choice makes it obvious that the sort of rationality I have in mind is one relativized to the time we've got. If I say that it is right to choose arbitrarily, the same features of the conversational context narrow the choice of interpretations in the same way.

22. Talk about different senses of a term can be a bit sloppy and I want to retain that

sloppiness here because I think that much of what I say remains true no matter how we fill out the details. Consistent with the idea that 'is rational' can capture any one of the more subjective or objective "senses" of rationality here postulated will be a number of different semantic proposals for explaining exactly how we can get across which sort of rationality is at stake. On one approach the terms in question which admit of different relativized "senses" have one core sense which is their literal meaning, either on the fully objective or the fully subjective end of the spectrum. Still, various features of the context could allow us to imply things we don't literally say by employing Gricean mechanisms to make clear that we are trying to communicate related facts about one of the other notions of rationality. Another approach would suggest that 'is rational' has a literal meaning which contributes an incomplete relational property to an utterance. The information so conveyed is then completed (again via various Gricean mechanisms) by the context. For example, if due to features of the context we cannot really mean to say that something makes sense to do given the actual facts, we will interpret the intended claim as being made relative only to what the agent could take account of. And still further views might treat the term like an indexical with features of the context determining which of several candidate contents is being literally expressed. For some relevant discussion regarding different options in other domains of discourse see, Bach (1994), Stanley and

Szabó (2000), and of course Grice (1989).

23. There is still the possibility of a gap since ideal evidence might be misleading. In normal cases if you think some claim is true, you should think that better evidence about it will continue to justify you in accepting it.

24. This ice cream example is parallel to Michael Smith's ill-tempered squash player (borrowed from Watson). Both have reasons they would not have if they were fully rational. (Smith 1995) This may seem to cause a problem with one intuitive idea about reasons, that in order to both justify and explain actions appropriately, reasons must be such that rational agents would act as they suggest. Since completely rational people would not be in the states that ground reasons of this sort it looks like this internalist idea about reasons confronts the "conditional fallacy" and must therefore be abandoned. (Johnson, 1999, van Roojen, 2000) This is why Smith abandons this intuitive formulation of internalism and replaces it with the idea that a fully rational version of the agent would want that the agent do what she has reason to do in those conditions. (Smith, 1995) I have argued elsewhere (2000) that we need the different relativized notions of rationality I'm employing here to capture the explanatory and justificatory aspects of reasons for action in these sorts of examples while escaping the conditional fallacy.

25. Jeffrey King's (1995, 1998, and 2007) account of philosophical analysis is motivated partly by the account's ability to avoid the paradox of analysis. King's

idea is that most such analyses can be captured by biconditionals relating structured propositions. Such biconditionals can be informative because the propositions represented on either side of the 'iff' either have different structures or different constituents at different points within those structures. For example, it can be informative to find out that something is a vixen just in case it is a female fox because the proposition that something is a female fox is a structured proposition with a component of that structure representing a complex whereas the proposition that the same thing is a vixen has a simpler structure with vixenhood as the corresponding component. If the right version of rationalism is one on which the property of being morally right reduces to the property of being rationally right/required based on certain grounds, rationalism can just take King's ideas on board. But there may be no principled way to divide moral grounds from non-moral grounds so that the best overall account will reduce moral rightness to rational rightness. If that's the case we will still have Frege puzzles but no difference in structure or in the components to do the work King puts them to. Millians have to treat such cases in whatever way they treat the analogous story about groundhogs and woodchucks, or Hesperus and Phosphorus.

26. The example is obviously due to Burge (1979). The special role of experts in Burge's example makes the analogy I am drawing a bit strained, but since this is the most well-known example of the general phenomenon I am trying to take advantage

of,it has expository advantages due to its familiarity.

27. This is not quite what Ralph Wedgwood (2006, 2008) proposes. It looks to me like Wedgwood requires that the term play the right action guiding role in an individual's psychology to count as expressing the right thought. That way of going makes it hard to explain the moderate internalism supported by the sadists example.

28. I recall reading a helpful analogy to the ability to recognize a house from different angles. One may without irrationality think that "that house" is white, and also think that "that house" is old, without recognizing that there is a single old white house of which one is thinking. Yet the content of 'that house' in the two thoughts may be identical. I have searched and searched for the source but can't find the example anywhere, though I continue to suspect it was in Soames (2002).

29. This round about way of picking out the relevant state of mind is needed because on Millian views thinkers in a position to think about morality or practical rationality will already believe the propositions that morality is morality and this proposition is also the content of the thought morality is practical rationality. The point of the position is that this is not irrational if one fails to know they have the same content. But Millians can still think that once one knows what one would normally express with the relevant identity claim, one is then rationally committed to treating the things picked out by the terms in exactly the same way, since they are (as one now recognizes) one and the same.

30. It is important to keep in mind that Frege's puzzle and the complications it generates will effect the relations between any judgements with this content expressed using any form of words you like. A speaker could be unaware that 'rational' or 'makes sense to do' designate this property and hence remain unmoved by judgements made using those forms of words, just as much as when she makes a judgement with that content using the words 'morally right'. I thank Nadeem Hussain for emphasizing this Millian commitment to me, and Mark Schroeder for helping me see its upshot for my view.