Firth and Hill: Two Dispositional Ethical Theories

Margaret Chiovoloni

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Philosophy.

Chapel Hill
2006

Approved by
Advisor: Thomas E. Hill, Jr.
Reader: Gerald F. Gaus
Reader: Geoff Sayre-McCord
Dispositional accounts of morality seem a promising way to capture certain central features of morality. While Roderick Firth and Thomas Hill, Jr. each offer accounts of this sort, the result in each case is very different. In this paper, I will describe both views. I will criticize Firth for making reactions the central moral notion, and describe how that limits the extent to which his theory can explain moral motivation. I will show how Hill’s account can solve this problem, but argue that his view is problematic in requiring ideal legislators to be ignorant of their life situations, for this move limits the extent to which his theory can explain moral motivation. I will offer a suggestion that I think his account can easily accommodate, and briefly discuss how the resulting view is superior to Firth’s.
ACKNOWLEDGMENTS

I would like to thank Tom Hill for being a wonderful advisor, and David Landy for his untiring help and support.
TABLE OF CONTENTS

Two Dispositional Ethical Theories................................................................. 1

Bibliography........................................................................................................ 28
Dispositional accounts of morality seem a promising way to capture certain central features of morality. While Roderick Firth and Thomas Hill, Jr. each offer accounts of this sort, the result in each case is very different. In this paper, I will describe both views. I will criticize Firth for making reactions the central moral notion, and describe how that limits the extent to which his theory can explain moral motivation. I will show how Hill’s account can solve this problem, but argue that his view is problematic in requiring ideal legislators to be ignorant of their life situations, for this move limits the extent to which *his* theory can explain moral motivation. I will offer a suggestion that I think his account can easily accommodate, and briefly discuss how the resulting view is superior to Firth’s.

Firth presents a dispositional absolutist ethical theory. While his theory is an analysis of the meaning of ethical statements, he does seem to want to commit himself to certain ethical views, and to this extent it will be useful to discuss the particular features of his account.

Firth’s motivation for adopting a dispositional (and absolutist, although that isn’t my concern here) account of ethical statements seems to be that other approaches are open to objection. He takes the fact that his analysis succeeds in key respects as reason to accept it.¹ He writes, explaining his attraction to dispositional accounts,

There is one kind of analysis of ethical statements, however, which has certainly not been examined with the thoroughness that it deserves—the kind of analysis, namely, which construes ethical statements to be both absolutist and dispositional…. [T]his kind of analysis seems to be capable of satisfying the major demands of certain schools of ethical thought which are ordinarily supposed to be diametrically opposed to one another. (Firth, 318)

¹At least, this seems to be the best way to understand the text.
Later in the paper, he gives reasons for adopting the particular dispositional account that he does. He makes his account an *absolutist* dispositional one so as to avoid objections commonly posed to relativist dispositional accounts (these objections are not our concern here). Furthermore, his account “construe[s] ethical statements…as assertions about the dispositions of all *possible* beings of a certain kind (of which there might in fact exist only one or none at all)…” (Firth, 320) to avoid serious objections posed to accounts that ethical statements as assertions about the dispositions of all actual beings, or as assertions about the dispositions of a majority of actual or possible beings. (Again, these objections are not our concern here.)

Firth claims that the meaning of ethical statements can be expressed by statements about the dispositions of ideal observers. He writes,

> If it is possible to formulate a satisfactory absolutist and dispositional analysis of ethical statements, it must be possible, as we have seen, to express the meaning of statements of the form “x is right” in terms of other statements which have the form: “Any ideal observer would react to x in such and such a way under such and such conditions” (Firth, 329).

Firth’s view (which he gestures at above) is that ethical statements have the same meaning as corresponding statements about the reactions of ideal observers. Out of this view, and certain plausible assumptions about meaning, an ethical view of sorts can be extracted. Such a view would claim that an action is right iff any ideal observer would react to it “in such and such a way under such and such conditions,” i.e. with approval, and an action is wrong iff any ideal observer would react to it with disapproval. It is important to note here that the rightness or wrongness of an action depends on the *reaction*, and not the judgment, of the ideal observers. This point will be of some importance later.
Firth describes in great detail the characteristics of these ideal observers. An ideal observer is omniscient, omnipercipient, disinterested, dispassionate, consistent, and “in other respects he is normal” (Firth, 344).

Firth stipulates that his ideal observers are “omniscient with respect to non-ethical facts” (Firth, 333). He characterizes them in this way for two related reasons. Firstly, he notes that in our everyday moral judgments, we think knowledge of the facts of the case help our judgments approach the truth. He writes,

We sometimes disqualify ourselves as judges of a particular ethical question on the ground that we are not sufficiently familiar with the facts of the case, and we regard one person as a better moral judge than another if, other things being equal, the one has a larger amount of relevant factual knowledge than the other (Firth, 333).

So, an ideal observer must at least have knowledge of the relevant facts.

Secondly, Firth notes that his analysis cannot define the ideal observers as possessing only relevant knowledge.

To say that a certain body of factual knowledge is not relevant to the rightness or wrongness of a given act, is to say, assuming that an absolutist dispositional analysis is correct, that the dispositions of an ideal observer toward the given act would be the same whether or not he possessed that particular body of factual knowledge or any part of it. It follows, therefore, that in order to explain what we mean by ‘relevant knowledge,’ we should have to employ the very concept of ideal observer which we are attempting to define (Firth, 333-4).

In other words, knowledge is relevant to the morality of an action if and only if the knowledge affects the reaction of an ideal observer. But, in this case, to define an ideal observer as having only relevant knowledge would render the account of ideal observers as circular. To avoid this difficulty, Firth grants omniscience to the observers. A clarification: while Firth characterizes the observers as omniscient with respect to non-ethical facts, he writes that he does not mean to
exclude the possibility that they have knowledge of ethical facts, but rather that they do not need to possess such knowledge for their reactions to be appropriate.

Firth also describes the ideal observers as omnipercipient. By “omnipercipience,” he means perfect power to imagine what others feel.\(^2\) He includes this characteristic because he thinks we ordinarily regard this ability as morally important. He writes,

> We sometimes disqualify ourselves as judges of certain ethical questions on the ground that we cannot satisfactorily imagine or visualize some of the relevant facts, and in general we regard one person as a better moral judge than another if, other things being equal, the one is better able to imagine or visualize the relevant facts (Firth, 335).

So, it seems, omnipercipience is a way to make the omniscience of the ideal observers even more useful.

Firth also describes the ideal observers as being disinterested. In elaborating what he has in mind, he denies meaning that the ideal observers cannot consider the individuals involved in the action, because this would rule out, for example, the possibility that relationships between people can affect the rightness or wrongness of an action. He writes,

> But to analyze ethical statements by reference to this kind of impartiality, would rule out, by very definition of the words ‘right’ and ‘wrong,’ the moral theory...that the rightness or wrongness of an act is determined in part by irreducible obligations arising directly from certain personal relationships; such an analysis would entail, for example, that there is never any moral justification, other things...being equal, for making a decision which favors one’s mother or friend or creditor at the expense of a greater benefit to someone else (Firth, 337).

Rather, he has in mind something somewhat different. He writes,

> Now it seems to me that a large part of what we mean when we say that an ideal judge is impartial, is that such a judge will not be influenced by interests of the kind which are commonly described as ‘particular’—interests, that is to say, which are directed toward a particular person or thing but not toward other persons or things of the same kind...(Firth, 337).

\(^2\)He notes that, on some uses of the term “omniscience,” omnipercipience may be included.
His discussion here is quite complicated, and not worth rehearsing here. It is sufficient to note that he means merely that the reactions of the judges won’t be *partial* in a way that we would take to conflict with morality. That is, it seems that morality is essentially *impartial*, and the characteristic of disinterestedness is meant to capture that quality in the ideal observers. The smaller details of this characteristic are not relevant to the rest of this paper.

Firth’s ideal observers are also dispassionate. He writes,

> The concept of impartiality cannot be exhaustively analyzed in terms of interests, for an impartial judge, as ordinarily conceived, is a judge whose decisions are unaffected not only by his interests, but also by his emotions (Firth, 340).

By introducing this requirement in terms of impartiality, Firth seems to imply that the problem with emotions is that they hinder the impartiality of a judge. Indeed, he confirms this, by writing,

> And we can say that an ideal observer is dispassionate in the sense that he is incapable of experiencing emotions of this kind—such emotions as jealousy, self-love, personal hatred, and others which are directed towards particular individuals as such (Firth, 340).

So, Firth’s ideal observers are dispassionate in the sense that they do not have “particular” emotions, that is, emotions directed towards particular individuals.

Firth’s ideal observers are consistent. However, consistency is not some quality *required* of the observers. Rather, as he notes, consistency falls right out of an adequate analysis of their other characteristics. He writes,

> …an ideal observer will indeed be consistent if an adequate dispositional analysis can be formulated; but his consistency will be a derivative characteristic—a consequence of his other characteristics together with certain psychological laws (Firth, 344).

So, given that the other characteristics of the ideal observers are all meant to eliminate the sources of disagreement, consistency will result from this elimination.
Firth also describes his ideal observers as “normal.” He writes, “in other respects [the ideal legislator] is normal,” and continues,

An examination of the procedures by which we attempt to decide moral questions, reveals that there are a great many conditions which we recognize, though not always explicitly, to be favorable or unfavorable for making valid moral judgments (Firth, 344).

He further writes,

For however ideal some of his characteristics may be, an ideal observer is, after all, a person….Most of us, indeed, can be said to have a conception of an ideal observer only in the sense that the characteristics of such a person are implicit in the procedures by which we compare and evaluate moral judges, and it seems doubtful, therefore, that an ideal observer can be said to lack any of the determinable properties of human beings (Firth, 344).

Firth isn’t explicit about the boundaries of this category. I think he uses it, in part, as a way to include many minor important aspects of being a “moral judge” (as he says). The examples he gives (such as enough exercise, lack of exhaustion, and lack of distraction) seem to imply that the features he means consist mostly in requiring the presence of appropriate conditions (and he says almost as much).

However, I think he uses it mostly to round out the ideal legislators. That is, he gives as criteria for ideal observers the possession of disparate traits, and these traits do not really have a unity. It also is not clear exactly how the various traits interact with each other, influence each other, and so on. The requirement for normality is, I think, a convenient way to unify these traits. It is simpler, it seems, than spending an inordinate amount of time describing the way that all of these traits interact with each other, a task that may turn out to be impossible anyway.

This “rounding out” may also be necessary to get substantial results.³ Since, besides the attribution of the quality normal, Firth only attributes to them many abstract qualities, it isn’t

³Thanks to Tom Hill, Jr. for pointing this out.
clear what about *them* would cause them to react in particular ways, apart from the fact that they are otherwise normal, that is, otherwise normally human. In other words, Firth has characterized the ideal observers so that they have perfect, undistorted perceptions of the *case*, or of the action. But apart from this normality, there is nothing in them that would do the reacting, and there is nothing in them, as far as I can see, that would cause them to react in one way rather than another.

Who are these ideal observers? There is little evidence, in Firth’s account, that the ideal observers are us. It is true that (on his account) the fact that we take our reactions in ideal situations to be evidence for the fact that an action is right or wrong is supposed to ground the fact that what is right or wrong depends only on the reactions of ideal observers (we will discuss his account in much greater detail, below). Furthermore, it is true that the ideal observers are supposed to be human, as he explains when he describes them as normal in other ways. However, Firth does say that they are better versions of particular real agents. They are generic, ideal humans, and not particular humans made better. I do not mean to imply that Firth’s account precludes that the ideal observers are us, but that he does not emphasize this feature, and even if they are us (as opposed to generic ideal humans), this fact does not play a central role in his theory. This feature of his account will be important later.

Let us consider more carefully the ethical theory that can be extracted from Firth’s analysis of the meaning of ethical statements. Above, I wrote that this ethical theory is as follows: an action is right iff any ideal observer would react to it with approval, and an action is wrong iff any ideal observer would react to it with disapproval. Let us note that this formulation needs clarification: three inconsistent claims are compatible with it. It will be useful to discuss which readings *not* to attribute to Firth, in order to distinguish them from what *should* be attributed to him.
In the first possible reading, this formulation could leave open the possibility that there is some underlying criteria for rightness or wrongness. For example, it may be the case that properties x, y and z each make an action right, and the ideal observers react to those properties with approval. In this case, it still would be true to say that an action is right if any ideal observer would react to it with approval, since the ideal observers would always react with approval to actions with properties x, y or z. Here, though, it is the features of the action, if this is the case, then, that explain the rightness of the action, and not the reactions of the observers.

In the second possible reading, this claim could mean that the (hypothetical) reactions of the ideal observers make the action right. That is to say, the reactions of the ideal observer are what make the action right, and the reactions explain the rightness. The features of the action, in this case, are not what make the action right or wrong, except via the reactions of the observer: perhaps feature x led the ideal observers to react with approval. In this mediated sense, we can say that feature x makes the action right. Even so, however, the important fact is the reaction of the observers, and not the features of the action. Now, of course, even if the reactions are what determine the morality of the action, there can still be consistency among the features of right actions, and among the features of wrong actions, respectively. As we saw earlier, consistency results from the other qualities of the observers. The point, rather, is that (even though there are consistent features among right actions, or among wrong actions) the features are not what make the actions right, but rather, it is the fact that they are actions that elicit approval from ideal observers that makes them right.

The third possible reading is only subtly different from the second reading. On this reading, while it may be true to say that the hypothetical reactions of the ideal observers make the action right or wrong, it is misleading to say only this much. Rather, on this reading, it is the reactions
of the ideal observers that separate really right (or wrong) actions from only apparently right (or wrong) actions. This reading is explained in greater detail, below.

Let us consider the first reading: that while it is true that ideal observers would react with approval to all right actions, it isn’t the reactions of the observers, but rather the features that lead to such reactions that make the actions right. If this reading is what Firth had in mind, he would have given us no explanation of morality, or of the rightness or wrongness of actions at all. In this case, his account could hardly be considered an ethical theory at all. Since (on this reading) he has not told us which features of morality would cause approval or disapproval in the hypothetical ideal observers, this kind of account wouldn’t explain what right actions have in common, qua right actions. His account would be helpful as a heuristic, but it would leave much unexplained.

However, this reading does not reflect Firth’s account at all. Consider his emphasis on the reactions of the observers (as opposed to their judgments). He writes,

Perhaps it would seem more natural to call such a being an ideal judge, but this term could be quite misleading if it suggested that the function of an ideal observer is to pass judgment on ethical issues. As an ideal observer, of course, it is sufficient that he be capable of reacting in a manner which will determine by definition whether an ethical judgment is true or false. And it is even conceivable, indeed, that an ideal observer, according to some analyses, should lack some of the characteristics which would make it possible for him to pass judgment on ethical issues—which would mean, of course, simply that he would not be able to judge the nature of his own dispositions” (Firth, 321).

Firth does not here explain his reasons for thinking that the reactions, and not the judgments, are important for the rightness or wrongness of actions. However, the last sentence of the above quotation hints at the reason for this feature of his account: if the judgments of the ideal observers are what matter for the rightness or wrongness of actions, then for plausibility’s sake the observers would need independent criteria on which to base their judgments. That is, the
very fact that judgment would be important would imply independent criteria for the rightness or wrongness of actions. On one plausible construal of judgments, a judgment just is the sort of thing that one produces upon considering some independent criteria. However, in that case, presumably, the rightness or wrongness could be expressed in terms of such independent criteria, and so reference to such ideal judges would be explanatorily redundant. Firth, on the contrary, seems to assert that the reactions of the observers are the only criteria for rightness and wrongness, and he expresses this claim when he writes that the judgments of an ideal observer would just be judgments about the nature of his dispositions and reactions.4 It is on this basis that I reject the first reading as a reading of Firth.5

Let us now consider the second reading of Firth. On this reading, remember, it is the fact (itself) that an action would be approved by an ideal observer that makes it right (or, it is the fact that it would be disapproved of by an ideal observer that makes it wrong). The reason, on this reading, that an action is right or wrong is that it would cause approval or disapproval in the ideal observers. However, this reading does not capture the similarity between his dispositional account of right and a sense-data theorist account of color. Only the third reading captures this similarity, one that Firth emphasizes.

The key to understanding Firth’s analysis is to notice the importance he places on the analogy between a dispositional account of yellow and a dispositional account of right. He writes,

4And the (reasonable) thought seems implicitly to be that the only two options are the reactions or the judgments of ideal observers.

5We find confirmation of this rejection of this reading in his discussion of the intuitionists. He argues that intuitionists use some conception of an ideal observer in distinguishing true and false intuitions, but he goes on to clarify: “But this fact does not make intuitionism any less dualist, of course, for Ewing and other intuitionists will maintain that in formulating these ideal conditions they are merely formulating a test for the validity of an ethical statement, and not an analysis of the statement” (Firth, 326). The fact that he distinguishes between a test and an analysis, and only attributes the test formulation to the intuitionists, is strong evidence against the reading that his analysis is a test.
And it is quite possible not only that the term "right" is [as yellow is] ambiguous, but also that in one of its senses it designates a characteristic of human experience (apparent rightness) which in some important respect is just as simple and unanalyzable as the property of apparent yellowness. And thus we might even decide by analogy with the case of 'yellow' that 'really right' must be defined in terms of 'apparently right'--i.e., that the experiencing of apparent rightness is an essential part of any ethically-significant reaction of an ideal observer (Firth, 324).

So, for Firth, “really right” is a species of “apparently right.” Once we connect this idea with the idea in the following quotation, we will be in a better position to understand his analysis. He writes,

It seems clear that these reactions, if the analysis is to be at all plausible, must be defined in terms of the kind of moral experience which we take to be evidence, under ideal conditions, for the truth of our ethical judgments (Firth, 326).

In other words, we ordinarily take our reactions, our experiences, to be evidence for our moral judgments. Therefore, the reactions of the ideal observers will need to be defined in terms of our reactions—they will need to be just like our reactions, but had under ideal conditions.

So, Firth’s picture appears to be as follows. We have certain experiences (which Firth calls “moral data”), and being the kind of action that causes those experiences is a necessary (but not sufficient) condition for being a right (or wrong, for the opposite experiences) action. It is important to note this fact—that while causing certain reactions is a necessary condition, it is not a sufficient condition. To be sufficient, this condition must be supplemented by the following condition; the conditions under which it causes those experiences are ideal conditions (i.e. the conditions described earlier, of omniscience, omnipercipience, and so forth). So, while the reactions of ideal observers do (in a sense) determine which actions are right and which actions are wrong, this determining is not causal: something's being right just consists in the fact that an ideal observer would react to it with approval. It is not that ideal observers react to action x with
approval, and thereby x is right (as was the case in the second reading of Firth). Rather, x's being right consists in the fact that ideal observers would react to it with approval.

On Firth’s view, our reactions (under ideal circumstances) limit what rightness can be, that is, they set the boundaries of the category [apparent and real rightness]. Real rightness is a specific kind of apparent rightness—all real rightness is also apparent rightness, but not all apparent rightness is real (it fails to be real if it would not appear right under the appropriate conditions). To weed out the mere apparent rightness that is included in our reactions, we have to weed out all reactions that are not had under ideal circumstances. While these reactions then determine which actions can count as right, this determining is not causal; something's being right just consists in the fact that an ideal observer would react to it with approval.

We are now in a position to consider Firth’s emphasis on reactions (as opposed to judgments). He thinks that ordinarily, we consider our moral experiences (which are our reactions) to ground our moral judgments. That is what he meant when he wrote (also quoted above),

It seems clear that these reactions, if the analysis is to be at all plausible, must be defined in terms of the kind of moral experience which we take to be evidence, under ideal conditions, for the truth of our ethical judgments (Firth, 326).

So, the thinking seems to be, if our moral judgments ground themselves in certain experiences, it is natural to think that the rightness or wrongness itself would be grounded in our reactions. And, the thinking seems to be, his account can further justify our care for reactions: we care about them, because, at least when had under appropriate circumstances (i.e. ideal circumstances), they determine which actions are right and which are wrong.

He considers an objection to the claim that our moral experiences are evidence for our moral judgments. The objection is that we only have those experiences as a consequence of the moral
beliefs we have from passing moral judgment. If it is true that all of our moral experiences are only consequences of our judgment and resulting beliefs, then it seems circular to define rightness and wrongness in terms of our reactions, because these reactions would only result from our judgments. To address this objection, he argues that not all of our moral experiences are mere consequences of our moral judgments. He writes,

The crucial question is whether it is possible to feel an emotion of moral approval toward an act when we are in doubt about whether it is right or wrong. And surely this is possible. It is not uncommon, for example, to find ourselves feeling moral approval toward an act, and then to begin to wonder whether our reaction is justified....At such times we may continue to experience the emotion of moral approval although in doubt about the rightness of the act (Firth, 328).

He succeeds in showing the epistemological possibility that moral approval is evidence for (and not always just a result of) moral judgment. Sometimes, he thinks, we feel moral approval or disapproval even when we have not passed moral judgment on an action. Because we sometimes have reactions that are not the result of our judgments, he thinks, it is not for the above reason objectionable to define rightness and wrongness in terms of moral experiences or reactions.

So, the reason Firth thinks we have for making reactions central both to our analysis of ethical statements and to our ethical theory is that we ordinarily consider our moral experiences (and reactions) to ground our moral judgments. However, we do not take our reactions, our moral experiences, to be evidence under just any conditions (just like we do not take our experiences of apparent yellow under just any conditions to be evidence for actual yellowness). We take our moral experiences to be evidence “under ideal conditions,” or, the more ideal the conditions, the more the experience can count as evidence. The importance of good conditions in this account further explains why the analysis must be in terms of ideal observers. The closer our conditions are to being ideal, the closer we are to being ideal observers. So, on Firth’s view, thinking that
our moral experiences are evidence for our moral judgments, under ideal conditions, is the same as thinking that the hypothetical reactions of ideal observers determine the truth of our moral judgments. The fact that we regard our experiences in good conditions to be evidence for our moral judgments is a good reason for thinking that the reactions of ideal observers determine the rightness or wrongness of actions.

I have several criticisms of this account. To begin with, while Firth does not explicitly take it as a desiderata that his account should explain moral motivation, an adequate account should, and his fails to. The problem lies in the fact that reactions are not as central to our moral thought as Firth claims they are. Both the fact that Firth explains the rightness and wrongness of actions in terms of reactions, and the fact that the important reactions are those of ideal observers leads to a gap between the strength of the reason to perform right actions and not to perform wrong actions his explanation of morality would provide, and the strength of the reason we take ourselves to have (and morality to provide) to perform right actions and not to perform wrong actions.

Firth uses what he took to be the centrality of reactions in our moral thought to explain why reactions might ground morality. However, I argue below that he has not given us sufficient reason to believe that reactions are of central importance to morality at all. If reactions are not of central importance to morality, though, any theory that treats reactions as central cannot provide an adequate explanation of the reasons we have to perform right actions and not to perform wrong actions. If reactions are not central, then the fact that something causes a reaction of approval or disapproval (in me or in ideal observers) does not seem to capture the strength of the reason I have to perform or to refrain from performing that action.
The fact that his account is in terms of generic ideal observers, and not in terms of versions of us, compounds this problem. Ordinarily, if we were to care about the reactions of anyone, it would be the reactions of those in our community of agents. By making morality about the reactions of independent, uninvolved observers, Firth makes morality seem as if it is independently imposed by some other being(s), and not about our interactions with each other. Besides altering the phenomenology of morality, this approach leaves unexplained what reason we have to conform our behavior in light of the reactions of this (or these) independent being(s).

As I said earlier, Firth’s explanation of the ethical importance of our reactions is unsatisfactory. The most compelling evidence he presents for the claim that ethical statements and morality are about the reactions of ideal observers is the experience of moral approval in cases where one has not yet formed moral judgment. However, this claim is controversial enough for us to resist, at least until we are satisfied there is no better alternative. While the experience he describes may well be familiar, I do not think it captures all there is to morality. I certainly do not think it is either familiar enough or central enough to ground a complete ethical theory.

We usually take moral thought to be propositional in nature. For example, we can argue that certain features of an action make that action right or wrong, and when we do so, we mean more than that those features caused us or would cause ideal observers to react in a certain way. We usually think that the features of an action contribute more directly to the rightness or wrongness of an action—we do not ordinarily take the relationship between the features of an action and its rightness or wrongness to be mediated by the psychology of some independent hypothetical observer.
Now, he might argue that we have reason to adopt a dispositional theory on the grounds that other ethical theories are open to greater objections. He may continue, a dispositional account of the sort he proposes cannot rely on the *judgments* of the ideal observers, as we saw earlier, because in such an account, such judgments would need to be made on the basis of independent criteria. If so, and if (as seems likely) a dispositional account must explain the rightness or wrongness of an action in terms of either the judgment or the reactions of ideal beings, then he might insist that we have reason to adopt a dispositional account that is in terms of reactions in virtue of the fact that we have reason to adopt a dispositional account. Remember, we were able to construct an argument on Firth’s behalf, that explained why the reactions and not the judgments of the ideal observers were important to morality, of the following sort: if the judgments of the ideal observers are what matter for the rightness or wrongness of actions, then it is plausible to think that the observers need independent criteria on which to base their judgments. That is, the very fact that judgment would be important would imply independent criteria for the rightness or wrongness of actions. In this case, then, it would not be the *judgments* that are doing the work, but rather this independent criteria that does so.

However, this response is more or less question-begging for the whole enterprise. That is, we accepted a dispositional account contingently, on the supposition that it would explain ethical statements (and morality) satisfactorily. However, if the account *does not* explain ethical statements, and morality, satisfactorily, then we are under no obligation to continue to accept the dispositional account. Firth’s dispositional account has not given a good enough reason for us to think that ethical statements, and morality, are about the *reactions* of ideal observers, over and above the reasons that led us to accept only *contingently* the dispositional account. For us to continue to accept his dispositional account, Firth would have to demonstrate (or we would have
to demonstrate within the confines of his theory) that the analysis really does capture what we mean by certain ethical statements, and that the analysis really does capture the heart of morality.

It would be better to have a theory that captured more of our intuitions about morality. We have just seen that he misses several of these intuitions. Firth may be right that there are some times when our moral experiences or our moral reactions inform our moral judgment. However, the features of actions that ground our judgments are much more central to morality than are our moral reactions. We usually think that the features of an action contribute more directly to the rightness or wrongness of that action, and we do not normally think that rightness just is a species of apparent rightness.

Firth seems to think that the fact that his analysis is dispositional legitimates the analogy between yellow and rightness. Remember, it is that analogy that leads him to think that real rightness is a species of apparent rightness. However, I will propose that it is possible to have a dispositional account that does not make that move, and yet does capture the fact that often the features of an action ground our moral judgments without being mediated by our reactions.

The kind of dispositional account I have in mind relies on a distinction between reasonable and moral (or morally right). Roughly, the idea is (in Firth’s terms): An action is right if any ideal observer would judge it to be reasonable under ideal conditions. Now, certain aspects of this idea can be improved on (for example, it is better to speak of principles than actions), but for now, this captures the basic idea. Now, let us consider an account that’s more in line with this approach.

Thomas Hill’s ethical theory is a Kantian ethical theory. We will discuss here his interpretation of the kingdom of ends formulation of Kant’s categorical imperative. In Hill's account, the kingdom of ends is a collection of ideally rational individuals who choose laws to
govern their behavior. They are bound by all and only the laws they legislate collectively, and these laws are the principles that make up morality. On his view, the rightness or wrongness of an action depends on whether the action accords with or violates the moral principles, or laws, that are decided on by these ideal members in the kingdom of ends.

It will be useful to elaborate on the specifics of this view. Firstly, the laws legislated in the kingdom of ends must be universal in form. This condition helps to prevent laws that are partial to particular individuals. There will be no laws that only apply to certain individuals, and there will be no laws that apply to all but certain individuals. Secondly, when the legislators decide on the laws that will govern their behavior, they “abstract from personal differences” (Hill, 1992: 60). He writes,

That is, they will disregard the various factors which distinguish individuals: for example, differences in appearance, height, weight, sex, race, family, heritage, special talents, social roles, and so on (Hill, 1992: 60).

This feature also seems intended to prevent partiality. Hill notes that this condition restricts the principles to those that are “universal in intent” (Hill, 1992: 60).

Thirdly, the members, when they decide on the principles, must decide on them as rational beings. Hill here elaborates on his use of the term “rational.” He writes,

From this we can infer that they think logically, that they make no rules without a reason, that they are legislating for persons whose private ends are more or less ranked, and that the set of rules they seek is consistent, never gives conflicting orders for the same situation, and is not self-frustrating or self-defeating (Hill, 1992: 60).

He further stipulates, here, that “they will follow the rules that they adopt” (Hill, 1992: 60). This stipulation, he notes, avoids the need to worry about practical considerations. 

---

6What follows is a summary of the conditions of legislation Hill gives (Hill, 1992: 59-66).

7One might worry about how literally one should understand this condition. Firstly, one might wonder whether to think that they legislate as if they will follow the laws. Such a difference may not amount to anything, although the
Fourthly, Hill stipulates that the members of the kingdom are autonomous.

This is Rousseau’s ideal that they make laws such that each man in obeying them is obeying only himself. For Kant, this implies that the legislators are not causally determined, or even sensuously motivated, to adopt the rules which they do (Hill, 1992: 61).

This condition ensures that we get principles that are necessitated by reason and nothing else. The principles are not contaminated by the legislators’ being moved by what we would take to be inappropriate forces.

Lastly, Hill claims that each member chooses principles that treat each member (including himself) as ends in themselves. He gives three ways in which this condition should be elaborated.

First, [the above implies] that each member grants to every other whatever rights and respect he is entitled to.…. Second, the legislators regard the rationality of each member as unconditionally and incomparably worth preserving, developing, and honoring. The set of traits Kant calls ‘rational nature’ or ‘humanity’ is not something which distinguishes one man from another but is something which men have in common and which marks them off from animals…. A third (and perhaps derivative) implication of the idea that members regard each other as ends in themselves is that they have a prima facie concern to see each person’s ends realized, or at least to ensure each person freedom to pursue his ends (Hill, 1992:.61).

Now that we understand how the principles are chosen, let us turn to the question of why the principles of morality are those that would be chosen in this way. It may seem obvious that moral principles are those that we would agree on collectively if we had no flaws, but it is worth investigating why these principles would be *moral* principles. Hill writes,

“as if” version does not lead to questions about whether these principles are principles to govern our behavior or not. Secondly, it seems possible that we want the legislators to legislate each law as if they will follow it, but that we want them to legislate other laws that take into consideration the fact that some laws may be broken. That is, it is plausible to suppose that there should be principles governing the behavior of those who have been wronged by the breaking of other moral principles. For example, suppose they legislate a moral principle against physical restraint. Further suppose there is a moral principle against stealing. We would, presumably, like for the principle against physical restraint to allow restraint in cases where another has stolen. However, such a modification would not be possible if the legislators assumed that the principle against stealing would never be broken.
Since the legislators in the kingdom of ends are meant to represent abstractly basic features of a reasonable attitude regarding moral rules, the idea is that one should accept the norms that would be adopted by legislators in the kingdom (Hill, 2000: 46).

He continues, shortly thereafter,

Assuming this ideal convergence of judgment, reasonable, conscientious persons could see the resulting moral rules as, in a sense, self-imposed, that is, not merely demands from others but as reasonable applications of their own deep commitments (Hill, 2000: 46).

So, there seem to be two closely related ideas. Firstly, Hill seems to be suggesting that the fact that these moral rules are reasonable means that we are compelled by reason to accept these rules. However, this cannot be all, because there seems to be a significant difference in the force of rules of reason and rules of morality (and, correspondingly, there seems to be a significant difference in weight between breaking rules of reason and breaking rules of morality). What Hill writes in his second quotation, above, is aimed at filling this gap. The answer seems to be in the person’s “own deep commitments.” That is, the reason there seems to be a difference between rules of reason and rules of morality is that rules of morality are the product of applying rules of reason to our commitments to others and their humanity.

This view provides the resources to correct much of what is problematic in accounts like Firth’s. In what follows, I will lay out ways in which I think a view like this can be filled in. Much of what I say, Hill is not explicit about, although I regard it as a virtue of his account that it can accommodate these ideas. While I certainly do not mean to put ideas that he would reject in his mouth, he may have had much of what I say in mind.

This approach has the virtue of combining the justification of moral rules with the explanation of the motivation to follow them. The moral rules are justifiably applicable to the agents bound by them because the rules are the result of applying rules of reason to the agents’ deep commitments. (We discussed earlier the commitment each of us has to others in the
community.) The agent implicitly accepts these rules as justifiable, because acting in accordance with them is required to satisfy his commitments. Therefore, the agent cannot reject these rules without risk of compromising either his reasonableness and rationality, or his commitments to others. (We will come back to this.)

Moral motivation can be explained similarly. An agent is motivated to follow these principles to the extent that he is reasonable and to the extent that he is committed to others. The force of moral principles in deliberation is explained by a combination of the force of principles of reason and the force one’s commitments have. In addition, the force of reason leads any reasonable agent to conform to the rules that it is justifiable to demand that he conform to (and it is in this way that justification and motivation are connected).

Before elaborating on the ideas I have just mentioned, it will be useful to note that this view captures the phenomenology that much of morality is about other people. That is, for me at least, it seems that the reason to refrain from performing a wrong action almost always comes down to how that action treats another. On this view, that fact is easily explained. One has a commitment to the value of other people, and often (although, I claim below, it isn’t necessary to have) a commitment to being a reasonable person. These commitments together constitute at least implicit acceptance of the principles that are decided on in the kingdom of ends.

Here is another way to think of it. Being reasonable means being willing to act in a way that satisfies one’s commitments. But once one realizes this, and given that one is committed to others, one has a further obligation to be reasonable about actions that involve others, because not being reasonable about this is a way of disregarding one’s commitments to others. Part of satisfying one’s commitments to others is being reasonable.
Note, however, that a robust rationality or reasonableness on the part of the agent is not necessary for this conclusion. Really, all one needs is a commitment to others, combined with some modicum of rationality—enough rationality to realize broadly what is required of one to satisfy one’s commitments, in order for one to accept implicitly being bound by the principles adopted in the kingdom of ends.

In addition, it may be possible to demonstrate that reason requires a commitment to others. For example, Hill and Kant think that each individual has immeasurable value. Hill writes,

Independently of talents, accomplishments, and social status, each person is to be regarded as having a special worth that conscientious agents must always take into account. This value is not derived from one’s being useful or pleasing to others, and it takes precedence over values that are contingent in those ways. More controversially, dignity is not value that can be quantified but is ‘without equivalent’ (Hill, 2000: 42).

If it is true that each individual has immeasurable value, then this fact may provide a reason to adopt such a commitment, depending on one’s view of reasons, and reason may require one to recognize this reason and so adopt a commitment to others. I myself find this approach particularly plausible.\(^8\) If it is possible to demonstrate that reason requires such a commitment, then either the commitment or reasonableness (or, even more weakly, a commitment to being reasonable) are sufficient starting places.

Now, one may wonder why we need a kingdom of ends. That is, one may wonder why it is not enough just to act in ways that reflect one’s commitment to others. Why do we need to act in accordance with principles that everyone would legislate, instead of just acting in accordance with principles that the person involved in my action and I would legislate? Well, to begin, acting in accordance with principles that were decided in the kingdom of ends is a way of

---

\(^8\)An approach less likely to yield this conclusion, but nevertheless perhaps one worth mentioning, could appeal to the practical value of adopting such a commitment. That is, if it is true that acting in ways that demonstrate a commitment to others will help one achieve one’s ends more fruitfully, then perhaps any agent who has ends has reason to adopt such a commitment to others.
satisfying one’s commitments to everyone all at once. Further, it is a way of giving substance to how one should fulfill one’s commitments. One has commitments to all others, and so the fact that everyone (or, rather, everyone’s ideal versions) would adopt certain principles together means that fulfilling one’s commitments to others requires one to act in accordance with such principles. The fact that ideal versions of everyone would accept them is the reason we have, via our commitment to them, to act accordingly. That is, we have commitments to everyone, and not just to one individual. If we only consider one individual at a time, however, we risk satisfying the commitment to one while at the same time neglecting another. By acting only on principles that would be accepted by everyone, we prevent this situation from arising.

Now, in what I have written, I have assumed that the ideal legislators are ideal versions of each of us. It is time to discuss just what this means and whether it is possible. The ideal legislators need to be ideal versions of us, and recognizably so, for two reasons: Firstly, the fact that an ideal version of me would accept these principles is ultimately what grounds the fact that I have reason to act in accordance with these principles in particular. One’s commitment to others means (as we saw earlier) that one has a commitment to acting reasonably toward others. But, this means that one has a commitment to acting as one’s ideal version would act, that is, in accordance with the principles decided by one’s ideal version. When I commit to being reasonable, I commit myself to being me-reasonable—I will be reasonable “in my shoes.” And so when I am committed to being reasonable, this amounts to a commitment to adopt the principles for action that my ideal version would adopt in the kingdom of ends (we will come back to this point). Secondly, the versions of others need to be recognizable versions of them, or else acting in accordance with these principles is not really fulfilling my commitment to others, it is rather fulfilling some commitment to abstract agents, a commitment I may well lack.
As we saw earlier, Firth’s ideal observers are not recognizably us. I said earlier that his ideal observers are generic ideal humans, and not particular humans made better. We are now in a position to see why it is a problem that they are not recognizably us: because much of the motivation to be moral seems to depend on an actual commitment we have to others, an ethical theory that does not take advantage of this fact leaves moral motivation unexplained.

Hill, in contrast, seems to respect the idea that these ideal legislators need to be recognizably us, as is evidenced in his respect for the force of the following objection: “Any two-level theory that calls for impartial thinking…alienates the living agent from what allegedly gives authority to moral rules” (Hill, 2000: 54). And, he criticizes rule-utilitarianism on the following grounds:

Seeing rules as prescribed from an alien perspective would not help us to see why we count them important, that is, why it makes sense for us as conscientious agents to accept that we must constrain ourselves by them (Hill, 2000: 55).

He characterizes the relation between the real and the ideal versions of us as follows:

We are to think of each member of the kingdom of ends as having a set of personal ends. Since we must ‘abstract from personal differences,’ however, we can know nothing of the particular nature of these ends. Members of the kingdom, for example, are not thought of as power-seekers, peace-lovers, money-grubbers, or bird-fanciers (Hill, 2000: 59).

Here, he explicitly connects the ideal versions with the real versions of us when he writes that the ideal versions have a set of “personal ends.” Making this connection is one way in which his theory is superior to Firth’s.

However, one might worry about his stipulation that the decisions must be made in ignorance of the particular features of the agent. One’s commitment is to being reasonable—but that means one must act reasonably, or act as one’s ideally reasonable version would. One’s commitment is not necessarily to acting in accordance with the principles that some unrecognizably reasonably ideal person would adopt. However, if my ideal version must make decisions in ignorance of our
particular features, then *when deciding on principles*, he is no different than the unrecognizable reasonably ideal person. The fact, then, that the principles are chosen by legislators who are ignorant of the features that define them seems to undermine the justification of moral principles to the agents who are bound by them.

It also seems to undermine the motivation a real agent would have to act in accordance with such principles. Part of the reason each agent has to act in accordance with the principles rests on his commitment to being a reasonable version of himself, that is, his commitment to acting according to principles that a reasonable version of himself would adopt. But if it is not clear that a reasonable version of himself *would* adopt them, then to that extent he loses motivation to act accordingly. Furthermore, another part of the reason he has to act according to these principles is that acting in accordance with these principles is a way of fulfilling his commitment to others, insofar as these principles are chosen by ideal versions of others. Again, however, if they are chosen by legislators that bear no particular resemblance to those he has a commitment to, then his commitment to others does not seem to require acting according to these principles. If acting accordingly is only acting according to principles that abstract agents would choose, then (it seems) he does not have a commitment to acting according to these principles *just* in virtue of having a commitment to others. So, in this respect, one might find Hill’s theory unsatisfactory (at least as spelled out in Hill, 1992: 59).

However, it is worth questioning whether this feature of his account is necessary. That is, I doubt whether there is a difference between, on the one hand, the principles decided in a kingdom of ends where the legislators are not aware of their particular features and, on the other hand, the principles decided in a kingdom of ends where the legislators are aware of their particular features but are also committed to the task of coming up with principles that will
govern their behavior, a task that requires discipline of not giving inappropriate weight to one’s own personal ends. And, while there may not be a difference between the principles decided on, there is a difference in the degree to which real agents are bound to act in accordance with those principles.

Now, it may seem that allowing the legislators to have particular knowledge of themselves would allow them to be unfair, or to aim to legislate principles that would favor the individual whom they ‘represent.’ However, recall that the legislators are stipulated to have a concern for others. This feature of the legislators is captured by the fact that they each regard each member as an end in himself (or at least, they are only willing to adopt principles that treat each member as an end in himself). This commitment will prevent each of them from inappropriately giving weight to the particular features of the individual of whom they are ideal versions. If the ideal legislators are willing to legislate with an appreciation for everyone’s position, then it is not necessary for them to abstract from all personal differences.

Note that this resulting account is a dispositional account that is not subject to the difficulties we saw in Firth’s account. The emphasis Firth placed on reactions (and not judgments) was problematic. In this account, judgment plays a crucial role. The ideal legislators judge that certain principles are the best, or at least as good as any other, principles for governing behavior. They judge that a set of principles is the most reasonable set of principles, and then they adopt that set of principles.

It is true that on this account there is not a direct relation (however there is clearly an indirect relation) between the features of an action and its rightness or wrongness (as I said ordinarily we expect). However, it is a virtue of this account that there is not a direct relation. Why? Because this account sacrifices the direct relation to gain the element of hypothetical ideal agreement of
the community, which is a much larger gain. Thus, we manage to retain the advantages of a dispositional account, without the drawbacks of the dispositional account Firth proposed.
Bibliography


