

【논문】

Firth's Ideal Observer Theory and Its Problems*

Chung, Hun

【Subject Class】 Ethics, Metaethics

【Keywords】 Ideal Observer Theory, Firth, Brandt, Dispositional Theory of Value, Ethical Naturalism

【Abstract】 There are three main attractions of an ideal observer theory in ethics: (a) it is less ontologically committed by insisting that moral properties are nothing more than the subjective psychological reactions of an idealized human subject, (b) it, nonetheless, preserves our common sense intuition that the truth of ethical statements are *prima facie* absolute, and (c) it naturalizes ethics by making it an empirical theory that can be pursued in accordance with other scientific disciplines. The purpose of this paper is to introduce and critically analyze an early prototype of such ideal observer theory that was presented by Roderick Firth in 1952. In this paper, I will show how Firth's ideal observer theory, despite his laborious effort to succeed, fails to meet (b) and (c).

I . INTRODUCTION: FIRTH'S IDEAL OBSERVER

There was a time when many variants of “the ideal observer theory”

* This is a shortened version of a much longer paper. If you wish to read the original version, please send me an email at: hunchung1980@gmail.com

were quite popular among ethical theorists in the analytic tradition.¹⁾ However, it seems that ideal observer theories in general have been relatively neglected in the Korean philosophical community. One purpose of this paper is to introduce ideal observer theories to Korean philosophers. Among the many variants of ideal observer theories, this paper will focus on one of the early prototypes of the theory that was presented by Roderick Firth in 1952. I hope to introduce other variants of the ideal observer theory in different papers.

In his article “Ethical Absolutism and the Ideal Observer”,²⁾ Firth presents a very interesting analysis of ethical statements. According to Firth, the meaning of ethical statements of the form “X is right (or good)” can be adequately expressed by statements of the form “Any ideal observer would react to X in such and such a way under such and such conditions.”(p.329) Here, we can see that Firth is interpreting moral properties or moral facts to be equivalent to certain types of psychological reactions of an ideal observer. Then, what kind of being is an “ideal observer”?

According to Firth, an ideal observer is basically a human being who possesses certain kinds of characteristics to *an extreme degree*. The major characteristics that Firth attributes to an ideal observer are the following: First of all, an ideal observer, according to Firth, is *omniscient* with respect to non-ethical facts.³⁾ That is, an ideal observer possesses a

1) See (Railton 1986a; Railton 1986b; Smith et al. 1989; Firth, 1952; Firth, 1955; Brandt, 1955; Brandt, 1998; Harsanyi 1977). Among these people, some have used the apparatus of ideal observer theory to analyze individual goodness or value (e.g. Railton 1986a; Railton 1986b; Smith et al.) Others have used the approach to analyze moral goodness and moral rightness(e.g. Firth 1952; Firth 1955; Brandt 1955; Brandt 1998). And, some used the approach to render interpersonal comparison of well-being possible to justify utilitarianism(e.g. Harsanyi 1977; for a good introduction to Harsanyi’s utilitarianism, see Weymark 1991).

2) Firth, Roderick, “Absolutism and the Ideal Observer”, *Philosophy and Phenomenological Research*, Vol. 12, No. 3 (Mar., 1952).

complete set of full factual knowledge. Second, an ideal observer is capable of vividly imagining all of the relevant facts as if he or she were directly perceiving them. In this sense, an ideal observer is *omnipercipient*.⁴⁾ Third, an ideal observer is *completely disinterested* in the sense of lacking all particular interests.⁵⁾ Fourth, an ideal observer is *completely dispassionate* in the sense of lacking all particular emotions.⁶⁾ Fifth, an ideal observer is *completely consistent* in his or her ethical reactions in the sense that he or she will always show the same ethical reactions towards the same situation.⁷⁾ And finally, an ideal observer is *normal* in every other respect that has not been specifically characterized.⁸⁾

So, according to the ideal observer theory that Firth proposes, when somebody claims that “X is right” what that person is really claiming is that “an idealized human being, who has perfect knowledge of all non-ethical facts, and who is able to imagine all of those facts vividly as if he or she were directly perceiving them, and who is completely disinterested and completely dispassionate, and who is in every other respect normal, would *psychologically react to X in a positive way*.” Here, we can see that Firth’s ideal observer theory is basically an *analysis* of ethical statements; it analyzes ethical statements and reveals what they really mean.

3) Ibid. p.333.

4) Ibid., p.335.

5) Ibid., p.335.

6) Ibid., p.340.

7) Note that although Firth contends that the ideal observer is consistent in his or her moral reactions, Firth thinks that consistency has a rather different status than all of the other characteristics; that is, according to Firth, consistency is not really an independent characteristic of the ideal observer, but rather a *consequence* of all of the other characteristics (such as omniscience, omnipercipience, disinterestedness, dispassionateness, etc.) functioning properly. In this sense, consistency can be seen as redundant in defining the ideal observer. See Ibid., pp.341-344.

8) Ibid., p.344.

According to Firth, analyzing ethical statements in this way has a number of characteristics. First of all, according to Firth, the proposed analysis deems the truth of ethical statements to be *absolute* rather than relative.⁹⁾ This is because, according to Firth, any two ideal observers that met all of the defining characteristics would necessarily show exactly the same psychological reactions towards any given situation. Therefore, if a certain ethical statement turns out to be true according to Firth's ideal observer theory, then the truth of the ethical statement would be both universal and absolute. Second, the proposed analysis basically construes ethical statements to be *dispositional* statements.¹⁰⁾ That is, according to Firth's ideal observer theory, ethical statements mainly state that an idealized human being would be *disposed* to psychologically react to something in a certain type of way. Third, the proposed analysis construes ethical statements to be *objective*, where objectivity and subjectivity are defined in the *traditional ontological sense*.¹¹⁾ When ethical statements are deemed to be subjective in this traditional ontological sense, it means that whether or not there actually exists experiencing human subjects have direct bearing on the truth or falsity of ethical statements.¹²⁾ By contrast, if ethical statements are deemed to be

9) Ibid., p.318.

10) Ibid., p.320.

11) Ibid., p.322.

12) Firth explains that in a subjectivist analysis of ethical statements, all ethical statements would turn out to be false by definition if there happened to be no experiencing subjects. (See Ibid., p.322) However, whether all ethical statements would turn out to be false in a subjectivist analysis of ethical statements if there happens to be no experiencing subjects is a controversial issue. For ethical statements might simply *presuppose* the existence of experiencing human subjects. In this case, ethical statements would simply lack truth-value (rather than being false) in a subjectivist analysis. Furthermore, all ethical statements might turn out to be *vacuously true* in a subjectivist analysis if there were no experiencing human subjects. So, it is not obvious that all ethical statements would necessarily turn out false by definition if there were no experiencing human subjects even if we follow a

objective in this traditional ontological sense, it means that whether or not there actually exists experiencing human subjects is logically irrelevant to the truth or falsity of ethical statements. Firth's ideal observer theory construes ethical statements to be objective precisely in this way. Fourth, the proposed analysis construes ethical statements to be *relational statements*.¹³⁾ That is, according to Firth's ideal observer theory, what ethical statements are basically asserting is that there is a *lawful relationship* between the psychological reactions of an ideal observer and the situations to which an ethical term may correctly be applied. In this way, ethical statements, according to Firth's ideal observer theory, has the same structure of statements about secondary qualities.¹⁴⁾ Fifth, the proposed analysis construes ethical statements to be *empirical statements*.¹⁵⁾ As we have seen, Firth's ideal observer theory construes ethical statements to be expressing how an ideal observer would *psychologically* react to a certain situation. Here, the psychological reactions of an ideal observer are those that can, in principle, be investigated by scientific and observational techniques that widely used in empirical sciences.

subjectivist analysis of ethical statements. Therefore, I think that a better way to characterize a subjectivist analysis and an objectivist analysis of ethical statements would be (not whether all ethical statements would turn out to be necessarily false if there happens to be no experiencing human subject, but) whether the actual existence of experiencing human subjects have any logical bearing to the truth or falsity of ethical statements. I thank Nick Sturgeon for providing me with this insight.

13) *Ibid.*, p.324.

14) Just as a statement about secondary qualities, say "daffodil is yellow", is usually analyzed as expressing the same meaning as the sentence, "the daffodil will appear in a certain way to a certain kind of observer under certain conditions", Firth's ideal observer theory construes ethical statements such as "X is right" as expressing the same meaning as the statement "an ideal observer would be disposed to psychologically react to X in a positive way." (See *Ibid.*, p.324)

15) *Ibid.*, p.325.

It is quite easy to see why such an ideal observer theory could be appealing to many people.

First of all, it is less ontologically committed than certain views in ethics that posit moral properties, which are usually seen as properties that are intrinsically different from natural properties, existing independently as furniture of the world.¹⁶⁾ According to the ideal observer theory, moral properties are nothing more than certain types of psychological reactions of idealized human beings; or to put it in another way, the psychological reactions of idealized human being *make* moral properties. In this way, the ideal observer theory successfully avoids the complicated metaphysical issues that arise from the existence of moral properties.

Second, even though the ideal observer theory construes moral properties to be nothing more than certain types of psychological reactions, it, nonetheless, preserves the *prima facie* objectivity of the truth of ethical statements. This is because the ideal observer theory doesn't deem the psychological reactions of just any kind of being to be relevant in determining the truth of ethical statements; the psychological reactions that are relevant are only those of *idealized human beings* and the fact that these human beings are *idealized* implies that their psychological reactions to a given situation are somewhat correct and authoritative. In this sense, the ideal observer theory seems to be combining the major attractions of both cognitive and non-cognitive views in ethics.¹⁷⁾

16) Note that this advantage of the ideal observer theory doesn't apply to ethical naturalism which regards moral properties to be natural properties.

17) That is, the ideal observer theory acknowledges that there are moral facts and moral properties that make ethical statements objectively true or false. So, it accommodates one of the main attractions of cognitive moral theories. However, the ideal observer theory construes the moral facts and moral properties that determine the truth-value of ethical statements to be certain types of subjective psychological

One last attraction of the ideal observer theory comes from its *science-friendliness*. As I have briefly noted, the ideal observer theory construes the meaning of ethical statements to be expressing the psychological reactions that an idealized human being would feel towards a given situation. Here, the psychological reactions of an ideal observer are the properties that can be, in principle, studied by the various sciences, such as psychology, biology, neurophysiology, and so on. Of course, the ideal observer theory itself is, as we have seen, intended to be an *analysis* of ethical statements. However, once the analysis is accepted, the ideal observer theory makes it possible to investigate first-order moral judgments in an empirical way. This makes the study of normative ethics *empirical* and puts its theoretical status on par with other scientific disciplines. This is a major attraction of the theory to those who are scientific-minded.

So, the ideal observer theory, when developed successfully, can provide a very powerful tool for doing normative ethics. And Firth's ideal observer theory seems to come very close to such success. Unfortunately, Firth's ideal observer theory has a number of innate problems that don't seem to be easily overcome within his own framework. The main purpose of this paper is to expose these problems. Most of these problems stem from the many defining characteristics of the ideal observer as well as the many characteristics of the proposed analysis conflicting with each other. Some other problems stem from Firth's rather imperfect characterization of such notions as "impartiality" and "relativist". Now, let's look at those problems in detail.

responses of an ideal observer which are mainly positive or negative attitudes that are essentially non-cognitive in their nature. In this way, it also accommodates one of the main attractions of non-cognitive moral theories as well.

II. THE PROBLEMS

As we have seen, the main attractions of the ideal observer theory come from the following three features:

- (1) It is less ontologically committed in the sense that, by identifying moral properties with psychological reactions of the ideal observer, it doesn't have to posit independent (non-natural) moral properties that exist as furniture of the world in order to explain the truth of ethical statements.
- (2) Even though it construes moral properties as being psychological reactions of the ideal observer, it still retains our commonsense intuition that ethical statements are somewhat *objectively* true or false.
- (3) It fits well with our scientific picture of the world by making normative ethics essentially an empirical study.

It seems that any version of the theory should try to retain these three characteristics as much as possible. The problem is that this is not an easy task to do.

1. Absolutism vs. Relativism—(1): Is Firth's analysis really absolutist?

As we have seen, Firth thinks that his ideal observer theory is an absolutist analysis of ethical statements; that is, according to Firth's ideal observer theory, the truth or falsity of ethical statements are absolute and are not relative to the speaker or the situation. The main reason why Firth's ideal observer theory becomes an absolutist analysis is, according to Firth, because it does not use any "egocentric expressions" when it analyzes the meaning of ethical statements. By "egocentric expressions",

Firth is referring to what are commonly known as indexical terms and demonstratives.¹⁸⁾ The major characteristic of egocentric expressions is that they tend to change their meaning systematically depending on whom the speaker is as well as in what situation the statement containing such expressions are uttered. According to Firth, any relativistic analysis of ethical statements is relativistic simply because it cannot avoid using such egocentric expressions in its analysis of ethical statements.¹⁹⁾

So, a typical relativist analysis of ethical statements would construe ethical statements such as “X is right” as expressing the same meaning as sentences like “I would feel approval towards X” or “Most people *here* and *now* would feel approval towards X” and so on. Here, we can see that each of these analyses include egocentric expressions (i.e. “I”, “here”, “now”). If any of these relative analyses were correct, then it would be possible for one person to say that a certain act is right, and for another person to say that the very same act is not right *without logically contradicting* each other. This is a familiar characteristic of all relativistic analyses. However, according to Firth, this characteristic is not by itself what defines a relativistic analysis; rather this characteristic is a *necessary consequence* of the relativistic analyses containing egocentric expressions. So, what really makes a relativistic analysis of ethical statements relativistic is the very fact that it includes egocentric expressions.

According to Firth, whether or not an analysis contains egocentric

18) Examples of egocentric expressions include personal pronouns (“I”, “you”, etc.), the corresponding possessive adjectives (“my”, “your”, etc.), words that refer to spatial temporal location (“this”, “that”, “here”, “there”, “now”, “then”, “past”, “present”, “future”, etc.), reflexive expressions such as “the person who is speaking”, and the various linguistic devices which are used to indicate the tense of verbs.

19) Firth claims that “an analysis of ethical statements is relativistic if it includes an egocentric expression, and if it is incompatible with any alternative analysis which does not include an egocentric expression” (Ibid. p.319)

expressions is the sole indicator of whether or not an analysis is relativist.²⁰⁾ And this is precisely why Firth regards his ideal observer theory as an absolutist analysis; Firth's ideal observer theory does not include any egocentric expressions in its analysis of ethical statements.

We should note that Firth's characterization of a relativist analysis that focuses on the existence of egocentric expressions is a very convenient way to distinguish between a relativistic analysis and an absolutist analysis, since, according to this characterization, we can know whether a proposed analysis is relativistic or absolutist simply by inspecting its constituent expressions.²¹⁾ However convenient it may be, there are certain problems that are looming here.

The first problem is that Firth's characterization of a relativistic analysis might be *too restrictive*. This is because there is still a sense in which Firth's ideal observer theory is relativistic even though it does not contain an egocentric expression on its surface. Remember that Firth is trying to identify the meaning of ethical statements with the psychological reactions of an ideal observer. Here, 'an ideal observer' refers to idealized *human beings*; it does *not*, for example, refer to idealized dogs or idealized Martians. Firth himself makes it very clear that 'an ideal observer is, after all, a *person*.'²²⁾

Now, if there really were Martians, then it seems perfectly possible for

20) Firth claims "We may now define an absolutist analysis of ethical statements as one which is not relativist. The kind of analysis which I propose to discuss in this paper, therefore, is one which does not include an egocentric expression." (Firth, *ibid.* p.319)

21) Of course, this can be properly done only when the analysis is *fully* developed. This is because an analysis may only *implicitly* include egocentric expressions in which the egocentric expressions appear only after the analysis is developed in its full-fledged form. We will later see that Firth's own ideal observer theory could be seen as implicitly including egocentric expressions in precisely this sense.

22) *Ibid.* p.344. The criticism that I will be making afterwards will still apply even if we interpret the set of *persons* to include highly cognitive *primates*.

them to have a completely different psychological disposition than ordinary human beings-maybe they tend to be meaner and crueler than ordinary human being. If this were the case, then it is highly likely that an idealized Martian would show very different psychological reactions to a certain act than what an idealized human being would show to the same act. So, if Martians used their own version of the ideal observer theory and construed ethical statements such as "X is right" as saying that an idealized Martian would psychologically react to X in a positive way, then it seems that many ethical statements would turn out to be systematically ambiguous; "X is right" might be true for Martians while being false for human beings. However, Firth wouldn't want to say that ethical statements are generally ambiguous in this way; Firth would want to say that the different psychological reactions of the idealized Martian have no bearing whatsoever in determining the meaning of ethical statements. This means that Firth is restricting his analysis of ethical statements to the psychological reactions of a particular species; *Homo Sapiens*.

In this way, Firth's ideal observer theory resembles what philosophers call "appraiser-relativism"-the view that what counts as morally right or wrong is determined by the standards that are invoked by the appraiser's group. Or Firth's ideal observer theory can be regarded as a form of *speciesism*. By identifying the meaning of ethical statements such as "X is right" with the statement, "An ideal observer will be disposed to show positive psychological reactions towards X", where 'an ideal observer' specifically refers to an idealized human being, Firth's ideal observer theory is, in effect, interpreting "X is right" to mean "X is right *for human beings*". And if we regard speciesism as a form of appraiser-relativism, then it turns out that Firth's ideal observer theory can also be interpreted as a relativist analysis in precisely this sense.

However, according to Firth's characterization, we are unable to characterize speciesism as well as many other forms of appraiser-relativism as a form of relativism since many of these analyses do not include any egocentric expressions. In short, Firth's characterization of a relativist analysis is too restrictive in the sense that it leaves many legitimate forms of relativism out of the scope of the category of relativist analyses.

Second, more problematically, there is a sense in which Firth's ideal observer theory is relativistic even by Firth's own standards. This is related to Firth's claim that an ideal observer is *normal* in all other aspects that are not specifically defined.²³⁾ Here, again, what is 'normal' must mean what is normal for human beings, not what is normal for, say, Martians. However, even if we restrict our domain to human beings, it is quite evident that what counts as normal for human being differs from time and place.²⁴⁾ So, in order to determine how an ideal observer would react to a specific situation, we must fix a standard of normality reference to which we are regarding the ideal observer to be normal in all other aspects that are not specifically defined. There cannot be such thing as 'just normal' or 'universally normal'; that is, any viable notion of normality must be fixed by reference to a specific species living in a specific time and location.

This means that what is 'normal' might not be able to be properly defined without using what Firth calls *egocentric terms*; if an ideal

23) Ibid. p.344

24) To the very extent that people's psychological dispositions are determined by society and culture, what would count as a normal psychological reaction would at least be partly determined by the influences of the particular social structure and the culture in which people have been brought up by. To the very extent that people's psychological dispositions are determined by their biological make-up, what would count as a normal psychological reaction would be partly determined by the particular stage of human evolution people are in.

observer is normal, then he or she must be normal in a more specified sense; such as being “normal for *us* living *here* and *now*” or being “normal for human beings *beginning 5000 years ago*” or being “normal for human beings living in *modern* democratic societies” and so on in order for the analysis to generate any determinate results for the meaning of particular ethical statements. In this way, we can see that although Firth’s ideal observer theory might not include egocentric expressions on its surface, it may, nonetheless, implicitly include such egocentric expressions which will appear when the analysis is developed into its full-fledged form. If this is the case, then Firth’s ideal observer theory can be accused of being relativistic in precisely the same sense that Firth defines “a relativistic analysis” of ethical statements.

So, the problem for Firth on his characterization of a relativist analysis can be summarized as the following dilemma: either he would need to use ego-centric terms himself in order to fix the notion of normality which would make his ideal observer theory a relativist analysis by his own standards, or he might be able to avoid using any ego-centric terms by fixing the notion of normality to any specific time or location which would then make his analysis totally arbitrary.

2. Absolutism vs. Relativism—(2): Can Ideal Observers Differ?

As we have seen, Firth’s ideal observer theory construes ethical statements as expressing the same meaning as the statements about the psychological reactions of an ideal observer. Therefore, in order for the truth or falsity of ethical statements to be absolute, it is necessary that all ideal observers show *exactly the same* psychological reaction towards the same specific act. And this is exactly what Firth is trying so hard to

accomplish. That is, a major portion of Firth's article is devoted to finding what Firth thinks are the necessary defining characteristics that would *guarantee any ideal observer to show exactly the same psychological reactions to any particular act*. And Firth seems quite confident that the several characteristics that he has proposed in his article would actually guarantee any two ideal observers to show exactly the same psychological reactions to any particular situation.

The claim that any ideal observer would show exactly the same psychological reaction to any particular act is the point where Richard Brandt raises one of his major criticisms against Firth's ideal observer theory.²⁵⁾ According to Brandt,

The facts of ethnology and psychological theory suggests that there could (causally) be two persons, both "ideal observers" in Firth's sense, who would have different or even opposed reactions (approval, experience of apparent requiredness) with respect to the same act, say on account of past conditioning, as different system of desires, etc.²⁶⁾

If this were to happen, then it would significantly damage Firth's analysis, since, according to Firth's analysis, no ethical statements would turn out to be true if there happened to be one single ideal observer who showed different psychological reactions toward a specific act from the rest of the ideal observers. Brandt concludes that it is, therefore, better to take a *relativistic approach* in developing an ideal observer theory; ethical statements such as "X is right" is true if and only if the speaker regarded the following statement to be true if he or she uttered it: "If I were an ideal observer, then I would show positive psychological reactions towards X regardless of whether or not other ideal observers

25) Brandt, Richard, "The Definition of an "Ideal Observer" Theory in Ethics", *Philosophy and Phenomenological Research*, Vol. 15, No. 3 (Mar., 1955)

26) Ibid. p.408

would show exactly the same psychological reactions”.

Note that there are two separate things that are at issue here: one is whether the ideal observer theory properly captures the meaning of ethical statements, and the other is (given that the ideal observer theory properly captures the meaning of ethical statements) whether any two ideal observers would necessarily agree to any particular issue. Firth and Brandt both seems to agree on the first issue; that is, Firth and Brandt both think that the ideal observer theory properly captures the meaning of ethical statements. And this is part of the reason why both Firth and Brandt are so eagerly engaged in the issue on whether or not all ideal observers would necessarily agree to any given issue; by accepting that the ideal observer theory properly captures the meaning of ethical statements, Firth and Brandt are both committed to the view that the truth of certain statements are, at the least, not absolute whenever any two ideal observers disagree about the matter in question. Whenever any two ideal observers disagree, this means that there is *no moral fact* of the matter that would settle the issue on whether that thing is right or wrong in absolute terms. Brandt tries to resolve this issue by recommending a relativistic approach, while Firth tries to resolve this issue by denying that any two ideal observers can disagree about anything.

Firth, in his reply, dismisses the seriousness of Brandt's objection by saying that “any argument to prove, on the basis of psychological evidence, that it is causally possible for two ideal observers to have conflicting moral experiences, would necessarily be a very frail and debatable one.”²⁷⁾ According to Firth, in order for somebody to *prove* that two *truly* ideal observers really showed different psychological reactions towards the same act, it would have to be shown that the

27) Firth, Roderick, “Reply to Professor Brandt”, *Philosophy and Phenomenological Research*, Vol. 15, No. 3 (Mar., 1955), p.415.

different psychological reactions were not due to any one of the alleged ideal observers not meeting all of the necessary defining characteristics of an ideal observer to their full extent.

However, as one might expect, this is not an easy task to do. First of all, in order to prove that the two alleged ideal observers were both really ideal observers, it would have to be shown, in accordance with the omniscient requirement, that the two alleged ideal observers were in exactly the same cognitive states, having exactly the same beliefs in every psychological, biological, sociological, and theological fact. This is because any differences in the two alleged ideal observers' respective belief system might have caused them to show different psychological reactions towards the same act, which would, then, mean that at least one of them is not an ideal observer in Firth's sense.

Second, even if we were able to confirm (which is highly unlikely in real life) that the two alleged ideal observers were exactly identical in their respective cognitive states, we would, then, have to further show, in accordance with the disinterestedness and dispassionateness requirement, that their different psychological reactions were not due to any one of them having selfish interests or self-referential emotions.

What Firth is trying to argue is that it is virtually impossible to prove from any viable methodological technique that the two alleged ideal observers who showed different psychological reactions were *really* two ideal observers in Firth's sense. And unless this can be shown, Firth seems to think that it is more reasonable to suppose that the different psychological reactions were due to any one of them being less than ideal. In reply to Brandt, Firth claims that if I really believe that I am an ideal observer, then I should refuse to admit anybody else who showed different psychological reactions as also being an ideal observer as well.²⁸⁾

28) Ibid. p.416

However, I do not believe that the problem that Brandt raises can be dismissed that quickly. Suppose that there are two subjects: A and B. A is an ideal observer in Firth's sense; that is, A meets all of the necessary defining characteristics of an ideal observer that Firth requires. B is not yet an ideal observer. Suppose that we try to make B into an ideal observer by providing B with the full set of non-ethical factual knowledge, correcting every false belief in B's cognitive system, and conditioning B to become perfectly disinterested and dispassionate. Suppose that all of these processes were successful. Would B, then, show exactly the same psychological reactions as A towards every specific act or situation?

I do not think so. This is because even if B, after the process, became completely disinterested and dispassionate, and even if B's cognitive state came to perfectly match A's cognitive state belief by belief, the *causal histories* on how A and B acquired their respective factual beliefs were completely different;²⁹⁾ B acquired every other factual belief that he

29) David Velleman, (in his article "Brandt's Definition of "Good"(1988)") has raised a similar objection to Brandt's "Full-information account of the "Good"". According to Brandt (1979), what is good for a certain agent is what that agent would desire after the agent obtained full factual information through what he calls "cognitive psychotherapy". One of the many objections that Velleman raises against Brandt is that the same agent might desire things differently, even after obtaining full factual information through cognitive psychotherapy, depending on *what specific medium* that the full factual information was introduced. (See Velleman (1988); pp.365-368) Peter Railton, who also proposes a version of a full-information account of value in "Facts and Values (1986)", also considers the problem of the effects that the specific *mode* and *order* of factual representation would have on the agent who achieved full factual information. In the example that I have provided, we can see that the order as well as the mode of how A and B attained full factual information was completely different. Railton tries to appease this worry by arguing that the effect of the order of presentation would become very dilute by second-order mechanisms of neutralization (i.e. the neutralization process that takes place when one becomes aware of how different orders of presentation could have effected one's conative systems) as the total volume of information grows (Railton, 2003, p.58) and by

had not acquired by his own past experience by experimental conditioning, while this is not the case for A by assumption. Accordingly, A and B would each associate completely different experiences to each of the respective non-ethical factual beliefs they have, and the impact of these different associated experiences can very well cause them to show quite different psychological reactions towards the same act.

Of course, Firth has a device to counter this kind of result; and the device is his requirement of “omnipercipience”.³⁰ According to Firth, being omniscient respect to non-ethical facts (in the sense of being able to make true factual judgments) is insufficient to make that person an ideal observer; “The ideal observer must be able, on the contrary, simultaneously to visualize all actual facts, and the consequences of all possible acts in any given situation, just as vividly as he would if he were actually perceiving them all.”(Firth, 1952, p.335)

It might be the case that any two subjects who met the omnipercipience requirement would necessarily associate exactly the same perceptual experiences to any set of non-ethical factual beliefs. So, the impacts of the different causal histories associated to the non-ethical factual beliefs of A and B might as well, in practice, be neutralized by making B meet the omnipercipience requirement. However, it seems that the omnipercipience requirement is still insufficient to nullify all of *the lingering effects* of B’s past experiences past experiences and past false beliefs.

arguing that many different modes of presentation would essentially have the same effect as our conative systems is not infinitely fine-grained (Railton, 2003, p.60).

30) Or Firth might take a similar stance to Railton and claim that the different effects that are attached to the different causal histories of the subject’s factual beliefs would eventually be neutralized as the total volume of information grows. See above foot note.

To this, Firth might respond by *adding a separate defining characteristic* of an ideal observer; that is, not only should an ideal observer be omniscient, omnipercipient, disinterested, dispassionate, consistent, and in other respects normal, but an ideal observer should also be *completely free from the lingering as well as the subconscious effects of his or her past experiences as well as his or her past false beliefs.*

It might seem that this would actually guarantee any two ideal observers to show exactly the same psychological reactions towards any specific act. However, this is not yet entirely obvious. And this again is due to the inherent vagueness that is associated with the requirement that an ideal observer should be *normal* in all other respects that are not specifically defined. As we have seen, what counts as “normal” even for a particular species, (say, human beings), differs from time and place. Human beings might evolve into beings that have completely different psychological and dispositional natures in the future. The social structure of future societies might be so different from that of modern liberal democratic states that what counts as a normal psychological reaction of human beings who were brought up in future societies might be drastically different from what counts as a normal psychological reaction of human beings who were brought up in a modern liberal democratic state.

This means that it is not entirely obvious that two ideal observers, who met all of Firth's requirements as well as the additional requirement of not being influenced by the lingering as well as the subconscious effects of his or her past experiences and false beliefs, would show exactly the same psychological reactions towards the same act, if one of them happen to be normal in the sense of being normal for human beings born after A.D 5000 while the other happen to be normal in the sense of being normal for human beings living today.³¹⁾

We cannot simply say that, since the two ideal observers each meet different standards of normality, at least one of them is *not really* an ideal observer, since, then, we would have to justify why one standard of normality is more legitimate than the other. Unless we are able to provide this justification, dismissing any one of these ideal observers as non-ideal would be totally arbitrary. If Firth tries to avoid this problem by stipulating the meaning of “normal” as “normal for human being living *in the present*”, then Firth’s analysis will again contain what he calls an egocentric term, and would, thereby, turn into a relativistic analysis by his own standards. So, we are back to the dilemma that has been mentioned for Firth’s characterization of a relativist analysis.

3. Objectivist Analysis vs. Subjectivist Analysis

As one might have noticed, Firth’s ideal observer theory, even without the additional defining characteristics, is already quite impractical. It identifies the meaning of ethical statements with the meaning of

31) This is again related to one of Velleman’s objections towards “the full-information account of the good” that the type of dispositions that one happens to currently possess would significantly influence how certain factual information would affect one’s desires; according to Velleman, “Images of alien cultures may rouse us to wonderment if we’re cosmopolitan, whereas if we’re provincial, the same images may arouse contempt; the thought of a storm at sea may thrill us if we’re brave but sicken us if we’re timid; and our hearts may grow either hard or tender at the sight of other people’s pain”(“Brandt’s Definition of “Good””, p.360) Of course, these specific objections of Velleman would not apply to Firth’s ideal observer theory, since not only is Firth requiring full factual information, but also full perceptual vividness, complete disinterestedness and complete dispassionateness which would, in many cases, nullify the effects of people’s particular dispositional traits. However, Firth’s many characterizations of the ideal observer would still be insufficient to nullify the different psychological reactions of two ideal observers who each had different dispositions of normality based on different conceptions of ‘what is normal’.

statements about psychological reactions of an ideal observer; but in order for somebody to be an ideal observer he or she must be omniscient and omnipercipient with respect to non-ethical facts, perfectly disinterested and dispassionate, consistent in his or her ethical reactions, normal in all other respects, and now, he or she might also be required to be completely free from the lingering as well as the subconscious effects of his or her past experiences as well as his or her past false beliefs. Even without the last additional requirement, it seems virtually impossible for any living human being to even roughly satisfy all of the defining characteristics of an ideal observer that Firth requires. And if it is just plainly impossible for there to be any actual ideal observers in Firth's sense, then, one might wonder; shouldn't this fact influence the meaning of ethical statements in any way?

Firth's response to this question is quite simple. The answer is: *not at all*. Obviously, Firth doesn't want all ethical statements to turn out *truth-valueless* simply because there are no ideal observers that actually exist. And this is his main motivation for defining the distinction between an "objectivist analysis" and a "subjectivist analysis" in a very idiosyncratic way.

When Firth characterizes his ideal observer theory as an objectivist analysis of ethical statements, he claims that he is using the terms "subjectivist" and "objectivist" in a *traditional ontological sense*—"in the sense in which Berkeley's analysis of all physical statements is subjectivist, and Descartes' analysis of some physical statements is objectivist." (Firth 1952, p.322)

...a proposed analysis of ethical statements is subjectivist if it construes ethical statements in such a way that they would all be false by definition if there existed no experiencing subjects (past, present, or future). An analysis may be called "objectivist," on the other hand, if it is not

subjectivist. (Firth 1952, p.322)

Therefore, when Firth claims that his ideal observer theory is objectivist, what he is mainly saying is that, although the ideal observer theory construes the meaning of ethical statements as expressing the same meaning as the statements about psychological reactions of an ideal observer, whether or not an ideal observer actually exists is logically irrelevant to the truth or falsity of ethical statements.

Firth emphasizes that the adjective “ideal” is here used in approximately the same sense in which we speak of a *perfect* vacuum or a *frictionless* machine as ideal things(Firth 1952, p.321). The term implies that the thing in question is *conceivable* and possesses certain characteristics to *an extreme degree*; it doesn’t imply that the ideal thing in question *actually exists* in real life. Just as the fact that there are no perfect vacuums or perfectly frictionless machines in real life does not influence the truth or falsity of scientific theories that rely on these idealistic concepts, the fact that there are no ideal observers in real life does not, in any way, influence the truth or falsity of ethical statements. In this way, ethical statements can be interpreted as contrary-to-fact conditionals in the subjunctive mood; the ethical statement “X is right” means that “if there *were* an ideal observer, the ideal observer *would* be disposed to react to X in a positive way.” The latter sentence can be true, even if there were no ideal observers in actual circumstances. And it is precisely in this sense that Firth regards his ideal observer theory as *objectivist* rather than *subjectivist*.

As we have seen, the main motivation for characterizing the terms “objectivist” and “subjectivist” in this particular way stems from Firth’s desire to retain the meaning as well as the truth-value of ethical statements even when there happens to be no ideal observers in the actual

world. However, characterizing objectivism and subjectivism in this traditional ontological sense has a number of problems.

First, the characterization is somewhat misleading. As we have seen, one of the main attractions of an ideal observer theory comes from the very fact that it is *subjective*; that is, it interprets moral properties to be nothing more than the subjective psychological reactions of an idealized human being. The fact that the subjective psychological responses are what makes ethical statements true is a vital part of the ideal observer theory, since not only does this free an ideal observer theorist from being ontologically committed to non-natural moral properties that some philosophers deemed to be metaphysically queer,³²⁾ but it also *naturalizes* ethics and turns it into an *empirical science* that can be conducted in accordance with psychological, biological, and neural physiological research.

Moreover, as we have already seen, another part of the attraction of the ideal observer theory comes from the fact that the ideal observer theory retains the objectivity of ethical statements. That is, even though an ideal observer theory interprets moral properties to be nothing more than the subjective psychological responses of a certain being, it, nonetheless, retains the *prima facie* objectivity of the truth of ethical statements by confining the relevant psychological reactions, which determine the meaning of ethical statements, to those of an *idealized* subject. Here, the fact the psychological reactions that determine the truth of ethical statements come from, not just any sort of subject, but from an *idealized subject* makes those psychological reactions not *merely subjective*, but also *objective* and, therefore, authoritative as well. Firth's characterization of objectivity and subjectivity seems to make these major attractions of an ideal observer theory far less unclear.

32) See Mackie, John, *Ethics-inventing right and wrong*, Penguin Books, 1977.

4. The Empirical Inaccessibility of the Ideal Observer

However, the problem on whether or not the actual existence of an ideal observer influences the truth or falsity of ethical statements seems less important compared to the problem on whether the fact that it is impossible for an ideal observer to actually exist undermines the theory's *empirical accessibility*.

Since there are apparently no ideal observers in real life, there is no way for us to have any direct epistemic access to how an ideal observer would react to any given ethical situation.³³⁾ If this is the case, then, how is Firth's ideal observer theory supposed to make normative ethics an *empirical theory* that is on par with the other sciences?

In this sense, we should note that Firth's notion of an 'ideal observer' is quite different from the concept of 'perfect vacuum', 'frictionless surface', 'perfectly competitive market' or any other idealized concepts that are frequently employed in the natural and social sciences. Ideal concepts or idealized conditions in the natural and social sciences are not just conceivable entities that have certain characteristics to an extreme degree as Firth describes. These idealized concepts and conditions work as reference points in which *perfect prediction* of what would happen at any given moment is made possible by being aware of its initial conditions.

For example, in a perfect vacuum there would be absolutely no air resistance when a certain object falls. This means that a falling object in perfect vacuum would strictly obey the laws of gravity. Therefore, we would be able to provide a perfect prediction on the exact location as well as the exact velocity of the object at T_2 by being provided with the

33) In this way, Railton agrees that his notion of idealized desires (which is comparable to Firth's ideal observer theory) is unabashedly a theoretical one(Railton, 2003, p.57).

information of the initial location as well as the velocity of the object at $T1$.

Let's look at another example in economics. In a perfectly competitive market, there are absolutely no externalities. Therefore, any good would be produced to the extent that the marginal social utility that it brings matches the marginal social cost of its production. In this way, the social equilibrium point will always match the social optimum point. By knowing the demand function as well as the supply function for a certain good, we can always give a perfect prediction on how much as well as at what price the good in question would be produced and consumed in the market.

What the two examples illustrates is this; ideal concepts or idealized conditions in the natural and social sciences are theoretical models in which *the laws of the respective sciences are strictly and perfectly observed*, which makes *perfect prediction possible* at least theoretically.

This is not the case for the ideal observer theory that Firth presents. Firth's ideal observer theory cannot serve as a scientific model, since, although we know what conditions must be fulfilled in order for someone to be an ideal observer, *we do not know how fulfilling these conditions would influence the ideal observer in any way that would make us reasonably predict how the ideal observer will psychologically react to any given situation.*³⁴⁾ Since we are unable to know or even reasonably

34) Thus, empirical inaccessibility is one of the major problems of such moral theories that rely on such idealized conditions. Velleman raises a similar problem to Brandt's version of "the full-information account of the good" which he calls the *empirical indeterminateness* of such idealized theories. The objection is that if such idealized theory tries to nullify the effects of different modes of factual representation by requiring the subject to be exposed to facts from *all possible angles* (where this includes *all future modes of representation* that has yet been developed), then no scientific or observational method would be able to assess how having full factual information in such way would impact a given subject. (see Velleman, 1988, pp.368-370) Note how this is parallel to my objection that requiring complete

predict how an ideal observer would react to any given situation, we are unable to practically use this theory to evaluate whether one person is closer to an ideal judge than another in respect to a certain situation in real life.

The most devastating thing is that, since we are unable to know or even reasonably predict how an ideal observer would react to any given situation, we are unable to use Firth's ideal observer theory to figure out what the truth-value for any particular ethical statement would actually be. This means that the theory is *practically useless*.

Moreover, it also means that Firth's ideal observer theory practically turns out to be a version of *ethical agnosticism*; it contends that there are ethical statements that are objectively true, but we cannot know which ones they are. This is disappointing since we would want our first order moral theory to not merely claim that the truth of ethical statements are objective and absolute, but also to provide us with a practically feasible guidance to figure out which particular ethical statements are really true in this objective and absolute way. This is especially so for a moral theory that purports to be empirical and deems this as one of its major advantages to other non-empirical moral theories.³⁵⁾

factual knowledge of the distant future would undermine the ideal observer theory's empirical accessibility by rendering reasonable prediction of the ideal observer's psychological reactions virtually impossible.

- 35) We should note that even Railton, who is an advocate of such idealized theories of morality, acknowledges the problem of empirical inaccessibility of such theories. For example, Railton claims; "We are not ourselves fully informed, and thus do not know the answer to the question where an extrapolation of our desires would lead." However, "Given nothing more powerful than commonsense psychological theory, we may feel reasonably confident that there is an answer." (Railton, 2003, p.58) The question that I am raising here is that merely knowing that an answer exists to specific moral questions is not enough for such moral theories to work; in order for the ideal observer theory to have any practical use, (as any viable empirical theory should), we must, at the least, be able to reasonably predict what the theory would regard as answers to specific moral questions. The ideal observer theory loses its

III. SUMMARY

Until now, we have seen some of the major problems that I raised against Firth's ideal observer theory. Here is a summary of the problems that I have raised so far.

There are three main attractions of an ideal observer theory in ethics: (a) it is less ontologically committed by insisting that moral properties are nothing more than the subjective psychological reactions of an idealized human subject, (b) it, nonetheless, preserves our common sense intuition that the truth of ethical statements are *prima facie* absolute, and (c) it naturalizes ethics by making it an empirical theory that can be pursued in accordance with other scientific disciplines.

It seems that any ideal observer theory meets attraction (a) by default. Firth, therefore, concentrates on how his ideal observer theory could meet attraction (b). In order to show that the truth of ethical statements is strictly absolutist, Firth tries to convince us that no two ideal observers would show different psychological reactions to any given situation. In order to guarantee the exact coincidence of any two ideal observers' psychological reactions, Firth presents a very demanding set of defining characteristics that he thinks should be met by any being that can be properly regarded as an 'ideal observer'.

The extremely demanding set of defining characteristics that Firth presents causes a number of problems.

First, even though the set of defining characteristics that Firth presents is extremely (almost ridiculously) demanding, it still cannot guarantee that all ideal observers would show exactly the same psychological reactions towards any given act. This is mainly due to the vagueness that

appeal as an empirically viable first-order moral theory to the extent that it cannot provide any answers to specific moral problems by any empirically viable method.

is associated with the notion of ‘normality’. What this shows is that Firth cannot help his analysis from becoming some version of a relativist analysis.

Second, it is quite obvious that no actual human beings can even approximately meet the full set of defining characteristics that Firth proposes. Firth acknowledges this fact and tries to block the nonexistence of ideal observer in actuality from influencing the truth-value of ethical statements by interpreting the terms “objectivist” and “subjectivist” in the traditional ontological sense and labeling his ideal observer theory as objectivist. However, Firth is quite unaware that his ideal observer theory, by interpreting the truth of ethical statements to be *contingent*, can be regarded as a subjectivist analysis even by his own standards.

Third, the fact that no actual human being can even approximately meet the defining characteristics of an ideal observer theory renders the theory completely empirically inaccessible, which is a big problem for a theory that pays so many lip services to making first-order moral statements empirical.

In short, despite Firth’s constant effort to make his ideal observer theory work the effort is not successful; more generally, his ideal observer theory fails to preserve two of the main attractions of an ideal observer theory: (b) and (c).

References

- Brandt, Richard, "The Definition of an "Ideal Observer" Theory in Ethics", *Philosophy and Phenomenological Research*, Vol. 15, No. 3, 1955.
- _____, *A Theory of the Good and the Right*, Prometheus Books, 1998.
- Firth, Roderick, "Absolutism and the Ideal Observer", *Philosophy and Phenomenological Research*, Vol. 12, No. 3, 1952.
- _____, "Reply to Professor Brandt", *Philosophy and Phenomenological Research*, Vol. 15, No. 3, 1955.
- Harsanyi, John, *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge University Press, 1977.
- Mackie, John, *Ethics-inventing right and wrong*, Penguin Books, 1977.
- Railton, Peter, "Moral Realism", *Philosophical Review*, Vol. XCV, 1986a.
- _____, "Facts and Value", *Philosophical Topics*, Vol. XIV, 1986b.
- _____, *Facts, Values and Norms-Essay Towards Morality of Consequence*, Cambridge University Press, 2003.
- Smith, Michael, and Lewis, David, and Johnston, Mark, "Dispositional Theories of Value", *Proceedings of the Aristotelian Society*, Supplementary Volumes, Vol.63, 1989.
- Velleman, David, "Brandt's Definition of 'Good'", *Philosophical Review*, Vol.97, No.3, 1988.
- Weymark, John, "A reconsideration of the Harsanyi-Sen debate on utilitarianism" in Jon Elster and John E. Roemer (eds.), *Interpersonal Comparisons of Well-Being*, Cambridge University Press, 1991.

퍼쓰(Firth)의 이상적 관망자 이론과 그 문제들

정 훈

(메타)윤리학에서 이상적 관망자 이론이 가지고 있는 장점은 크게 세 가지이다. 첫째, 그것은 도덕적 속성들을 이상적인 인간들이 보여주는 심리적 반응으로 해석함으로써 다른 메타윤리학적 이론에 비해 존재론적 부담이 덜하다. 둘째, 그러면서도 그것은 도덕문장들이 가지는 참-거짓이 일견 절대적이라는 우리의 상식적인 입장을 유지한다. 셋째, 그것은 윤리학을 자연화시킴으로써 윤리학을 다른 과학분야들과 더불어 연구될 수 있는 경험이론으로 만들어준다. 본 논문의 목적은 1952년에 이상적 관망자 이론의 초기 전형을 제시한 퍼쓰의 이론을 소개하고, 그것을 비판적으로 분석하는 데 있다. 본 논문에서 필자는 퍼쓰의 이상적 관망자 이론이, 그것을 성공시키기 위한 퍼쓰의 엄청난 노력에도 불구하고, 방금 소개한 둘째와 셋째 장점들을 충족시키는데 실패했다는 것을 보일 것이다.

주요어: 이상적 관망자 이론, 퍼쓰, 브랜트, 가치에 대한 기질적 이론, 윤리적 자연주의