

Could Ideal Observers Disagree?: A Reply to Taliaferro

Author(s): Thomas L. Carson

Source: *Philosophy and Phenomenological Research*, Vol. 50, No. 1 (Sep., 1989), pp. 115-124

Published by: International Phenomenological Society

Stable URL: <https://www.jstor.org/stable/2108112>

Accessed: 17-04-2019 18:36 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

International Phenomenological Society is collaborating with JSTOR to digitize, preserve and extend access to *Philosophy and Phenomenological Research*

Could Ideal Observers Disagree?: A Reply to Taliaferro*

THOMAS L. CARSON
Loyola University of Chicago

Charles Taliaferro has written a sympathetic and generally accurate account of my version of the ideal observer theory (IOT), as set forth in *The Status of Morality*.¹ However, I remain unpersuaded by most of his main criticisms. The most important area of disagreement between us is over the question whether ideal observers (IOs) could disagree in their attitudes about any moral questions. Taliaferro claims that IOs (as characterized by Firth) would all agree in their attitudes about *all* moral questions.² This issue is very important within the context of the IOT, since, according to the IOT, questions about the objectivity of morals are to be answered by determining the extent to which IOs would agree (in their attitudes or judgments) about moral questions. Firth would accept roughly the following as a criterion for the objective truth of moral judgments:

A favorable (unfavorable) moral judgment about X is objectively true or objectively correct if and only if all possible IOs would feel moral approval (disapproval) for X.³

* This paper was written while I was on a research leave from Loyola University. I am grateful for this support. I would like to thank Paul Moser, Harry Gensler, and Michael Gorr for helpful comments. Firth's classic paper "Ethical Absolutism and the Ideal Observer" first appeared in *Philosophy and Phenomenological Research* 12 (March 1952): 317-45. A very valuable exchange between Firth and Richard Brandt also followed in *Philosophy and Phenomenological Research* 15 (March 1955): 407-23. I would like to express my personal and intellectual indebtedness to Professor Firth and note with sorrow his recent death.

¹ Charles Taliaferro, "Relativising the Ideal Observer Theory," *Philosophy and Phenomenological Research* 49 (September 1988): 123-38.

² Taliaferro does not claim that IOs (as characterized in my version of the IOT) would all agree in their attitudes about all moral questions. Indeed he would take it to be an objection to my characterization of the IO that it implies that IOs could disagree about many moral questions.

³ Firth describes "moral approval" and "moral disapproval" as felt desires and aversions

Firth ascribes the following characteristics to an IO: 1) omniscience with respect to all non-moral facts; 2) omniperception, or the ability to imagine vividly any events or states of affairs, including the experiences of others; 3) disinterestedness, i.e., not having any interests or desires that involve essential reference to particular persons or things (for example, an IO cannot desire his own happiness, he can only desire such things as the happiness of all human beings); 4) dispassionateness, i.e., not having any emotions that are directed upon objects because they are believed to have essentially particular features (an IO could not love someone because s/he is *his/her* child); 5) consistency; and 6) normality “in other respects.”⁴ Firth thinks that IOs will all agree in their “morally significant attitudes” (moral approval and disapproval) about all possible moral questions. Given this and given his version of the IOT, all moral judgments must be either objectively true or objectively false.

that are characterized by a “demand quality,” “Ethical Absolutism and the Ideal Observer,” in Hospers and Sellars, eds., *Readings in Ethical Theory*, second edition (Englewood Cliffs: Prentice Hall, 1970), p. 208. Firth takes the IOT to provide the correct account of the meaning of moral judgments and moral terms. According to Firth, “act x is morally right (wrong)” means roughly “an IO (any IO) would feel moral approval (disapproval) for x if he were to reflect on x.”

⁴ I take this last requirement to mean that an IO must be an otherwise normal human being or *homo sapiens*. In his discussion of this requirement Firth describes the IO as a “person.” This might seem to support Taliaferro’s interpretation. But Firth goes on to say the following: “our conception of the personality of an ideal observer has not yet undergone the refining processes which have enabled theologians, apparently with clear conscience, to employ the term ‘person’ in exceedingly abstract ways. 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 being” (p. 220). The last sentence seems to clearly imply that an IO must be an otherwise normal *human being* who is omniscient, omniperceptive, etc. There is another reason for thinking that Firth takes being a human being to be an essential feature of an IO. In his arguments to show that IOs would all agree in their “ethically significant reactions” to at least some moral issues, Firth appeals to general features of human psychology and thus presupposes that an IO must be a human being. (See “Reply to Professor Brandt,” *Philosophy and Phenomenological Research* 15 [March 1955]: 415-16.)

Given that an IO must be a human being, it would seem that the IOT cannot provide all possible rational beings with a standard for the truth or correctness of moral judgments. It cannot give us a standard by which a Martian could assess the truth or correctness of moral judgments, but, at most, only a standard for human beings to judge the truth or correctness of moral judgments (cf. *The Status of Morality*, pp. 75-76). I don’t think that any moral judgments are true for all possible rational beings, but I think that at least some moral judgments are true for all (empirically possible) *human beings*.

My conviction that IOs could disagree in their attitudes about a great many moral questions motivated me, in Taliaferro's words, to "relativise the IOT." If IOs would disagree in their attitudes (or judgments) about certain issues, then proponents of the IOT cannot say that there is any objectively correct (true) judgment about those issues. But I am not willing to conclude that all judgments about such issues are equally correct or equally valid. Nor am I willing to conclude that judgments about such issues cannot be correct or incorrect in any significant sense. I suggest the following as a criterion for determining whether a particular moral judgment is *correct (true) for a particular person*:

It is *correct* (in a sense that is opposed to mistaken) *for S* to hold a favorable (unfavorable) moral judgment about X if and only if, he would have a favorable (unfavorable) attitude about X, if he were an IO and if he were to consider or think about X. (Among other things, to have a favorable attitude about something is to be disposed, other things equal, to choose or prefer that it exist or occur rather than not.)

Note that this is perfectly compatible with saying that some or all moral judgments are objectively true. We could say that a moral judgment is objectively true (correct) if it is correct for everyone to accept it, or, to put it another way, a moral judgment is objectively true if it is subjectively true for everyone. Note also that I give a different analysis of the "morally significant attitudes of the IO" than does Firth.⁵

Taliaferro makes stronger claims concerning agreement among Firthian IOs than I take Firth to be making himself. Taliaferro claims not only that *human* ideal observers would all agree in their views and attitudes about all conceivable moral questions, but that all *non-humans* who satisfy all of the other conditions (other than that of being human) for being ideal observers would agree in their views and attitudes about all conceivable moral questions.

Taliaferro notes that many of the explanations of ordinary disagreements about moral questions — that one party has relevant information that the other lacks, that one has a more powerful and vivid appreciation of the nature of another individual's experiences — could not be operative if both parties were Firthian IOs (Taliaferro, p. 135). He goes on to ask "What, then, remains to account for IO disagreement in the Firthian theory? His [Carson's] suggestion that moral training might do so seems to be a nonstarter in light of the above considerations" (Taliaferro, p. 135). I am puzzled by the final remark about moral training. Taliaferro

⁵ See *The Status of Morality*, pp. 86-90 for a discussion of this issue and a rationale for my departure from Firth's formulation of the IOT.

does not mention the issue of moral training in any of the three pages which precede this passage. He has done nothing to counter my claim that Firthian IOs might disagree in virtue of differences in moral training.

Taliaferro considers the following possibility. Two ideal observers disagree “because the one has a brain tumor causing it to approve of that of which it would otherwise disapprove. Every time it is led by its grasp of the data, *et al*, to approve of X, it disapproves of X” (Taliaferro p. 135). According to Taliaferro, the individual with the brain tumor is disqualified as an IO, because his attitudes were caused at least in part by something other than his essential characteristics *qua* IO:

this is sufficient to disqualify the tumor-troubled observer from IOhood, for his approval and disapproval is not in virtue of the conditions specified as constituting the ideal moral point of view. Rather, his attitudes are in virtue of his nonethical knowledge, *et al*, plus a brain tumor inverting his judgments. IOs may be imagined to have any kind of biological makeup you like or even to be nonphysical. The essential feature of IOhood is simply that the conditions identified in the Firthian theory are those which properly ground, warrant, justify, and bring about moral attitudes. Odd neurological wiring can make for odd, not ideal, observers (Taliaferro, pp. 135-36).

Firth does *not* make it an explicit condition of one’s being an IO that nothing inessential to one’s being an IO can decisively influence one’s attitudes. In making this argument Taliaferro is significantly revising Firth’s theory. He offers no argument for this revision and assumes without argument that the essential features of an IO are always sufficient to determine the morally significant attitudes of moral approval or moral disapproval in response to the contemplation of moral issues. As I shall argue, there are many other factors which could influence the attitudes of Firthian IOs. Given that Firthian IOs could differ with respect to these factors, they could disagree in their attitudes about many moral issues.

I will defend the following two claims: 1) If we modify Firth’s theory as Taliaferro does by saying that being a human being is not a necessary condition of one’s being an IO, then it would seem that Firthian IOs might disagree in their attitudes about *any* moral questions. 2) If we take Firth to be saying that an IO must be a (“normal”) human being, then Firthian IOs could still disagree in their attitudes about a great many moral issues.

1) The essential features of a Firthian IO (excluding that of being a “normal” human being) radically *underdetermine* his/her attitudes and judgments with respect to moral issues. Given the bare knowledge that some being is omniscient with respect to all non-moral facts, omnipercipient, etc., etc., we cannot and could not in principle know how s/he would react to any moral questions. These highly abstract and indeterminate characteristics generate the morally relevant attitudes of moral approval

and disapproval only by interaction with complex psychological characteristics and dispositions of the moral appraiser. If we say that being a human being is not necessary for being an IO and say that *any* creature (human or non-human) who is omniscient and omniperceptive etc., counts as an IO, we make being an IO compatible with almost any kind of psychological character or make-up. And IOs with wildly differing psychological natures could disagree about anything.

The foregoing can be seen by considering the requirement of “omniperceptive” which is clearly a crucial feature of IOs either on Firth’s theory or my own. Why would anyone think that this requirement helps to insure some measure of agreement in the attitudes of IOs? The answer is that human beings have a deep-seated propensity to sympathize with the experiences of other creatures when they make themselves vividly and powerfully aware of those experiences. We tend to be displeased by the suffering of others and pleased by their pleasure when we make their experiences vividly present to us in thought. But clearly sympathy and benevolence are separable from omniperception — one’s *representation* of another person’s experiences and one’s *attitude toward* those experiences are two different things. That human beings who are omniperceptive are also sympathetic and benevolent is true (if it is true at all) as a matter of contingent fact.⁶ We can easily conceive of rational non-human beings who vividly represent the experiences of other creatures and satisfy all of Firth’s other conditions for being an IO, but who are relatively indifferent to the suffering and well-being of others. Such beings could disagree with human IOs in their attitudes about any conceivable moral questions. (We might imagine a race of intelligent beings whose biological, psychological and social nature is such that sympathy and benevolence do not significantly contribute to their survival or well-being. Such beings are clearly logically possible and there is no reason to suppose that they are not physically or empirically possible.)

⁶ See *The Status of Morality*, pp. 60-65, for a more detailed defense of this claim. R. M. Hare would reject my sharp distinction between understanding another person’s experiences or circumstances and one’s own attitudes, emotions, and preferences about the other person’s experiences. Hare claims that in order to fully understand what it would be like to be in another person’s situation one must *now* desire that if one should find oneself in the other person’s situation with the other person’s desires, then those desires (that one would have) should be satisfied. “I cannot know the extent and quality of others’ sufferings and, in general, motivations and preferences without having equal motivations with regard to what should happen to me, were I in their places, with their motivations and preferences” (*Moral Thinking* [Oxford, 1981], p. 99). I give a detailed criticism of this view in my paper “Hare’s Defense of Utilitarianism,” *Philosophical Studies* 50 (July 1986): 106-7.

2) Suppose that we take Firth to be saying that an IO must be a human being whose psychology is consistent with human nature. Firthian IOs could still disagree about many significant moral questions. They could, for example, disagree about the morality of lying in cases in which lying is necessary in order to avoid acute embarrassment and distress but not necessary in order to avert anything worse than this. (Let me stipulate that in the cases in question lying will not seriously contribute to the deterioration of one's character and is very unlikely to be found out by anyone else.) I have in mind the following kind of case. I hold religious and political views which would be abhorrent to an elderly grandparent and am questioned insistently and at length about these matters by the grandparent. I cannot avoid causing us both acute distress unless I lie. There are at least two kinds of factors which could lead Firthian IOs to disagree in their attitudes about such issues.

A) Our views and attitudes about moral questions are sometimes decisively influenced by the nature of the moral training that we received when we were young. Differences in moral training could easily lead Firthian IOs to disagree about the morality of lying in cases of the sort sketched above. Suppose that IO X was taught to observe very strict rules against lying and taught that lying is never permissible unless it is necessary in order to save the life of an innocent person. He was severely punished whenever he told any lies, and his parents and his loved ones constantly displayed contempt for dishonesty in any forms and great admiration for honesty even when it led to disastrous consequences. IO Y was taught to believe that the moral presumption against lying is very weak and admits of many exceptions. His parents would punish him or scold him for lying, but only if he couldn't offer any kind of plausible consequentialist justification for his actions. X and Y might very well disagree in their attitudes about lying. Given their omniscience, X and Y would know that they disagree only because they received different kinds of moral training. But knowing this would not necessarily cause either of the two to modify his judgments and attitudes about the issue unless he had reason to suppose that the moral principles in which he was indoctrinated were mistaken. But in this case neither X nor Y could have reason to so suppose. According to Firth's version of the IOT, a certain view about the morality of lying is objectively correct or objectively mistaken only if all IOs would agree in their attitudes about the morality of lying. But, by hypothesis, this is not the case; X and Y disagree in their attitudes about the morality of lying. To argue that an apparent ground of disagreement between IOs is not really possible because it would involve one of the parties knowing that he was indoctrinated in incorrect or mistaken moral judgments

would be to presuppose an answer to the question at issue. One would need independent grounds for thinking that IOs would all agree in order rule out my example.

B) Many people are disposed to adopt or identify with the attitudes of parents, peers, or loved ones. If all of X's peers and loved ones accepted something bordering on an absolute prohibition against lying and if all of Y's peers and loved ones accepted a much more permissive view about the morality of lying it would be possible for X and Y to have very different attitudes about the morality of lying.

In my book I defend the stronger claim that Firthian (human) IOs could disagree in their attitudes about *every* possible moral issue. I did not undertake the impossible task of considering every possible moral question and showing that it could be a source of disagreement between Firthian IOs. Rather, I took an issue about which the strongest possible case could be made for moral objectivism and unanimity in the reactions of IOs — the Nazi's attempted extermination of the Jews. I attempted to show that it would be possible for Firthian IOs to disagree in their attitudes about this and concluded that if it were possible for them to disagree about this then they could disagree about anything. I think it obvious that some (most) possible Firthian IOs would be profoundly moved by the tribulations of the victims of the Holocaust and would strongly disapprove of what the Nazis did. But because of certain inadequacies in Firth's conception of the IO, we can conceive of a psychologically plausible Firthian IO who would approve of (or at least not disapprove of) the Holocaust. Suppose that a Firthian IO possessed all of the following characteristics: 1) during his childhood he was indoctrinated in wildly false beliefs about the characteristics of Jews — beliefs to the affect that all Jews are vicious, dishonest, and deeply involved in monstrous plots which threaten the freedom and well-being of almost everyone else. (Given that he is now an IO he can no longer have these beliefs; but attitudes and feelings based upon false beliefs do not necessarily disappear when one comes to reject those beliefs, even if one understands that one has the attitudes only because one had the false beliefs.)⁷ 2) All of those he loves and admires are fanatical Nazis. Thus, simple conformity and identification with others will tend to cause him to adopt their attitudes about Jews. 3) He is a seething caldron of pent-up anger or *resentiment* and is unable to find any other adequate outlet for this anger apart from his hatred for Jews. Since he is an IO he would know that he is displacing hostility on Jews, but it cannot be

⁷ See *The Status of Morality*, pp. 66-67, and a very long footnote on pp. 178-79 for an extended defense of this claim with criticisms of Brandt's notion of "cognitive-psychotherapy."

assumed that mere knowledge or understanding that he is displacing hostility would necessarily cause him to cease doing so.⁸ His rancor and hatred are so great that the sympathy which one would expect on account of his omnipercipience is entirely lacking. 4) His hatred of Jews is so great that his impartiality (or disinterestedness) does not prevent him from being a Nazi. (He would consent to be killed if it could be shown that he himself is Jewish.)

I am no longer as sure of this argument as I was when I wrote the book. Still it seems at least an open question whether a (human) Firthian IO could approve of the Holocaust. My own version of the IOT more effectively rules out the possibility that an IO could approve of the Holocaust. For on my view: i) not only must an IO be fully informed about all relevant matters at the present, but also must have been fully informed at all times in the past (this means he cannot now have any attitudes which are dependent on his having had false beliefs or incomplete information in the past); ii) his attitudes cannot be influenced by association with people who are not IOs (this means that an IO cannot have acquired an attitude by virtue of conformity or identification with someone who acquired that attitude only as the result of some kind of irrationality or cognitive failing); and iii) his attitudes cannot involve the displacement of emotions. These are not arbitrary or *ad hoc* conditions. All can be defended as conditions for the (ideal) rationality of attitudes or emotions (see *The Status of Morality*, pp. 66-72).

In addition to our dispute about the possibility of disagreement among IOs, the other main difference between Firth's theory and my own is that my conception of the IO does not include any requirement of impartiality or "disinterestedness." Firth and Taliaferro are clearly correct to hold that impartiality is "deeply imbedded" in our ordinary moral point of view. In constructing my version of the IOT, I was not guided by a conception of an idealized version of our ordinary moral point of view, but rather by a conception of rationality in the formation of attitudes. (I defend Brentano's view that moral judgments are statements about the correctness of attitudes. A moral judgment is correct if, and only if, the attitudes which it asserts to be correct are correct or reasonable. Thus, the attitudes

⁸ Taliaferro claims that the displacing hostility is incompatible with the omniscience of the Firthian IO (p. 9). For if someone knew that he were displacing hostility he could no longer do so. But this seems to fly in the face of the facts. People sometimes do persist in displacing hostility in spite of the knowledge that they are doing so. I once knowingly displaced hostility aroused by other sources upon an umpire during the course of a baseball game. The felt psychological need to displace emotions can often be very great and cause one to persist in the displacement, even if one is aware of the fact that one is displacing hostility aroused by other sources (see *The Status of Morality*, p. 180 fn. 32).

of someone whose attitudes are ideally rational can serve as the criterion for the truth or correctness of moral judgments.) I dropped impartiality as a requirement of one's being an IO, because I cannot see any reason to suppose that attitudes which are not impartial are *ipso facto* unreasonable or mistaken.

In rejecting the requirement of impartiality I was also moved by the apparent vulnerability of Firth's theory to challenges by amoralists. It seems perfectly possible for someone to ask: "why should I care that all IOs would approve of ——?" "why is it irrational for me to be partial and reject the attitudes of an impartial rational judge? Granted, this is to reject the moral point of view as it is ordinarily conceived of, but what is unreasonable about *that*?" Given my theory, one cannot ask the same sorts of "why be moral?" questions. On my view, if a moral judgment to the effect that act x is wrong is *true for* all (possible) human beings, then all empirically possible human IOs would have an unfavorable attitude about x. Given that all empirically possible human IOs would have an unfavorable attitude about act x (remember that on my theory having an unfavorable attitude about x includes having the on balance preference that x not be performed), it follows that it would be irrational for any human being to have a favorable attitude about x or to prefer, on balance, that it be performed rather than not. For if all possible human IOs would share an unfavorable attitude about act x, then it follows that for any human being, if she were ideally rational, i.e., if she were an IO, she would prefer that x not be performed. Any on balance preference that x be performed can be attributed to some kind of cognitive failing.

There is no doubt that some requirement of impartiality would help achieve a much greater unanimity in the reactions of IOs than would otherwise be the case. This would enable me achieve a much stronger kind of moral objectivism than I was able to defend in the book. Because of this, I have had occasion to regret my decision to exclude impartiality as a condition of IOhood. However, the difference between the two versions of the IOT on this score is not as important as it might seem. To include impartiality as a condition of one's being a IO only succeeds in moving the amoralist's challenge to a different level; one cannot thereby avoid or answer that challenge. The amoralist can then ask "Granted that being moral involves being impartial, why should I be impartial?" or "Why should I do what I would want to do if I were impartial?" Roughly, a requirement of impartiality makes it easier to show that certain acts are (objectively) morally right or wrong (because it makes it easier to show that all IOs would accept a given moral judgment or attitudes consistent with that judgment), but it makes it more difficult to justify the moral

point of view itself. My position makes it much more difficult to show that a particular act is (objectively) right or wrong, but it enables us to avoid altogether the question “Is it rational for me to do what is (objectively) morally right?”