The absolutist criteria of Roderick Firth’s ideal observer theory

Mikael Barclay

Supervisor: Lars Samuelsson
Abstract

Meta-ethical theories take a number of different ontological, epistemic and semantic positions. In 1952 Roderick Firth published the article “Ethical absolutism and the ideal observer”, in which he defends and shares his own version of a theory on the meaning of ethical expressions, referred to as the ideal observer theory (IOT).

The IOT essentially suggests that the truth value of an ethical expression could in principle be determined by knowing the ethically significant reaction it would evoke on an ideal observer (IO), of certain ideal psychological characteristics, should such a being exist. These characteristics are being understood in terms of an ideal practice of justification for actions. For instance, we might hold that in order to be a competent moral judge, we must have sufficient knowledge of the circumstances which we are to assess, or that we are not somehow biased. Firth suggests that an ideal observer has the characteristics of omniscience to non-ethical facts, omnipercipience, disinterest, dispassion and consistency. The theory itself is described as being absolutist, dispositional, objectivist, relational and possibly empirical.

The specific research question of this paper regards the theory’s ability to give a plausible and meaningful explanation as to the meaning of ethical expressions, while maintaining its absolutist characteristic.

The presented conclusion holds that: (i) the ethically significant reaction of IOs cannot be conflicting, (ii) that knowing the characteristics of the IO is not in principle necessary for the form and validity of the theory, (iii) that such form presupposes actual IO characteristics based on an assumption about the human nature and (iv) that ‘IO’ designates a hypothetical reference through a circular definition. And that this, although perhaps not in principle refuting the theory, renders it without the ability to provide any real meaningful explanation regarding the meaning of ethical expressions. A dilemma suggested to be possibly addressed by the abandonment of the theory’s absolutist criteria.
Table of contents

ABSTRACT .................................................................................................................................................. II

TABLE OF CONTENTS .............................................................................................................................III

1 INTRODUCTION .....................................................................................................................................1
  1.1 RESEARCH QUESTION .......................................................................................................................2

2 BACKGROUND .......................................................................................................................................3
  2.1 THE IDEAL OBSERVER THEORY CHARACTERISTICS .........................................................................4
    2.1.1 Absolutist ......................................................................................................................................4
    2.1.2 Dispositional .................................................................................................................................4
    2.1.3 Objectivist ....................................................................................................................................5
    2.1.4 Relational .....................................................................................................................................5
    2.1.5 Empirical .......................................................................................................................................5
  2.2 THE CHARACTERISTICS OF THE IDEAL OBSERVER ..........................................................................6
    2.2.1 Omniscience to non-ethical facts .................................................................................................6
    2.2.2 Omnipercipience ..........................................................................................................................6
    2.2.3 Disinterest ......................................................................................................................................6
    2.2.4 Dispassion ......................................................................................................................................7
    2.2.5 Consistency ....................................................................................................................................7
    2.2.6 Normality ......................................................................................................................................7

3 INVESTIGATION ....................................................................................................................................8
  3.1 SOME FUNDAMENTAL INTERPRETATIONS .........................................................................................8
    3.1.1 Epistemology and moral words .....................................................................................................8
    3.1.2 Practice of justification ............................................................................................................... 9
    3.1.3 Ambiguity ....................................................................................................................................10
    3.1.4 Conflicting IOs .............................................................................................................................11
    3.1.5 The theory’s general form ...........................................................................................................12
    3.1.6 The IO and human nature ...........................................................................................................14
  3.2 INTUITIVE APPEAL ............................................................................................................................15
    3.2.1 Brandt’s explanatory merits ........................................................................................................15
  3.3 THE IO AS A REFERENCE ................................................................................................................16
    3.3.1 Circularity .....................................................................................................................................16
    3.3.2 The open question argument .......................................................................................................17
    3.3.3 Naming and referentialism ..........................................................................................................18
    3.3.4 Designating the IO .......................................................................................................................21
  3.4 REVISING THE THEORY ..................................................................................................................22

4 CONCLUSION .....................................................................................................................................24

REFERENCES ............................................................................................................................................25
1 Introduction

Meta-ethical theories take a number of different ontological, epistemic and semantic positions. *Cognitivists* (in one sense of the word), for instance, hold that there are such things as *moral facts* and by consequence that moral statements can be true or false. *Non-cognitivists* on the contrary suggest the opposite, that there are no such facts and that moral statements simply cannot be true or false. *Naturalists* would have it that there are no *non-natural* moral properties, such as irreducible goodness. This is something that *non-naturalists* refute, arguing that morals cannot simply be described in conventional naturalist terms. *Absolutists* argue that there are acts that are wrong for all agents to commit, while *relativists* hold that moral statements are true or false relative to the agent\(^1\).

In 1952 Roderick Firth published the article “Ethical absolutism and the ideal observer” (*Philosophy and Phenomenological Research*. 12(3):317-345), in which he defends and shares his own version of a theory on the *meaning of ethical expressions*, referred to as the *ideal observer theory* (I will use the abbreviation IOT). This is the main source of this thesis and the source I refer to when referring to the IOT according to Firth. I will refer to it with page numbers in brackets. In his article Firth suggests that, although the field of *moral philosophy* overall has been well examined, there is one position that has not, that of a theory that is *both absolutist and dispositional* (p.317). I will clarify later what Firth intends with ‘dispositional’. Another way of viewing it, is that the theory is a welcome contribution in its attempts to take the position of an *absolutist and naturalistic* theory. A position merited of both giving an explanation to the notion that we in disputes appear to hold that there is such a thing as an *actual* right or wrong, as well as allowing us not to presuppose some form of *supernatural* moral properties.

1.1 Research question

My particular research question is as follows:

*Can Roderick Firth’s ideal observer theory give a plausible and meaningful explanation as to the meaning of ethical expressions, while maintaining its absolutist characteristic?*

I will attempt to answer this question by showing that according to the IOT (i) the reactions of different IOs cannot be conflicting, (ii) that it is not the IO characteristics as presented by Firth that are necessary for the validity of the theory, but its *general form*, (iii) that this form presupposes something about the human nature, (iv) and finally that the defence of this form and the absolutist position leaves the theory with a fundamental dilemma.
2 Background

The IOT could be suggested to have a great deal of intuitive appeal, embracing what we commonly hold to be valid justifications for our actions, as well as explaining why we believe it possible to have moral disagreements\(^2\). In addition, according to Firth (317-318), it could satisfy meta-ethical positions that otherwise could be held as conflicting one another. Firth expresses the fruitless search of a feasible analysis of ethical statements as:

“…This concentration of effort by many acute minds has not produced any general agreement with respect to the solution of these problems; it seems likely, on the contrary, that the wealth of produced solutions, each making some claim to plausibility, has resulted in greater disagreement than ever before, and in some cases disagreements about issues so fundamental that certain schools of thoughts now find it unrewarding, if not impossible, to communicate with one another.” (p. 317)

In essence what Firth suggests with the IOT is that the truth value of an ethical expression could in principle be determined by knowing the *ethically significant reaction* it would evoke on a person, or *observer*, of certain *ideal* psychological characteristics, if such a person were to exist. Or very bluntly put, to claim that an action is *right*, is close to saying that it is an action that would be *approved* of by an IO, if such person were to exist. This last explanation is not necessarily precisely what Firth intended, since he doesn’t attribute any particular sentiments of *approval* to the IO. I do however find that it is a useful way of expressing an otherwise partly vague description of the nature of the reaction of the IO. The simplified expression I will use to describe the IO, is that he is a *hypothetical agent, reacting in a way that reveals moral correctness*. I will mostly use *correctness* for its convenient short form instead of using something like *rightness or wrongness* which would be another suitable simplification.

Before my attempt to analyse the theory, I will try and give a brief overview of it as presented in the two parts of Firth’s article.

2.1 The Ideal observer theory characteristics

In the first part of his article, Firth sets out to describe the characteristics of the analysis (p.318-329). I will here try and convey the essence of his description following the sections as they are presented in the article.

2.1.1 Absolutist

An analysis of ethical statements is described as relativist if it proposes ethical expressions to be dependent on egocentric expressions. By ‘egocentric expressions’ Firth intends expressions “that are ambiguous in abstraction from their relation to a speaker, but their ambiguity is conventional and systematic” (p.318). Essentially then, expressions whose meaning is relative to the speaker, typically including personal pronouns (such as I or they), possessive adjectives (such as mine or their), reflexive expressions (such as he who is speaking). Firth suggest that these expressions could be defined in terms of the word this.

By consequence, if one believes the meaning of ethical expressions to be understood in terms of egocentric expressions, one is a relativist. In such case one might translate the meaning of the sentence it is wrong to be impolite into impoliteness upsets me or I urge you to be polite or most educated people of today would disapprove of someone who is impolite. Firth, by contrast advocates a non-relativist or absolutist theory, suggesting ethical expressions to be true or false, consistent or inconsistent with each other, independently of its speaker. (p.318-320)

2.1.2 Dispositional

The theory is described as dispositional, by which Firth intends that it asserts that actual or hypothetical beings are disposed to react to certain acts in a certain way, and that ethical statement may always be formulated as being contrary-to-fact conditional. This is to say that a statement such as it is right to be polite would mean that if a certain being (IO) existed, he would react to politeness in such way that its rightness could be asserted. The conditional being is disposed to react in a certain way, and thus the theory is referred to as being dispositional. The analysis, Firth suggests, could involve assertions of three types: that of actual, possible or a majority of beings. While postulating the existence of an actual being of this kind would essentially mean a reliance on God, an analysis on a majority of beings would lead to empiricism, something that he refutes. What remains then is an IO corresponding to a
possible being, ideal in a similar sense as that of a machine being frictionless. This perfection of the IO is attributed his characteristics as such rather than some virtuousness of his person. The characteristics being such that they make him react in a way that reveals moral correctness. (p.320-322)

2.1.3 Objectivist
The theory is described as being objectivist, or non-subjectivist, in the ontological sense of suggesting that ethical expressions are not per definition false should there not exist an experiencing subject. In other words, the truth value of a contrary-to-fact conditional proposition, is not dependent on the existence of an actual IO. What matters is not whether or not the IO actually exists, but how he would react if he did. (p.322-324)

2.1.4 Relational
The theory is described as relational. This is to say that it asserts that when an ethical term is applied to a thing, that thing is then related to something else. In this case to the reaction of the IO. Firth uses an analogy with the colour of a daffodil to explain: “…to say that a daffodil is yellow is to say something about the way the daffodil would appear to a certain kind of observer under certain conditions” (p.324). Yellowness could then be seen as a relational property of the daffodil. Similarly, rightness would also be a relational property. (p.324-325)

2.1.5 Empirical
Firth refers to the IOT as empirical, explaining that an absolutist dispositional analysis very well might be. “Such an analysis would be empirical, for example, if the defining characteristics of an ideal observer were psychological traits, and if the ethically-significant reactions of an ideal observer were feelings of desire, or emotions of approval and disapproval, or some other experiences accessible to psychological observation.” (p.325). These ethically-significant reactions, he refers to as moral data. And much like when observing a yellow daffodil, we appeal to our colour sensation to justify its actual yellowness, we are suggested to appeal to our moral experiences or data when justifying a moral belief. The notion that it would be necessary for an ideal observer to have moral beliefs in order to have moral experiences, is refuted by Firth. He argues that one could easily imagine a situation in which one would feel approval about some moral act, only to afterwards be in doubt of the justification of that feeling. In other words, the IO could have moral experiences
2.2 The characteristics of the ideal observer

In the second part of the analysis, Firth presents what he refers to as the characteristics of the IO. Important to note is that Firth suggests that we can validate the form of the analysis without actually being able to formulate a correct analysis of an ethical statement. The extent to which we can formulate a correct analysis is dependent on the success of formulating details of the IO’s characteristics. These characteristics should reflect our rational decision procedures in moral questions, or as Firth puts it: “In analyzing ethical statements, for example, we must try to determine the characteristics of an ideal observer by examining the procedures which we actually regard, implicitly or explicitly, as the rational ones for deciding ethical questions.”

2.2.1 Omnipotence to non-ethical facts

Without knowing the facts relevant to a particular ethical question we would typically see our justification of it as inadequate. We simply need to know the factual circumstances of an event to have a valid moral opinion on it. In the case of the IO, this is handled by the attribution of omniscient knowledge with regards to non-ethical facts. Non-ethical to avoid circularity, since ethical facts is what is intended to be established.

2.2.2 Omnipercipience

The lack of imagination of such things as other people’s feelings, is also something practically deemed as a cause for a lack of adequacy of someone’s moral judgement. The IO is consequently attributed omnipercipience, perfect visualisation skills.

2.2.3 Disinterest

To be biased, is commonly seen as immoral. Firth suggests that the ideal observer needs to be in some way impartial to particular interests, such as the affection of family members. He argues that one would have to somehow find a degree of disinterest in these particulars that is not too narrow, nor too wide.
2.2.4 Dispassion
The ideal observer is suggested to need a lack of emotions, such as jealousy, which could obstruct correct moral judgement. To attribute him virtues such as love of mankind would not necessarily be circular, Firth argues, in so far as they were assigned as characteristics for other reasons than for being virtues in their own right. (p.340-341)

2.2.5 Consistency
A good judge arguably needs to be consistent in his judgements, reacting in the same way to the same particular act. Firth argues that consistency must be a trait of the IO. This is a trait that, in a successful dispositional analysis, has to be a consequence of the other characteristics, rather than some independent characteristic of its own. A more general consistency, governing responses to similar but different acts following the same ethical principles would become circular in that it would have to presuppose some ethical proposition. (p.341-344)

2.2.6 Normality
In other aspects the ideal observer is to be considered normal and very much human. What exactly this normality entails, is not defined by Firth, although he asserts that it in principle could be. I would interpret this section as a reassurance of us being able to imagine the IO as a person, albeit with the previous characteristics, rather than say a God. (p.325-329)
3 Investigation

In this chapter I will attempt to analyse the theory with a particular focus on the very meaning of the term ‘IO’. Before I start, I would like to clarify my interpretation of some key positons of the theory. I want to do this in particular since I have come to hold the IOT to be susceptible to interpretations that, although fundamentally different, are easy to assume as self-evident. After this I will comment on what would appear to be some intuitive support for the theory. I will then proceed to an analysis of the meaning of ‘IO. Finally, I will give my view on how the theory could be revised to avoid my criticism.

3.1 Some fundamental interpretations

3.1.1 Epistemology and moral words

In his article “On the moral epistemology of ideal observer theories”, Jason Kawall attempts “to defuse a set of epistemic worries commonly raised against ideal observer theories”\(^3\). I will not go into depth on this subject but rather clarify that I do not find epistemic critique relevant to the principle validity of the theory since the meaning of ethical expressions is not suggested to be dependent on our ability to make a correct analysis. Or as Firth clarifies, “Any plausible formulation is certain to be very complex, and there is no reason to suppose that philosophers could ever reach complete agreement concerning all the details of an adequate analysis. But this in itself should not prevent philosophers from agreeing that this general form of analysis is valid.” (p.330). Because my interpretation of the IOT is such that I consider it independent of such epistemological concerns, they shall be outside the scope of my discussion.

Nor will I treat particular clarifications on ethical expressions with regards to such things as prescriptiveness, being evaluative or descriptive or concerning instrumental or intrinsic value\(^4\). The reason being that it is far from obvious how and if these concerns would present problems for the IOT. Whatever the meaning of ethical expressions, it is analysed in virtue of

---


the reactions of the IO, and I shall later explain how concerns regarding the ambiguity of moral words is considered by the theory.

3.1.2 Practice of justification
Strömberg suggests that the IOT, as presented by Firth, indicates a sort of pragmatic view on the meaning of the moral language, as something being manifested by our “practice of justification”\(^5\). Garner on the other hand, refers to the theory as a proposal of what ethical expressions ought to mean, and stresses its distinction from our intentions. “It would be absurd to treat the ideal observer theory as an attempt to state what people actually mean by "intrinsically good" or "right" for obviously most people would deny that they mean anything at all like this when they employ those expressions”\(^6\). He also states “The point here is that the open question argument cannot be applied to a definist theory which is a proposal or a recommendation, since such a theory is not a theory about what normative expressions mean, but rather a proposal, put forward for consideration, about what they should mean.”\(^7\). It would appear then that while Strömberg argues in favour of the theory explaining the actual meaning of ethical expressions, Garner seems to suggest that it proposes a meaning of ethical expressions, and as such avoids Moore’s\(^8\) reductionist critique. So which of the two interpretations seem most plausible?

I find it unclear what a semantic proposal would entail. My position is that a theory either does, or does not explain the meaning of ethical expressions. I see no procedure of acceptance of a proposal. What more, Firth is perfectly aware of the ambiguity of ethical expressions and suggests that the absolutism of his theory stands nonetheless. He states: “The thesis


\(^{7}\) Garner, p.620.

maintained by the absolutist as such, is simply that ethical statements are not *conventionally* ambiguous in a manner which would require them to be analysed by means of an egocentric expression; and this thesis is quite consistent, of course, with the proposition that ethical statements are accidentally ambiguous—perhaps even more ambiguous, indeed, than most other statements.” (p. 331). My interpretation then is that the IOT is claimed to handle ambiguity and that it is intended as an *interpretation* of an ideal meaning, rather than a *suggestion* of the meaning of ethical expressions. I will further explain my position on what such ideal meaning must entail.

3.1.3 Ambiguity

One of the important interpretations of the IOT is with regards to its *absolutist* claim. As we have learnt, Firth intends that his analysis is not dependent on *egocentric expressions*, by which he intends “…expressions of which the meaning varies systematically with the speaker. They are expressions which are ambiguous in abstraction from their relation to a speaker, but their ambiguity is conventional and systematic.” (p.318). A relativist analysis is such that it’s interpretation of ethical statements is understood in terms of egocentric expressions. (p.319).

I would then like to argue, that Firth intends that the absolutist claim concerns the *analysis* of ethical expressions, as opposed to the *intention of speakers*. In other words, even if two speakers, A and B were to have absolutist beliefs and *intentions* when uttering an ethical expression, that in itself would not make our analysis absolutist. Firth consistently refers to an *analysis*, rather than the *intention* of speakers. He also acknowledges the fact that different people indeed can intend different things with the same ethical expressions, but that this would not affect the *analysis*. He states: “Any such analysis, if it is at all plausible, is certain to assign a number of complex characteristics of an ideal observer, and to refer to complex psychological phenomena in describing the nature of these ethically-significant reactions.” (p.330). In other words, according to Firth, if the IOT were at all possible, it would handle differences of intention as *ambiguity*, keeping the analysis from becoming relativist. The meaning of ethical expressions is in other words not suggested to be understood in terms of intention, but from an analysis referring to the IO.
3.1.4 Conflicting IOs

Although Firth admits the possibility of multiple IOs, I suggest that he does not intend that their reactions can be inconsistent with one another. He states: “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.” (p.329). In other words, all IOs must react in a non-disagreeing manner or consistently with one another, and it would seem that they as a consequence would have to share their fundamental characteristics. If we intend that, and I shall later argue that we must with reference to the theory’s absolutist claim, meaning according to the IOT is given with reference to actual dispositional IOs, as opposed to beliefs of the same, we must agree that the IOs must always react consistently with one another.

The requirement to reacting consistently with one another is also expressed with reference to the possibility of a God. Firth states: “And it is to be expected, consequently, that any plausible description of an ideal observer will be a partial description of God, if God is conceived to be an infallible moral judge” (p.333) (from now on, when I refer to God I shall intend God as an infallible moral judge). The notion of the IOT as compatible with the existence of God would similarly exclude the possibility of disagreeing IOs. If there is a God, all IOs would have to react in accordance to him/ her, or their reactions would not be the ideal and morally correct at all.

Non-disagreeing IOs of the same characteristics, seems to be a common interpretation of the IOT. Brandt\(^9\) defines the IOT in terms of the IO’s characteristics, Jonathan Harrison declares that “Either 'X is right' means 'All ideal observers approve of X,' which is a universal proposition about a null class, or 'X is right' means. If A is an ideal observer, he will approve of X,' which can only be known to be true if in fact all ideal observers do approve of X…”\(^10\). Taliaferro states “Unlike earlier ideal observer theorists, like Roderick Firth, Carson contends that IOs may disagree.”\(^10\), after which he defends the theory’s status as absolutist. I will argue that the characteristics as stated by Firth are not at all what defines the IO. I do however agree

\(^9\) Brandt, p. 407

that all IOs need in all essence share their characteristics in order for them to have reactions that are consistent with one another.

3.1.5 The theory’s general form

If the IOT does not permit relativism, and if we consequently believe that all IOs are non-conflicting, it would appear that we also must assert that all IOs essentially share the same characteristics. I would argue however, that which we on reflection find to be morally correct justifications for our actions, does not seem given or universally agreed upon, even after thorough contemplation. Philosophers for instance, do not seem to have concluded that they are in agreement with respect to these characteristics. Take for instance Harrison, who stated “I do not like Professor Firth's suggestion that the ideal observer must be dispassionate, since it seems to me that a being who had no passions - or, if moral reactions are themselves passions of a sort, no other passions - would have no moral reactions.”

11 Or Carson12 who suggested new IO characteristics altogether. Or we might consider a situation in which speaker A and B were to utter the same ethical expression, but with reference to fundamentally different justifications. Let’s pretend for instance that A is a utilitarian who subscribes to the idea of a disinterested IO who put equal emphasis on the experienced pleasure of all people. And suppose B on the other hand is a devoted believer of care ethics, arguing that that an IO could not be disinterested at all. In such case we could imagine that they, if analysing their justifications of actions, would not at all come to an agreement about what an IO would be like, and by consequence how such observer would react. And indeed Richard Hare, defends what could arguably be defined as a version of the IOT, in which the IO is defined as a utilitarian archangel13. What I suggest this demonstrates, is that our practice of justification is not at all consistent, not even when described by philosophers. Different individuals may have different ways of justifying acts and if we suppose there being an ideal and actual set of IO characteristics we must simply admit that our own attempts to describe these may be flawed. Our beliefs of these characteristics, or our intention upon uttering an ethical expression, will as I have previously argued not satisfy the theory’s absolutist criteria. If meaning is not given in virtue of an actual and idealised IO but from intention, meaning would be relative of the speaker.

11 Harrison, p.260
12 Taliaferro, p.129
Epistemic concerns regarding the actual characteristics of the IO would not however, necessarily speak against the IOT. It is not evident that the theory need be dependent of the characteristics as described by Firth, who states:

“Any plausible formulation is certain to be very complex, and there is no reason to suppose that philosophers could ever reach complete agreement concerning all the details of an adequate analysis. But this in itself should not prevent philosophers from agreeing that this general form of analysis is valid.” (p.330).

I do not suggest that Firth necessarily intends that his version of the IO need be correct, be that as it may. I do however find this important to underline simply because it is one of the interpretations of the IOT where one can make an assumption that others do not share. For instance, one could argue that Firth holds that it is only minor details of the characteristics that could be questioned. After all, both Brandt and Taliaferro introduce the IOT in terms of Firth’s characteristics. A comment on this however, is that I find it difficult to see how we could agree at all on characteristics too vague to be described in detail.

The interpretation of Firth’s characteristics as necessary, would however introduce a serious problem, which is illustrated by Firth’s question: ”…What are the defining characteristics of an ideal observer?…” (p.329). The question in itself reveals an acceptance of the theory’s independence of our ability to accurately describe the IO’s characteristics. When we set out to try and find the characteristics of the IO, we do not yet know them, but we try because we can imagine the IOT valid in principle.

As a consequence of the above, I would argue that the only thing we must know for certain about the IO, for the IOT to be valid in principle, is that he is such that he has the ability or property of reacting in a way that reveals moral correctness. Unlike the other characteristics of the IO, this property is the only thing that we know about the IO to be analytically true. While we have to accept that we could be mistaken, perhaps influenced by personal emotions, when describing the characteristics of the IO, it is impossible to imagine an IO without the property of reacting in a way that reveals moral correctness.

14 Brandt, p. 407
15 Taliaferro, p.124
3.1.6 The IO and human nature

In the article *Relativising the Ideal Observer Theory*[^16], Charles Taliaferro sets out to defend the IOT against Thomas Carson’s relativist version in which is suggested that “Some given act A, may be both morally right and wrong at once in that Eric, if an IO, would approve of it and Miriam, if an IO, would disapprove.”[^17] Carson further suggests that “Firth’s theory does nothing to rule out the possibility that different ideal observers could have received radically different kinds of moral training”. Martians and humans, he suggests, could have attitudes different enough to effectively make for disagreeing IOs. Taliaferro on the other hand maintains that “…impartial empathetic knowing, and attention to states of the world is enough to give rise to IO judgements.”[^18] And that the IO judgement have “…equally powerful grasp of the mental life of all involved, and not prompted by particular interests and passions.”[^19]

Not unlike Firth’s approach to ambiguity, Taliaferro referrers to the psychological characteristics of the IO.

So, while Carson attributes the characteristics of the IO to the *human nature*, Taliaferro hold them to be *universal*. I happen to agree with Taliaferro, but regardless of this, the discussion raises an important question. If the IOT is to be independent of the existence of God, then in virtue of what, or perhaps according to who’s ideal practice of justification would the IO’s characteristics be valid? If we agree with Carson, this would be in virtue of the human nature. But if we agree with Taliaferro, we must then suggest there being an ideal practice of justification for something like *all possible moral beings*. Regardless of who is right, it seems that there needs to be a practice of justification that is *relevant* for humans. After all Firth deems it meaningful for humans to, with reflection, try and characterise the IO (p.332). I would argue then, that if we believe that humans would react as or agree with IOs under certain circumstances, and if we believe that IOs cannot have conflicting reactions, we essentially believe something about the human nature. We believe that humans, and possibly other forms of moral agents, be it under ideal circumstances, would accept some *common*

[^17]: Taliaferro, p.133
[^18]: Taliaferro, p.137
[^19]: Taliaferro, p.135
practice of justification. In other words, the IOT as an absolutist theory, presupposes something about the human nature.

3.2 Intuitive appeal

The IOT would appear to have a number of intuitive merits. Perhaps this is the most appealing argument for the theory. For instance, it might be tempting to view the meta-ethical position of the theory as something meriting in itself. Firth (p.317) points out its position as absolutist and dispositional and Bergström introduces it as naturalistic.20 I will not argue that Firth nor Bergström explicitly argues that the position in itself is an argument for the theory, but merely clarify that it is not. In this section I will attempt to comment on some of the other intuitive merits of the theory.

3.2.1 Brandt’s explanatory merits

The theory’s capacity to give explanations to ethical questions, would appear another argument in its favour. Not least Richard Brandt emphasised the theory’s explanatory power21 and suggested the following merits about the theory:

“(1) that this theory enables us to regard as really relevant to ethics all the facts which on reflection we take to be relevant; (2) that it enables us to explain the heterogeneousness of the actions which we regard as right or wrong; (3) that it explains how ethical disagreement is possible even when there is agreement about the nature of the act being appraised; (4) that it explains why our feelings and attitudes-and especially our sympathies—are (and properly are) engaged in ethical reflection, and why moral philosophers have thought that moral experience is distinctively a union of cognition and emotion; (5) that it enables us to hold that moral opinions are subject to objective criticism and are correct or incorrect; that (6) it explains why we value the advice of knowledgeable, impartial, and consistent persons at times of moral decision, and why we reject previous moral opinions of our own which we think reflect self-interest, inconsistency, or lack of information; and that (7) it enjoys advantages over the emotive theory such as the capacity to give a satisfactory analysis of "ethical relevance", and the

21 Brandt, p. 407-413
ability to explain why ethical judgments do not always correspond with favorable or unfavorable attitudes on the part of the judge.”

I will not argue that explanatory power isn’t relevant, but it is important to point out that it is not conclusive evidence and doesn’t rule out other possible explanations to the presented problems. It would for instance appear that Brandt’s list does not exclude relativism. The criteria would be equally met by a group of speakers who attributed a correct practice of justification to a presumed IO, reflecting the norms of that particular group. But as this would allow for conflicting IOs, it could not be permitted by an absolutist IOT.

3.3 The IO as a reference

Harrison states that he agrees with Richard B. Brandt, when suggesting an alternative interpretation to the IOT, that bases the ethically significant reactions exclusively on the characteristics of the IO. This, he suggests, would avoid the theory’s dependency of the IO, which he argues is a null class which would render the analysis meaningless. I shall instead give the theory the benefit of a doubt as to the null class argument since I have suggested that it is rather the IO that is indispensable to the theory and not the characteristics.

In this section I intend to analyse what ‘IO’ must refer to, given that we so far seem to know nothing of substance for certain.

3.3.1 Circularity

Firth himself raises the concern of circularity, and states: “But if the ethically-significant reaction of an ideal observer were the belief (or judgment) that a certain act is right or wrong, it is evident that an absolutist dispositional analysis would be circular: it would contain the very ethical terms which it is intended to define.” (p.326). Instead he suggests that the reactions of the IO are not judgements, but rather feelings with which acts are justified.

However, when we talk about the IO we cannot do so in terms of some specific characteristics, as we cannot claim to know them for certain. With absolute certainty, we know the IO in terms of having certain ethically-significant reactions that reveals moral

---

22 Brandt, p. 407
23 Harrison, p.258
correctness. We must now suppose that the meaning of ‘ethically-significant reactions that reveal moral correctness’, is different than that of ‘moral judgement’, and that it is different in a way that lets us escape circularity dilemma. However, I would argue that despite such supposed difference, the circularity dilemma remains in one important sense. It remains in the sense that, if we cannot define the IO in terms of his characteristics, we seem forced to do so in terms of his reactions alone. However, this cannot help us since in such case we do not know what ‘moral correctness’ means. Such attempt would look something like:

The IO is a hypothetical agent, reacting in a way that reveals moral correctness

**Moral correctness** is that which is revealed by the reactions of the IO

Another way to put it is that, to be meaningful, we require a definition of the IO other than “an agent who always gets it right.”\(^\text{24}\). Of course, one might argue that the IO after all is defined in terms his characteristics, whichever these may be, regardless of whether we know them for certain or not. But in such case the characteristics are such that they define an IO, but that we when talking about one, would still not know for certain about what we were talking. In other words, the meaning of ethical expressions would be something that had a meaning, only one that we could not know anything about. I will argue that not even this is necessarily a problem for the form of the theory, but that it is when it comes to its meaning. Because however we twist and turn this matter, it would seem we are forced to define the IO without the substance that we would need to be enlightened as to the meaning of ethical expressions.

### 3.3.2 The open question argument

One common argument against naturalist theories, is that of G. E. Moore’s *open question argument*. This could be explained as follows: Say that ‘A’ is a naturalistic expression such as ‘pleasure’, and ‘B’ a moral one such as ‘good’. If we propose that the claim *X is A* is the same as claiming that *X is B*, this would mean that the question as to whether *A is B* would be the same as asking if *A is A*, which would be meaningless. Moore by contrast suggests that to ask if *pleasure is good* is not at all meaningless, but an *open* question. The argument more

specifically aims its critique against reductionism, the notion that morals can be reduced to something else\(^25\).

Following the same reasoning, we might be able to shed some light as to our understanding of the IO. Say that ‘A’ is the naturalistic expression ‘approved by an observer who acted like God, the infallible moral judge.’ or else ‘approved by an agent that reveals moral correctness’. And let ‘B’ be the moral term ‘right’ and ‘X’ an act of some sort. If \(X \text{ is } A\) has the same meaning as \(X \text{ is } B\), this would imply that the question if \(A \text{ is } B\) would be closed. In these examples this does not seem problematic, but rather an analytical truth. If we by contrast were to say that ‘A’ is the naturalistic expression ‘approved by an observer with the characteristics \(\varphi\)’, it would leave the question open. If Moore was right, perhaps the IO cannot be reduced to a reaction based on a number of rational characteristics. But if he was not, it would nevertheless appear that we would need to show why the first two examples actually designates the same IO as the last. Or put differently: why would an IO defined by some characteristics \(\varphi\), react in the same way as God the infallible moral judge? I will later argue that these characteristics \(\varphi\) must be a reflection of our human nature, which will increase the stake. And, in the case we define the IO without a God, how can we do so, if the only thing we know for certain leaves us with circularity?

3.3.3 Naming and referentialism

```
"I saw a Heffalump today, Piglet"

'What was it doing?' asked Piglet.

Just lumping along,' said Christopher Robin.

'I don't think it saw me.'\(^26\)
```

In this section I shall use an analogy to introduce a terminology, to show that the IOT is a referentialist theory with a circular reference, but that even this in itself need not be a


problem. Instead the fundamental problem is that, even if there in fact were some set of universally valid IO characteristics relevant to humans, even if they also in principle were consistent with the will of a good God, and even if we could refer to such an IO without knowing anything about him for certain, apart from a circular definition, this would leave us with a theory that although possibly right in form, lacked the ability to provide meaning to ethical expressions.

In A.A. Milne classic *Winnie the Pooh*\(^{27}\), the readers are introduced to what appears to be an imaginary animal referred to as a ‘heffalump’. First mentioned by Christopher Robin, the heffalump becomes a general subject of conversation by the other characters, all acting as if they *actually* knew the meaning of the word ‘heffalump’. Little by little they start to explore the characteristics of the animal. Pooh asks himself whether heffalumps are *fierce* or if they *come when you whistle* and Piglet supposes that they *like acorns*. The two friends decide to set up a trap, hoping to catch one.

The *referentialist theory* suggests that the meaning of a name is the very thing that it *designates*\(^{28}\). One theory on the Milne’s ‘heffalump’ word, is that it came from a child’s attempt to say ‘elephant’. Assuming that Christopher Robin with ‘heffalump’ really intended to say ‘elephant’, we could with the referentialist theory argue that the meaning of ‘heffalump’ in fact would be *elephant*. Gottlob Frege\(^{29}\) points out one of the difficulties with this view. If a=b, in virtue of designating, or referring to, the same thing, then a=b is *analytic*, just like a=a. In other words, the sentence “a heffalump is an elephant” would have the same meaning as “a heffalump is a heffalump” which would not be meaningful. Instead Frege suggests that a name ‘a’ not only has a reference but also a *sense*. That is, the sense of ‘heffalump’ and that of ‘elephant’ differs, although their common reference does not, rendering the sentence meaningful after all. Another dilemma with the referentialist theory is that of names that do not denote any objects at all, such as ‘Pegasus’\(^{30}\). In the case that the heffalump in fact is an *imaginary* animal, the meaning of ‘heffalump’ would then be hard to explain with the referentialist theory.

\(^{27}\) Milne, *Winnie the Pooh*

\(^{28}\) Lycan, p.3


\(^{30}\) Lycan, p.3
Now, let’s say that Pooh and Piglet actually did manage to catch something in their trap. In virtue of what could they claim that this something was such a thing that ‘heffalump’ designated? Could they, if they caught something claim that this was what they had been talking about all along? The idea is amusing, because after all it seems completely unclear what they actually refer to when they use the word ‘heffalump’.

Saul Kripke\textsuperscript{31} gives his account on naming and how names or designating expressions relate to their reference. This is not, he argues, in general through corresponding properties $\varphi$ of the names or our beliefs, but rather through a sort of history of how we came to know the particular name, through others in our community. Say for instance that we know the name ‘Columbus’ as the first European to reach the western hemisphere and as the man who discovered that the earth is round. Kripke argues that even if these things turn out not to be true, this would still not mean to say that ‘Columbus’ refers to the persons who actually fulfilled these properties. Designating properties, he argues, are not always necessary truths, but rather contingent. We can, he suggests, simply refer to something in virtue of hearing someone else making a reference. If this were true, we could then in the Heffalump case make the Kripkean interpretation of arguing that Pooh and Piglet are referring to whatever Christopher Robin was referring to. And if Christopher Robin indeed overheard his father, who actually was speaking about elephants, we could say that the reference of ‘heffalump’ in fact is elephant. The meaning of ‘heffalump’ would simply be elephant. However, in the case that Christopher Robin just imagined the animal, we end up in the situation of trying to characterise something that completely lacks reference, other than that in the minds of our friends.

If instead we were to believe that the ‘heffalump’ was the scariest animal in the forest, this would, given that there were any scary animals in the forest at all, pick out one of these animals. Only our friends would not know which one. Of course they might try to imagine its properties, but if we subscribe to Kripke, our descriptions of the Heffalump would not be necessary but rather contingent. After having designated an actual animal to the name, it would hold even if it later turned out that this animal was actually not the scariest animal in the forest.

\textsuperscript{31} Martinich, p. 53-66
3.3.4 Designating the IO

Let us translate the previous illustration to the IO. Firstly, it appears that the IOT does not intend that meaning is given by *sense* of the IO for, as previously argued, it is not our beliefs on a correct practice of justification that constitutes his characteristics. The IOT would instead appear *referentialist*, referring to a hypothetical agent in which virtue we are to understand the meaning of ethical expressions. Harrison states: “Professor Firth can do no better by accepting Mr. Strawson's suggestion that the question of the truth or falsity of universal statements only arises if they are not about null classes, for Professor Firth wishes to maintain that the question of the truth or falsity of ethical statements arises whether there are ideal observers or not.”\(^{32}\), and further that this would be problematic. I would argue that referring to hypothetical agents, or *null-classes*, should be possible because the IOT is defined in such a way that lets a hypothetical IO play the exact same role as an actual IO would. And I shall assume that we might still, in line with Kripke, designate the IO without knowing his characteristics, or *designating properties*. The problem is not designating the IO, the problem as we shall see, is to maintain that the IOT does what it sets out to do.

3.5 The IO-reference dilemma

Richard Hare suggests that “The crucial difference between linguistics and moral institutions is that the latter, unlike the former, claim to establish matters of moral substance.”\(^{33}\)

Linguistics, he suggest, is such that it either empirically renders the native speakers its authority or that it, in philosophical logics, has it that it is the *intention* of the speaker that is authorative. Neither of which, he stresses, can “…yield either statements or precepts of substance about morals”\(^{31}\). In other words, they do not establish the correctness of what the speakers affirm. Such reasoning, he argues, would lead to *subjectivism* or *relativism*. Now, as we have seen, the IOT is not concerned with empirically establishing the use of our language, nor the intention of the speakers. It intends to produce an *analysis* of the meaning of ethical expressions, in virtue of the reactions of the IO.

As we have seen, it is not enough to simply say that the IO is *an agent that reveals moral correctness*. On the other hand, we cannot claim to know the IOs characteristics. This however need not refute the theory since it might still be valid in virtue of its *form*. And we

\(^{32}\) Harrison, p.257

might argue that we, in line with Kripke, could designate and refer to the IO without knowing any of its properties and with a circular designation. Concerns regarding differences in linguistic practice, we dismiss with reference to non-conventional ambiguity. However, if we successfully defend the theory against all of these concerns, it seems it leaves us with one fundamental dilemma regarding an absolutist IOT.

If we neglect clarifying the IOTs positon on human nature and its relation, through the IO, to God, all with reference to it being a semantic theory, it would appear contradictory to at the same time neglect clarifying its linguistic position, for instance defending referentialism. Central to all of this, is the very understanding of the meaning of ‘IO’. As I have previously argued, we must admit that with theory’s absolutist claim, follows an ontological assumption about human nature. And because of the presumed relationship between this human nature and a possible God and infallible moral judge, the theory also introduces a meta-ethical assumption. In essence, the IO represents our human nature which in turn is in principle consistent with the will of God. Again it is important to stress that the analysis of the IOT does not reflect what we intend with an ethical expression but what that ethical expression actually means in virtue of the IO. And the reactions of the IO must be consistent with that of a possible infallible moral judge. If this is correct, we after all must admit that the IOT is first and foremost about moral substance, its analysis in principle telling us what is right and what is wrong. Its semantic position not being more complicated that presupposing referentialism of a hypothetical agent.

I would therefor argue that the dilemma of the IO reference, is that while its dispositional nature seemingly lets us avoid meta-ethical and ontological claims, its absolutist claim does not. And if this is correct, the theory ought to either occupy itself with a clarification of these claims or undergo a revision.

3.4 Revising the theory

Hare’s presents what he considers to be three acceptable forms of purely linguistic hypotheses on the meaning of moral words. The first one, completely linguistic, considers the speech practise of native speakers and what they would accept as a non-self-contradictory use of moral words. The others are what he refers to as hypotetico-deductive procedures. The second

34 R.M. Hare, Moral Thinking, New York, Oxford University Press, 1981, chapter 1.3. E-book
presupposes that most humans share certain common desires and valid form of reasoning. A hypothesis about the meaning of moral words, consistent with the actual moral reasoning and opinions of the speakers, would allow us to provisionally hold the hypothesis to be true. This, form, he argues, very much in line with my previous arguments on the IOT, would rest on an assumption about humans. The third form being a form of anthropological research in which hypotheses on moral opinions could be checked. However, neither anthropology nor linguistics, he argues, could tell us anything about any actual moral correctness.

If my previous interpretation of the IOT is correct, essentially we must understand ethical expressions with reference to an IO which we cannot claim to know for certain. We might hold our own ideas of the characteristics of the IO as provisionally true. But this would not define a proper IO, which is what I argue that the analysis must concern. This however, would not refute the theory in principle, but rather render it without any real ability to shed light on the meaning of ethical expressions.

One possible revision of the theory would be to abandon the absolutist criteria and instead attribute meaning to our non-self-contradictory speech practice. In such a case, meaning would rely entirely on our use of words, in turn reflecting our norms. This would, as I have previously argued, open up for relativism. And since we would no longer analyse meaning in virtue of a universal and ideal reference, with reactions aligning with an infallible moral judge, the very concept of referring to an IO would appear redundant. It might of course be the case that such norms would happen to satisfy a provisionally held absolutist theory. However, the theory as such would not need to depend on this as a principle. In the same way it might be that our human norms just happen to align with actual rightness or moral correctness, in line with the reactions of a possible infallible moral judge. But similarly, the semantic theory would not be dependent on this. The only semantic assumption we would need to defend, would be that of attributing meaning to our speech practice. Any notions as to the nature of universal idealised human norms or moral substance, could be attributed non-semantic theories.
4 Conclusion

I have argued that: (i) as a consequence of the IOT’s absolutist criteria, the ethically significant reaction of IOs cannot be conflicting, and that this means that the IOs characteristics must be fundamentally the same. (ii) that we cannot know for certain the characteristics of the IO, but that this is not in principle necessary for the form and validity of the theory. (iii) that such form presupposes actual characteristics of the IO based on an assumption about the human nature. (iv) that ‘IO’ designates a hypothetical reference through a circular definition, but that this need not in principle refute the theory. But that it however renders the theory with little or no substance, by which I intend a lack of real ability to clarify the meaning of ethical expressions. This dilemma, I suggest, could be addressed by abandoning the theory’s absolutist criteria.

In attempting to analyse the IOT, I have taken part of several of its interpretations and I have myself changed foot several times. And while the theory might escape much of its possible critique, I believe that it does so at the expense of substance. If for instance we believe the IO characteristics, as presented by Firth, as necessary for the IOT, we might view them as a progress of sort, helping us to better understand or analyse the meaning of ethical expressions. But if we, in defending the theory, refute their strict necessity and instead rely on the theory’s form, that same substance will escape us. If we hold that the theory is based on our practice of justification, that too could be suggested to help us better understand the nature of the moral language. But if we, in defence of the theory’s absolutism instead attribute the meaning of ethical expressions to the ideal practice of justification of the IO, that position too escapes us since our own practice of justification is not necessarily ideal at all. We might defend the theory, claiming that any such certain knowledge, or substance, be not necessary for the principle validity of a semantic theory. However, in such case it would appear that we refer to an IO we know nothing about for certain, save a circular designation. Without a revision or clarification regarding the absolutist characteristic, this would seem to leave us with an analysis that, although perhaps not flawed in principle, can provide no real meaningful explanation regarding the meaning of ethical expressions. And in the event that I am mistaken in some or all of my interpretations, and I do not think I am, I would hold that the IOT is in need of further clarification.
References


Hare, R.M., Moral Thinking, New York, Oxford University Press, 1981. E-book


Jollimore, Troy, "Impartiality", The Stanford Encyclopedia of Philosophy,


