CAN PROGRAM EXPLANATION CONFER ONTOLOGICAL RIGHTS FOR THE CORNELL REALIST VARIETY OF MORAL REALISM?

by ANDREW FIELD

A thesis submitted to
The University of Birmingham
For the degree of
MPHIL(b) OF PHILOSOPHY

School of Philosophy,
Theology and Religion
College of Arts and Law
The University of
Birmingham
October 2010

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Acknowledgements

I would like to acknowledge and thank the following persons who made the completion of the following this thesis possible. I am grateful and fortunate to have Professor Alex Miller as a supervisor, with his timeless patience of reading many inadequate drafts and offering suggestions for their improvement. Also, I appreciate his constant support and enthusiasm for this project , this became invaluable throughout, and without him who knows what mess this thesis would have become.

I'd like to extend my appreciation to my family for their continual encouragement, genuine interest and support for the thesis throughout and in every facet of my life. Completing this thesis is only possible with such loving and caring foundations, and this thesis is dedicated to them.

Lastly, I offer my gratitude to Anu for her advice throughout the thesis and making me forget possible worlds, counterfactual conditionals and program explanation. Finally, I thank Anu's family for helping me at one particular difficulty at the very end of writing the thesis.

Contents

Can Program Explanation Confer Ontological Rights for the Cornell Realist variety of Moral Realism?

Introduction

This thesis is primarily concerned with the question as to whether program explanation can confer ontological rights for the Cornell Realist variety of moral realism? The aim of this thesis is to argue and defend the claim that program explanation *cannot* confer ontological rights for moral facts and properties. For a bulk of the time in the thesis in Chapter 4, in order to answer the question above, I will discuss and examine the papers or book excerpts of Alex Miller (2003),(2009), Mark Nelson (2006) and Paul Bloomfield (2009). Specifically, in the thesis I end up arguing against Bloomfield and Nelson and defend the arguments of Alex Miller. My argument is that program explanation is not needed to explain why a certain counterfactual conditional, which we discuss in the thesis, is true because the context-sensitivity of counterfactuals undermines the need for program explanation to explain this. Moreover, Bloomfield's argument against the fact that the context-sensitivity of counterfactuals undermines the need for program explanation to explain why the counterfactual conditional is true, does not work because the epistemically unlimited agent vis-à-vis the lower level properties and the process explanations in which they figure already knows all the contextual features needed for the context-sensitivity of counterfactuals to undermine Bloomfield and Nelson's arguments.

In Chapter 1, I present an historical overview of metaethics and examine the main metaethical positions and objections. Towards the end of the chapter I claim that Cornell Realism is not prone to the objections just discussed, and Cornell Realism

comes out best. Chapter 2 examines Cornell Realism at length and focuses on the 1980's debate between Gilbert Harman and Nicholas Sturgeon, towards the end of the chapter I argue that Harman wins the debate between the two and that Sturgeon needs an alternative reply to Harman other than the one that he gives. Chapter 3 is devoted to Sturgeon's reply to Harman that comes in the form of a program explanation; I start off by discussing what program explanation is, what its benefits are and finally how Sturgeon can use it to reply to Harman's objection[1]. Chapter 4 discusses the responses as to whether Sturgeon's use of program explanation is successful in its reply to Harman, and in doing so, confers ontological rights for moral facts and properties. In Chapter 5, I develop an objection to Bloomfield and finish the thesis with a conclusion.

---

[1] Though it should be noted that this is not a reply that Sturgeon himself makes explicitly: it is investigated on his behalf in Miller (2003).

Chapter 1

This chapter will present a highly selective overview of Meta-ethics in historical

order. The aim of this chapter is to move from the Open Question Argument to

Cornell Realism and to make plain what justifies interest in Cornell Realism through

the failings of other meta-ethical theories. Hopefully, it will also trace the movement

of meta-ethical theory through the failings of previous meta-ethical theories. So

throughout this chapter I will make clear the position of major meta-ethical theories

and the influential objections against each theory.

The meta-ethical theories and major objections that I discuss are: the Open

Question Argument, Non-Naturalism, The Frege-Geach objection, Quasi-Realism,

Error Theory, and then finally in the next chapter, Cornell Realism.

Cognitivism and Non- Cognitivism

Consider a typical moral judgement, such as the moral judgement that murder

is wrong. *Cognitivists* think that a moral judgement expresses a belief where beliefs

can either be true or false. For instance, take the belief that the tea is hot – this belief

is true if and only if the tea is hot. *Non-Cognitivists*, on the other hand, think that a

moral judgement expresses something like a desire or emotion. Desires and emotions

cannot be either true or false. So according to the non-cognitivist, a moral judgement

such as the judgement that murder is wrong is neither true nor false.

The Open Question Argument (OQA) and Moral Non-Naturalism

Meta-ethics begins with G.E. Moore. Moore in his *Principia Ethica* [1903] (1993) produces the OQA in which he argues that "good" cannot be defined in terms of natural properties, such as, pleasure or desire. By natural properties I mean properties that are either causal or detectable by the senses. By this characterization, natural properties are the properties dealt with in the natural sciences or psychology. This is called *definitional naturalism*: moral properties, by definition, are identical or can be reduced in terms of naturalistic properties. To try and define moral properties in terms of naturalistic properties commits, what Moore calls, the "naturalistic fallacy", the fallacy of trying to define "good" in naturalistic terms.

The OQA is as follows. Lets assume that the predicate "good" is synonymous (or analytically equivalent) to a naturalistic predicate N, say, pleasurable. So it is part of the meaning of "x is pleasurable" that "x is good." So if someone asked if that which is pleasurable is also good? they would betray conceptual confusion; they would appear not to understand the concept that the predicate "pleasurable" and the predicate "good" are synonymous (or analytically equivalent). Just in the same way, if someone asked is an "unmarried man" a "bachelor"? we would think that they didn't understand the concept that a "bachelor" is synonymous (or analytically equivalent) to "unmarried man." Now this next step is why the argument is called the "open question argument." However, it is always a *significant* question whether something that is pleasurable is good (*unlike* whether asking someone who is unmarried is a bachelor), thus someone asking this question actually betrays no conceptual confusion. So it is not true that the predicate "pleasurable" is synonymous or analytically equivalent to the predicate "good."

The OQA though is not as strong as it seems, at least in its original formulation. The main objection and problem of the OQA was offered by William Frankena (1938), as Frankena says, quite concisely:

"[T]he charge of committing the naturalistic fallacy can be made, if it all, only as a conclusion from the discussion and not as an instrument of deciding it." (Frankena, 1938, 465).

Basically, Frankena's objection is that the OQA begs the question against the moral definitional naturalist. Moore can only appeal to the OQ if the conviction that the question "is an x which is N, good?" is open is well substantiated. The problem is that the OQ is only well substantiated if it is the case that moral and naturalistic predicates are not synonymous or analytically equivalent. However, the definitional naturalist will say that that conviction that the question is open is *not* well substantiated. So, according to the moral definitional naturalist, the question "is an x which is N, good?" is a closed question, and Moore's argument against naturalism cannot run.

Since it roughly took 35 years – Moore's OQA in the *Principia Ethica* [1903] and Frankena's objection (1938) – for a convincing response against the OQA on behalf of the moral definitional naturalist, as this chapter is working historically through the metaethical views and movements, the OQA will be assumed successful against the moral definitional naturalist for our purposes[2].

 Moore's favoured brand of cognitivism was moral non-naturalism. Non-naturalism argues that moral properties are non-natural, irreducible and sui generis[3].

---

[2] The two main contemporary versions of moral reductionism are: Analytic Moral Functionalism and its leading proponents are Frank Jackson and Philip Pettit (1995), and Peter Railton's Moral Reductionism (1986a), (1986b). I will not have space to discuss these here.

[3] Contemporary versions of non-naturalism are advocated by David Wiggins (1987), John McDowell (1988) and more recently Russ Shafer Landau (2003). Again, I cannot discuss these here.

Moore's non-naturalism claims that moral judgements are truth apt: moral judgements are either true or false. According to Moore what makes a moral judgement true is the instantiation of a non-natural, unanalysable and simple property of moral goodness. Moore thinks, then, that we access this non-natural, unanalysable, simple property via the operation of a special faculty of "intuition" (hence Moore's position is sometimes called "intuitionism"). So take the moral judgement that giving to the poor is good, it is true because of the instantiation of a simple, sui generis property moral goodness by acts of giving to the poor.

The main problem for non-naturalism that makes it an unattractive position is that it has no plausible epistemological account as to how we cognitively access moral properties. If by "intuition", as we will assume, Moore means a cognitive faculty that works as a kind of perception, similar to sense-perception, then such a view makes no sense. It amounts to a form of perception that apparently is not, ex hypothesi, causal and this is difficult to understand. Moreover, non-naturalism is viewed within widespread suspicion amongst philosophers and later in this chapter in the section "error theory" we will see that John Mackie expounds on the same general problem of intuitionism.

Assuming for the time being the success of the OQA and the standard rejection of non-naturalism the prospects of cognitivism look bleak; moral properties are neither identical nor reducible to naturalistic properties or non-natural properties themselves. For these various reasons many philosophers turned towards trying to find a viable non-cognitivist theory. It is now that we look at Quasi-Realism.

Quasi-Realism (QR)

Before the in-depth discussion of Quasi Realism, I will focus briefly on Emotivism and its main objection. The main objection is the Frege-Geach objection which threatens not only Emotivism but all non-cognitivist theories. I am dealing with Emotivism first because Quasi Realism was set up to try and solve the objections that defeated Emotivism, and Emotivism and Quasi Realism both share the same underlying expressive semantics.

A.J. Ayer's Emotivism (1936) denies that moral judgements express propositions. According to Emotivism, in the standard moral judgement that murder is wrong, I am not saying that murder is wrong or anything else, but rather I am expressing my disapproval. Just as I would cry !%*$ in anguish when I burn my hand: crying !%*$ does not express the proposition that I am in pain, it is an expression of anguish.

The central objection to emotivism and non-cognitivism is the Frege-Geach objection (Geach 1960; 1965). The problem for the non-cognitivist is how to account for the appearance of a moral judgement in an unasserted context in a way that does not render intuitively valid arguments guilty of a fallacy of equivocation. Take an intuitively valid moral modus ponens argument. For instance,

(1) Murder is wrong

(2) If murder is wrong, then getting your little brother to murder is wrong

(3) Therefore, getting your little brother to murder is wrong.

In order for the argument to be valid the occurrence of "murder is wrong" in (1) has to have the same semantic function as "murder is wrong" occurs in (2). The argument is *not* valid if semantic function in (1) is different than how it occurs in (2). This is the equivocation fallacy.

The equivocation fallacy is committed when the same word or phrase is used in an inference but each token of the word or phrase has a different meaning. For instance,

(4) My clock has hands

(5) If something has a hand, then it must have fingers and a thumb

(6) Therefore, my clock has fingers and a thumb.

As you can see here the semantic function of the word "hands" within (4) is different from how it occurs within (5). In (4) "hands" means (usually) black instruments that point to numbers on a clock, but in (5) "hands" means a part of the human anatomy.

Now the problem for the emotivist and non-cognitivism is that it renders intuitively valid arguments guilty of equivocation. To see this, take the valid moral modus ponens argument but with an underlying expressive semantics:

(1) Murder is wrong

(2) If murder is wrong, then getting your little brother to murder is wrong

(3) Therefore, getting your little brother to murder is wrong.

The semantic function of "murder is wrong" as it occurs within an *asserted* context in (1) is different from its semantic function as it occurs within an *unasserted* context in

(2). This commits the fallacy of equivocation and the argument is not valid unless the semantic function is the same as it occurs within (1) and (2). In (1) "murder is wrong" is asserted to express disapproval of murder, but in (2) "if murder is wrong, then getting your little brother to murder is wrong," "murder is wrong" is being used as an antecedent of the conditional. It is *not* being used to express disapproval of murder. The argument thus appears to turn out invalid.

Quasi-Realism is developed in order to defuse this worry**.** What is QR? As Simon Blackburn puts it:

"Quasi-Realism is the enterprise of explaining why our discourse has the shape it does, in particular by way of treating evaluative predicates like others, if projectivism is true. It thus seeks to explain, and justify, the realistic seeming nature of our talk of evaluations – the way we think we think we can be wrong about them, that there is a truth to be found, and so on." (Blackburn, 1984, 180).

It is useful to note that there are two parts to the QR enterprise: *modest QR* and *ambitious QR*. Blackburn's above quote concerns modest QR. Modest QR attempts to explain why we talk as if it is true (or at least truth-apt) that, for example, murder is wrong and giving money to the poor is good even though it is assumed that such moral predicates do not refer to actual moral properties. That is, even though we judge and express the belief that murder is wrong, moral predicates do not refer to actual moral properties, but rather serve to express feelings and sentiments. So for instance, when we say that "murder is wrong" we are talking as if we are expressing the proposition that murder is wrong, even though we are expressing no proposition at all but actually an expression of disapproval.

How then does such a position work? Lets take a closer look at modest QR. The QR believes that there is a distinction between the *surface form* of a region of discourse and the *deep form* of a region of discourse. The surface form of a discourse is usually suggested by the syntax. The surface form of a moral discourse is *cognitive* or *propositional*: "murder is wrong" and "giving money to the poor is right", for example are *declarative sentences*, suggesting that moral sentences represent states of affairs, and the *predicates* "wrong," "good" and "permissible", for example suggest that predicates denote properties. Furthermore, "Mark believes that murder is wrong" and "Patricia believes that giving money to the poor is good" are syntactically well-formed, suggesting that moral judgements express beliefs. However, the projectivist wants to deny that the surface form of a moral discourse is an accurate guide to its deep form, although moral statements appear propositional or cognitive on the surface, their essential role is expressive. (Miller, 2003, 60)

Ambitious QR, on the other hand, argues that moral judgements, such as, the judgement that murder is wrong or the judgement that giving money to the poor is right, have truth conditions, but this is done solely on a projectivist basis. So it is either true or false that murder is wrong or giving money to the poor is right, but a moral truth is constructed by our expressions of attitudes and moral sensibilities. The shift from modest QR to ambitious QR is the shift from talking "as if" there is moral truth when really there is not, to allowing that there is such a thing as moral truth but viewing it as constructed solely on a projectivist basis[4].

Since QR was designed to meet the problems that emotivism faced, how then does Blackburn try and find a solution on behalf of the QR against the Frege-Geach

---

[4] Since I'm concentrating on modest QR due to its application to the Frege-Geach problem, the discussion of ambitous QR is far more brief and is intended merely to show what the other part of the QR enterprise is.

problem**?** In what follows I borrow heavily from Miller's (2003) exposition on Blackburn's attempted solution to the Frege-Geach problem.

Blackburn's solution consists of three sections. Firstly, consider a simpler connective than "if … then", such as "and". A conjunction is true only when both of its conjuncts are true, and false otherwise. We use "and" to join commitments: for instance "murder is wrong and giving money to the poor is good." What can a projectivist say about such a conjunction since he does not want to assess the conjunction in terms of truth and falsity? Blackburn answers:

"[we should] expand the way we think of "and". We have to do this anyway, for it can link utterances when they certainly do not express beliefs which are genuinely susceptible of truth-value e.g. commands: "hump that barge and tote that bale." We would instead say something like this: "and" links commitments to give an overall commitment which is accepted only if each component is accepted." (Blackburn, 1984, 191-192)

So a projectivist account of "murder is wrong and giving money to the poor is good", is as follows: this conjunctive sentence serves to express my disapproval of murder and my approval of giving money to the poor.

Before I move on, secondly, I will quickly explain the notion of a moral sensibility. A moral sensibility is a disposition to respond to various situations with different attitudes. For instance, we may be pleased by selflessness, inspired by ambition, repulsed by megalomania and indolent about dictatorships. The complete set of dispositions is a person's moral sensibility. We can also take attitudes towards a person's moral sensibility: for example, we can regard it as sensitive or insensitive, fickle or stubborn and admirable or abhorrent. (McNaughton, 1988, 183)

Projectivists can do something similar for conditionals, such as, "if murder is wrong, then getting your little brother to murder is wrong." In uttering a conditional like the one we are concerned with you are expressing an attitude towards a moral sensibility. So in the conditional "if murder is wrong, then getting your little brother to murder is wrong" I am expressing my attitude of approval towards moral sensibilities which combine disapproval of murder with disapproval of getting little brother to murder people.

Finally, in order to see how Blackburn's attempted solution works consider a language in which its surface form is expressive (call such a language Eex) instead of its surface being propositional.

"It might contain a "hooray!" operator and a "boo!" operator (H!, B!) which attach to descriptions of things to result in expressions of attitude. H!(the playing of Tottenham Hotspur) would express the attitude towards the playing, B!(lying) would express the contrary attitude towards lying, and so on". (Blackburn, 1984, 193)

In order to talk about an attitude of approval or disapproval we use square brackets. So [H!(the drinking of a cup of tea)] refers to the sentiment of approval of drinking a cup of tea. Moreover, to indicate two attitudes together we place a semi-colon between them: so [[H!(drinking of a cup of tea]); [B!(eating crisps)]] refers to the joining of attitude of approval towards the drinking of cups of tea *with* the attitude of disapproval towards the eating of crisps.

Now what would the conditional (2) in the original argument look like in Eex? The projectivist interprets (2) in terms of the expression of approval of moral sensibilities which combine disapproval of murder with disapproval of getting little brother to murder. In Eex (2) would be represented as: H![[B!(murder);[B!(getting

little brother to murder)]]. So the argument of (1), (2) and (3) comes out as:

(1ex) B!(murder)

(2ex) H![[B!(murder);[B!(getting little brother to murder)]].

- Therefore –

(3ex) B!(getting little brother to murder)

The argument then appears to be valid. Somebody who was committed to the premises but who rejected the conclusion would have, as Blackburn puts it, a "clash of attitudes". He would fail to have the combination of disapproval towards murder and disapproval towards getting little brother to murder, whilst approving of that combination.

However, Blackburn's attempted solution is not successful. Crispin Wright raises the following worry:

"Anything worth calling the validity of an inference has to reside in the inconsistency of accepting its premises but denying its conclusion. Blackburn does indeed speak of the "clash of attitudes" involved in endorsing the premises of the modus ponens example, construed as he construes it, but in failing to endorse the conclusion. But nothing worth regarding as inconsistency seems to be involved. Those who do that merely fail to have every combination of attitudes of which they themselves approve. That is a *moral* failing, not a logical one. (Wright, 1988, 33 [original italics])

In essence Wright's problem is that the "clash of attitudes" shows a moral failing not a logical failing. And for the validity of the inference to be accepted, being committed to the premises and not the conclusion must be a logical failing.

Despite Blackburn's attempts the Frege-Geach problem is a strong and powerful objection against QR and non-cognitivist theories. There has yet to be a convincing solution to the problem and the worry that non-cognitivism cannot account for the validity of a valid modus ponens argument still persists; if we cannot find a convincing solution to the Frege-Geach objection, then all the worse for QR and other non-cognitivist positions[5].

Error theories are supposed to be an alternative to both Moorean non-naturalism and non-cognitivism so I am now going to move onto John Mackie's Error-Theory.

Error Theory

John Mackie's *Error-Theory* (1977) argues that moral judgements express beliefs, but that atomic positive moral judgements are systematically and uniformly false. So, for instance, when we assert the moral judgement that murder is wrong this moral judgement comes out as false: it is not true that murder is wrong.

How then does Mackie argue for such a radical claim? By the conjunction of a conceptual claim with an ontological claim. I'll discuss the former in detail first and then move onto the latter. Mackie's conceptual claim is that our concept of a moral requirement is a concept of an *objectively categorical prescriptive* fact. To say that moral requirements are *prescriptive* is to say that it tells us how we ought to act, or

---

[5] For instance, Allan Gibbard's *Norm Expressivism* (1990) which is the theory that moral judgements express our acceptance of norms.

that it gives us reasons to act. For instance, if something is morally good, the prescriptive requirement of moral properties entails, it is something that we ought to act towards, or have reasons to act towards. And conversely, if something is morally bad it is something that we ought *not* to act towards, or have reasons *not* to act towards. Moreover, to say that moral judgements are *categorical* is to say that the reasons for actions are not dependent on our desires, wills or inclinations. If something is morally good, then irrespective of our desires or inclinations we ought to act towards it. Or if something is morally bad, no matter our desires, wills, inclinations, then we ought *not* to act towards it. "*Objective*" here means - for simplicity's sake for nothing will turn on it in this discussion, as Mackie says a lot of different things – is a property that is independent of our choices and preferences, and it is part of the fabric of the world.

The ontological claim is simply that there are no categorical objectively prescriptive facts in the world; there are no moral facts in the world. Both of these claims together (the conceptual and ontological) entail a cognitivist error theory view of morals[6].

Mackie's main argument for the ontological claim is the argument from queerness. Mackie has two reasons as to why there exist no moral properties, which must fit the objective and categorically prescriptive requirement: metaphysical and epistemological. I will look at each one in turn respectively.

Firstly, the metaphysical argument: if such moral properties were to exist they would be metaphysically queer, moral properties would be utterly and completely different to the properties that we are familiar with in the universe as described in science and common-sense. As Mackie explains:

---

[6] Which, obviously, due to the ontological claim is not a version of moral realism.

"An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. Similarly, if there were objective principles of right and wrong, any wrong (possible) course of action would have not-to-be-doneness somehow built into it." (Mackie, 1977, 40)

And, according to Mackie, no instantiation of moral state of affairs exist in the world. Secondly, the epistemological argument (which expounds on the problem of "intuitionism"), backs up the metaphysical argument:

" If we were aware [of objective values], it would have to be by some special faculty of moral perception or intuition, utterly different from our ways of knowing everything else. These points were recognised by Moore when he spoke of non-natural qualities, and by the intuitionists in their talk about a "faculty of moral intuition." Intuitionism has long been out of favour, and is indeed easy to point out its implausibilities. What is not so often stressed, but is more important, is that the central thesis of intuitionism is one to which any objectivist view of values is in the end committed: intuitionism merely makes unpalatably plain what other forms of objectivism wrap up." (Mackie, 1977, 38)

The only way in which we can come into contact into such metaphysically queer moral state of affairs is to include unexplanatory and unpalatable conceptions of how we come into cognitive content with such moral state of affairs, such as that exhibited by moral intuitionism

       Wright's objection is that error theory's story about the point of moral discourse is inherently unstable given its moral scepticism about moral truth. Wright

and Mackie both realize that the error-theorist need a subsidiary norm which statements of moral discourse aim at and can satisfy besides moral truth. (Wright, 1996, 2) The point of moral discourse, according to Mackie, is to secure the benefits of social co-operation[7]. (Mackie, 1977, chapter 5) Suppose there's a plausible account of the subsidiary norm taken from the story of moral discourse – to secure benefits of social co-operation – that governs the practice of forming moral judgements, then moral judgements will be aimed at satisfying this subsidiary norm, whatever it is. Not all moral judgements will satisfy the subsidiary norm in equal manner, for example, the moral judgement "murder is good" will frustrate the subsidiary norm, whilst the moral judgement "murder is bad" will facilitate the subsidiary norm. The problem that Wright raises for Mackie's error theory concerns whether it can plausibly combine a story about the benefits of practicing moral judgements with the central negative claim of the error theory. Wright thinks not:

" [I]f, among the welter of falsehoods which we enunciate in moral discourse, there is a good distinction to be drawn between those which are acceptable in the light of some subsidiary norm and those which are not – a distinction which actually informs ordinary discussion and criticism of moral claims – then why insist on construing *truth* for moral discourse in terms which motivate a charge of global error, rather than explicate it in terms of the satisfaction of the putative subsidiary norm, whatever it is. The question may have a good answer. The error-theorist may be able to argue that the superstition that he finds in ordinary moral thought goes too deep to permit any construction of moral truth which avoids it to be acceptable as an account of *moral* truth. But I do not know of promising argument in that direction." (Wright, 1996, 3 [original italics])

---

[7] This is a simplification, but the details need not concern us here, nothing turns on it.

Wright's problem for Mackie is that if there is such a thing as a subsidiary norm, then why not construct moral truth in terms of the satisfaction of the subsidiary norm, instead of constructing truth for moral discourse in terms of that motivates a global error-theory? According to Wright there is no plausible reason or good argument to do so, in which case, Mackie might as we construe moral truth in terms of the satisfaction of the subsidiary norm which will enable him to bypass the global error-theory of the practice of moral judgements.

Another problem for error-theory is that opposing metaethical views can simply reject the metaphysical aspect of Mackie's argument, denying his formulation of moral properties is the correct version. To reject the metaphysical aspect of Mackie's argument from queerness, we need to find a plausible account of moral properties that does not see moral properties as being categorical objective prescriptive requirements[8] [9].

Cornell Realism

Finally we come to Cornell Realism. In the next chapter I will discuss Cornell Realism more fully, but for now let's see how Cornell Realism manoeuvres past the objections and problems that beset the other metaethical theories.

Cornell Realism is a cognitivist theory that argues that moral facts and properties are irreducible (non-reductive) sui generis natural properties. By claiming that moral properties are irreducible properties this enables Cornell Realism to

---

[8] One option is Crispin Wright's Judgement Dependent Account (1988; 1989; 1992) which argues "that our best judgement about morals determine the extension of moral predicates" (Miller, 2003, 6) This is a dispositionalist account of moral values.

[9] Throughout this paper I am quiet on whether certain meta-ethical views can offer a plausible epistemological account of morals to diffuse the epistemological aspect of Mackie's argument from queerness.

sidestep the OQA and the naturalistic fallacy because Cornell Realism does not claim

that moral predicates are analytically equivalent or synonymous to other natural

predicates. Also by claiming that moral properties are natural properties the Cornell

Realist avoids being caught up in moral non-naturalism and the bizarre faculty of

"intuitionism" in cognitively accessing moral facts or properties. The fact that Cornell

Realism is a cognitivist theory means obviously that it does not face the same problem

as the non-cognitivists. The Cornell Realist escapes the Frege-Geach problem because

they can explain the semantic function of "murder is wrong" in an unasserted context

without making intuitively valid argument guilty of equivocation. Finally, Cornell

Realism does not fall into a global error theory since it does not accuse moral

judgements of being systematically and uniformly false. Cornell Realism is a type of

moral realism in that it allows that moral judgements are objectively true. However, is

Cornell Realism as attractive as it initially appears? The rest of this paper is devoted

to looking at one recently discussed aspect of this question.

Chapter 2

In this chapter we turn to Cornell Realism and discuss whether it is a viable meta-ethical theory. Starting off by discussing the Cornell Realist position in depth, I move onto an influential objection by Gilbert Harman against the Cornell Realist. I will then give an overview of the debate discussing Nicholas Sturgeon's replies and counter replies[10] to Harman's objections and counter-objections. This chapter will provide background for the next chapter when I discuss a recent argument in response to Harman's objection, which is the central topic of this paper

Detailed account of Cornell Realism

According to Cornell Realism moral properties are natural but irreducible properties. Moral properties supervene on other natural (non-moral) properties, but moral properties themselves are irreducible and cannot be reduced to a particular or complex set of non-moral natural properties. Firstly, in this discussion of Cornell Realism, I will solely discuss what it means for moral properties to supervene on natural properties. Cornell Realists prefer a strong supervenience claim, which is as Jaegwon Kim explains:

"A strongly supervenes on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and necessarily if any y has G, it has F." (Kim, 1984, 165)

---

[10] Throughout this chapter we mainly concerned with Sturgeon's 1980's presentations.

To say that moral properties strongly supervene on non-moral natural properties is so to say that if an object has a moral property, such as being good, then there is some non-moral natural property, say being N, such that, necessarily, if some object has N, then that object is good. For example, suppose that giving aid to the hungry is good. Then according to strong supervenience, acts of that type have some non-moral natural property N such that necessarily, any act which has N is also good. Moreover, this means that in any possible world, suppose someone gives aid to the poor, then acts of this type have the property N, and if some object has N then that object is good.

Strong supervenience is usually distinguished from weak supervenience: to say that moral properties *weakly* supervene on non-moral natural properties is only to say that within any particular possible world if two objects have the same non-moral properties, then they have the same moral properties. For example, suppose that two actions are instances of giving aid to the hungry and are identical in all other natural respects. Then, within the possible world in question, either both are good or neither are.

Strong supervenience is consistent with multiple realization: a good action A might be good in virtue of having N_1, while another good action B might be good in virtue of having N_2. Note that multiple realization is usually held to rule out reduction.

Later in the paper the terms "higher-level property" and "lower-level property" will show up. I'm not going to give a precise definition of what higher and lower level properties are since the best way to understand them is through illustration.

In any supervenience relation between the base properties and the supervening property, the base properties are the lower-level properties and the supervening

property is the higher-level property. For instance, the moral property supervening onto the non-moral base properties is the higher-level property and the non-moral natural properties are the lower-level properties. To take another example, chemical properties are widely held to supervene onto microphysical properties, in this case the supervening property, the higher-level property, is the chemical properties and the base properties, the lower-level properties, are the microphysical properties. And as a last example, to show, for instance, that the same property will not always be the higher-level supervening property, take the supervenience relationship between biological and chemical properties. The biological properties supervene onto the chemical base properties, where the biological properties, in this case, are the higher-level property and, in this case, the chemical properties are the lower-level properties.

Now I will approach the second part of the Cornell Realist argument: that moral properties are *irreducible* (or non-reductive) to a particular or complex set of non-moral natural properties.

Cornell Realists hold the view that moral properties can be multiple realized by other natural properties. Contrary to moral reductionists, Cornell Realists hold that moral properties are not reducible to other natural properties. For instance, to take an archaic reductionists position, Cornell Realists deny that "good" is analytically equivalent or synonymous to "pleasure".

So what does multiple realizability mean? To help explain this I will first illustrate the famous example of multiple realizability in philosophy of mind. Take a mental type, pain (for example): many mind-brain identity theorists held that pain was reducible to c-fibre stimulation. Just as, analogously, moral reductionists hold that a moral kind was reducible to a particular or complex non-moral natural kinds.

But as Hilary Putnam (1967) pointed out: many life-forms – terrestrial, extraterrestrial or robotic – could feel (or eventually could be known to feel) pain. Now, for this to be compatible with the mind-brain identity theorist, then they would have to argue that the particular or complex physical kind was common to all the life-forms that could feel pain. This seems absurd because the mental kind, pain, could plausibly be realized by many *distinct* neurophysiological types, such as, green slime in aliens, silicon in cyborgs, or an electronical state in a supercomputer. Multiple realizability about the mental, then, is that given a mental kind, pain, this can be realized by many distinct neurophysiological types: c-fibre stimulation in humans, green slime in aliens and silicon in cyborgs etc. So, it is not just organisms that have c-fibre stimulation that can feel pain, being in pain is a state which is multiply realizable by neurophysiological types and cannot be reduced to a specific neurophysiological type. Regardless of the philosophy of mind example above, multiple realizability is commonly held to block reduction separately[11]

Analogously Cornell Realists block reduction because they argue moral kinds, wrongness (for example), can be multiply realized by distinct non-moral natural kinds. For example wrongness can be realized by unprovoked physical violence, stealing money, unjustifiably mocking someone, deceiving and lying etc, but in each of these examples, wrongness, cannot be reduced to a particular or complex non-moral natural kind. This is still consistent with strong supervenience, as Cornell Realists are prepared to say that necessarily, anything else which is an instance of unprovoked physical violence, for example, is wrong.

---

[11] Not all philosophers accept this view, for example, a notable and influential philosopher, such as, Jaegwon Kim. Kim, briefly, believes that all multiple realizablity does is make reduction "messy" (cf. Jaegwon Kim (1989) "The myth of nonreductive materialism.")

How, then, do the Cornell Realists justify the existence of irreducible natural moral properties:

"[The Cornell Realists] have pursued analogies with natural and social science to argue that moral properties might be both irreducible and explanatorily efficacious. One might, for example, argue that various chemical or biological "natural kinds" – acid, catalyst, gene, organism – are not obviously type reducible to the natural kinds of physics, and yet play a good role in good scientific explanation." (Darwall, Gibbard and Railton, 1992, 169-170)

So, as we can see from the above quote, analogously the Cornell Realists argue that although moral properties are not type reducible to other natural properties, moral properties still play a distinctive role in the best moral explanations.

So the Cornell Realist, then, argues as follows:

(1) "P is a real property if and only if P figures ineliminably in the best explanation of experience.

(2) Moral properties figure ineliminably in the best explanation of experience. Therefore:

(3) Moral properties are real properties." (Miller, 2003, 140-141)

The Cornell Realist argues that moral properties and facts can justifiably be included in our ontology of the natural world in the same way that physical, chemical, biological, psychological and sociological properties and facts are, in that they pull their weight in our best overall explanatory picture of the world. Just as we can explain why a physicist forms the belief that there is a proton in the cloud chamber in

virtue of their being a proton in the cloud chamber, moral properties can contribute to the explanation of moral belief.

Harman's objection to Cornell Realism

Gilbert Harman (1977) argues that whilst physical properties figure ineliminably in the best explanation of experience, as demonstrated by the case of the physicist's belief in the proton, he denies that moral properties figure ineliminably in the best explanation of experience. Harman's motivation for the argument is to show how moral explanations are problematic in a way that scientific explanations are not.

To see why consider this example. Godzilla kidnaps a young maiden, takes the maiden on top of the empire state building (the maiden is scared of heights) and eats her, and Anu in seeing this immediately forms the belief that what Godzilla is doing is wrong. But do we need to cite wrongness as explanatorily relevant as to why Anu thinks what Godzilla is doing is wrong? According to Harman, we do not: we only need to cite the natural non-moral facts, such as, the unwanted kidnapping, the extreme pain, the suffering the maiden experiences, and facts about our psychology (specifically our character and upbringing).[12]

The difference between the moral and the physical case is as follows. In the physical case we have to assume the existence of the proton to explain why the physicist forms the belief that a proton is in the cloud chamber. In the moral case, however, we do not need to assume the existence of wrongness to explain why Anu forms the belief that Godzilla's action of kidnapping and eating the maiden is wrong.

---

[12] In other works, Harman (1999; 2000) denies that there is such a thing as character! This is an odd fact since he believes, above, that you can explain why something is wrong, in part, due to our character.

As Harman explains:

"You need to make assumptions about certain physical facts to explain the occurrence of the observations that support a scientific theory, but you do not seem to need to make assumptions about any moral facts to explain the occurrence of the so-called moral observations I have been talking about. In the moral case, it would seem that you need only make assumptions about the psychology or moral sensibility of the person making the moral observation[13]." (Harman, 1977, 6)

So according to Harman, we can deny premise (2) of the Cornell Realist argument by arguing that: (2a) Moral "properties" do not figure ineliminably in the best explanation of experience. By dropping reference to moral facts in the explanation of moral beliefs we suffer *no* explanatory loss, the unavailability of moral facts does *not* entail explanatory impoverishment. The use of moral facts in explanation does not, in the relevant sense, count as best, and so we do not earn ontological rights for moral "properties."

Sturgeon's reply to Harman

Nicholas Sturgeon (1986) argues, in defence of Cornell Realism[14], that wrongness is explanatorily relevant in the forming of Anu's belief that what Godzilla was doing is wrong. Sturgeon's reply is that if wrongness passes the *counterfactual test*, a way to test claims about explanatory relevance, then wrongness *is* explanatorily relevant to Anu forming her belief

---

[13] "moral observations" here are really just moral beliefs
[14] Other leading proponents of the Cornell Realist position that I will not discuss in this paper are Richard Boyd (1988) and David Brink (1989)

The counterfactual test is as follows. If (if a had not been F, then b would not have been G) then a's being F is explanatorily relevant to b's being G. To illustrate this, consider this example: Godzilla's ferocious roar is explanatorily relevant to the town's citizens screaming in terror because if Godzilla had not ferociously roared, the town's citizens would not have screamed in terror.

So the pertinent question now is: - when we apply the counterfactual test to Godzilla and the maiden – if Godzilla had not done something wrong, then Anu would not have believed that Godzilla did something wrong? Sturgeon points out firstly that when reading this counterfactual conditional that we have a normative moral theory that conveys the relationship between non-moral facts and moral facts. According to this normative moral theory, an act that possesses the non-moral properties of unwarranted kidnapping and suffering needless extreme pain, or some other non-moral properties upon which moral properties supervene, also possess the moral property of wrongness. Suppose this normative moral theory is correct, then for the action not to be wrong, it would have to lack the non-moral properties of needless extreme pain and unwarranted kidnapping, or some other non-moral properties upon which moral properties supervene according to our correct normative moral theory. Now it is incorrect to assert that Anu would have still thought that Godzilla's actions were wrong even though the act lacked the non-moral properties of needless extreme pain and unwarranted kidnapping and so on. So, holding the assumption that our normative moral theory is correct, due to the counterfactual test, it seems true that the wrongness of the action is explanatorily relevant to Anu's belief that Godzilla's actions were wrong.

Sturgeon is well aware that moral sceptics will deny that he is entitled to assume a correct (or roughly correct) normative moral theory, but according to

Sturgeon this move of denying the normative moral theory does not help Harman. For, if Harman denies the assumption that the normative moral theory is correct, then we can derive a sceptical conclusion about the explanatory relevance about *physical properties*. Consider the case of the physicist and the proton in the cloud chamber. The physicist has a theory about what phenomena signal the existence and presence of the proton, namely, the vapour trail in the cloud chamber. Suppose this theory is correct, if there hadn't been a proton present there wouldn't have been a vapour trail, the physicist would not have formed the belief that a proton was present. So by this counterfactual test, assuming the truth of the physical theory, the proton is relevant to the physicist's belief that a proton was present. This is exactly the same as the moral case and the normative moral theory.

Now suppose that the physicist's theory is *incorrect*, that it is false that a vapour trail signifies the presence of a proton. That is, even without the presence of a proton there still would have been a vapour trail and due to the physicist's belief in a mistaken physical theory the physicist would have still believed that a proton was present even though it was not. So, on the assumption that the physicist's theory is incorrect we can derive the conclusion that if a vapour trail is present, then the physicist would believe that a proton was present when in fact it is not. If we assume that the physicist's theory is mistaken, then the presence of the proton is explanatorily irrelevant, by the counterfactual test, to the explanation of the physicist's belief that a proton is present[15].

---

[15] If some cannot get their head around the fact that a modern physical theory could be mistaken**,** there are two additional points that can be made. Firstly, an example of a now defunct scientific theory can illustrate this, such as, phlogiston. Not going comprehensively into phlogiston theory, scientists saw certain phenomena as signals that phlogiston was present. But since phlogiston theory is infact incorrect and phlogiston does not exist, the scientists who believed in phlogiston theory would still have believed that phlogiston was present due to certain signalling phenomena, when in fact it was not. Secondly, scientific hypotheses' are continually being up-dated and shown to be false, before a new scientific theory and hypothesis takes its place.

Harman has the option of "biting the bullet" and accepting that science and morality are on a par. So by "biting the bullet" Harman would be showing that moral and scientific theory are problematic in the same way, i.e. that they can be used to derive sceptical conclusions. However, Harman's motivation for his argument was to show that moral theory is problematic in ways that scientific theory is not. So instead of "biting the bullet" Harman needs a counter reply.

Harman's counter reply

Harman (1986) denies Sturgeon's claim that wrongness is explanatorily relevant to the formation of Anu's belief that Godzilla's action of kidnapping and eating the maiden is wrong. The truth of the counterfactual conditional that appears in the counterfactual test is *not* sufficient to yield genuine explanatory relevance. As Harman says:

"What is needed is some account of how the actual wrongness of [Godzilla's] action could help to explain [Anu's] disapproval of it. And we have to be able to believe in this account. We cannot just make something up, saying, for example, that the wrongness of the act affects the quality of the light reflected into [Anu's] eyes, causing her to react negatively. That would be an example of wrongness manifesting itself in the world in a way that could serve as evidence for and against moral claims, but it is not something we can believe in." (Harman, 1986, 63)

Harman's thought is that the counterfactual test is not sufficient for genuine explanatory relevance because for wrongness to be explanatorily relevant what we need is an account of the mechanisms that underlie such counterfactual dependence.

That is, we need to know what makes moral properties causally efficacious in the production of our experience and it cannot be the idea that "wrongness of the act affect the quality of light reflected into Anu's eyes" as Harman, above suggested. Or that moral protons, "morons" as Ronald Dworkin (1996) calls them, emanate from wrong actions, such as Godzilla kidnapping a maiden and eating her, and affect the observer in such a way that they end up thinking that what Godzilla did was wrong. Since Sturgeon cannot use the above examples and has no alternative plausible story about how irreducible moral properties play a genuine explanatory role in people's moral beliefs and judgements or how wrongness is causally efficacious to the production of our experience, it is acceptable to conclude that moral properties do not play a genuine role in explaining the formation of moral beliefs.

Harman re-affirms this point when he argues that the counterfactual dependence of Sturgeon's purported moral properties cannot justify the explanatory relevance of moral properties because a moral *epiphenomenalist*- someone who accepts the explanatory impotence of moral properties – can accept that the relevant counterfactual conditional is true. The moral epiphenomenalist accepts that moral properties supervene onto the non-moral properties (for instance, that wrongness supervenes onto unprovoked kidnapping, torture, abuse etc.), but that moral properties have no explanatory relevance themselves. That is, the counterfactual dependence of moral belief upon moral properties is consistent with the denial that moral properties are explanatorily relevant. So Sturgeon's claim that moral properties are explanatorily relevant in explaining why Anu formed her belief that what Godzilla was doing is wrong is unconvincing. The non-moral properties of the Godzilla situation were causally efficacious in the production of Anu's belief that what Godzilla was doing is wrong, but wrongness itself was not causally efficacious in producing Anu to form

her belief that what Godzilla was doing was wrong. So although a moral epiphenomenalist can accept that the relevant counterfactual conditional is true, this does not show that moral properties are genuinely explanatorily relevant in production of Anu forming her belief that what Godzilla was doing is wrong.

Sturgeon's response against Harman's counter-reply

Sturgeon (1986) agrees that a moral epiphenomenalist would pass the counterfactual test, but argues that we need to find an independently plausible reason to accept moral epiphenomenalism, otherwise moral epiphenomenalism does not undermine the intuition we have that moral properties are explanatorily relevant. The intuition remains because all supervening facts would be epiphenomenal since the only lower-level properties that have a causally efficacious mechanism are microphysical properties. The properties of the special sciences (chemical, biological, psychological, and social properties) will all be rendered epiphenomenal since they supervene onto the microphysical properties and are causally inefficacious and explanatorily impotent. According to Sturgeon this is implausible.

Furthermore, Sturgeon (2006) claims that this epiphenomenalist argument against the supervenience thesis proves too much. Physicalists take psychological properties to supervene on physical properties, that the base physical properties are sufficient for psychological states, for example, pain. However, and importantly, for Sturgeon, they resist the conclusion that psychological states are epiphenomenal or that psychological states are causally inefficacious in the production of our

experience[16]. Why should we do so purely in the moral case and not in the other cases? Bringing both responses together, Sturgeon is arguing that although moral epiphenomenalism is a logical possibility since it passes the counterfactual test, by itself this does not damage the claim that moral properties are explanatorily relevant. Epiphenomenalism is not accepted in other areas of philosophy; we need an independent reason to accept moral epiphenomenalism in this case.

Miller (2003) finds Sturgeon's reply here to Harman implausible for another reason. Miller points out that Sturgeon's original claim was that the counterfactual dependence of "b's being G upon a's being F is *sufficient* to justify the idea that the property F is explanatorily relevant to b's being G: this is what allows him to argue that moral properties are explanatorily relevant." (Miller, 2003, 149 [original italics]) So the fact that moral epiphenomenalism is logically possible does undermine Sturgeon's sufficiency claim: the counterfactual test is *not* sufficient to show the explanatory relevance of moral properties because moral epiphenomenalism asserts that moral properties are explanatorily impotent at the same time as accepting the claim about counterfactual dependence. So, this means that Sturgeon needs to find some alternative way of responding to Harman. In the next chapter we will look at program explanation as a way of responding to Harman on behalf of Sturgeon.

---

[16] This is just one example out of many that Sturgeon can run – for instance, he could use the example of chemical properties supervening on physical properties, or social properties supervening on psychological properties amongst others. In those examples, the supervening property will not be seen as epiphenomenal or as causally inefficacious to the production of our experience.

Chapter 3

In this chapter I will suggest a possible solution for Sturgeon in response to Harman's objection by the use of program explanation (this is derived from Miller 2003). Firstly, the need for program explanation will be motivated by examining the role of properties in causal explanations, which eventually yields a counter-intuitive conclusion that can only be assuaged by the use of program explanation. Secondly, what exactly program explanation is and the importance of program explanation will be discussed by showing that it imparts information not imparted by the relevant process explanations. Finally, it will be shown how and why program explanation can be used as an attempted solution to Harman's objection, showing that a program explanation can be run to show that the unavailability of moral properties will lead to an explanatory impoverishment.

Program explanation

To explain what program explanation is we start off with Frank Jackson and Philip Pettit. In Jackson and Pettit (2004) they identify three common and plausible background assumptions about the role of properties in causal explanations. They then suggest that if we add a fourth assumption we get a problem. The notion of program explanation emerges via the rejection of the fourth assumption.

(1) A causal explanation of something must direct us to a *causally relevant* property as opposed to a *causally irrelevant* property of the factor it identifies as explanatory: a property relevant to the causal production of the effect explained.

For example: a causal explanation of why a keycard unlocked a secured door must direct us to the property of having a suitable electromagnetic strip rather than the colour of the keycard, the electromagnetic strip is relevant but the colour is not.

(2) One way in which properties are causally relevant is by being causally efficacious. A causally efficacious property in virtue of whose instantiation, at least in part, the effect occurs; the instance of the property helps to produce the effect and does so because it is an instance of that property.

One way to be causally relevant is to be causally efficacious. For example: the property of having a suitable electromagnetic strip is causally efficacious because it is in virtue of the keycard's having the strip that the door unlocked. The intuition is that if the property is causally efficacious the effect occurs, at least in part, because of the instantiation of that property (in which case the property is obviously causally relevant).

(3) A property F is not causally efficacious in the production of an effect e if these three conditions are fulfilled together.

Intuitively, if conditions (3i) (ii) and (iii) are satisfied, then property F is not causally efficacious in the causal production of the effect occurring**,** since all the causal work will be done by the relevant G property.

(i) there is a distinct property G such that F is efficacious in the production of e only if G is efficacious in its production;

(ii) the F-instance does not help to produce the G-instance in the sense in which the G-instance, if G is efficacious, helps to produce e; they are not sequential causal factors

(iii) the F-instance does not combine with the G-instance, directly or via further effects, to help in the same sense to produce e (nor of course, vice versa): they are not coordinate causal factors. (Jackson and Pettit, 2004, 120)

We now add a fourth assumption:

(4) The only way for a property to be causally relevant to the production of a certain effect is by being causally efficacious in the process of production (Jackson and Pettit, 2004, 124).

Taken together these four assumptions are incompatible with strong intuitions about the role of certain properties in explanation. To illustrate this point, consider the following example[17]: suppose that we boil water in a closed glass container, and the water reaches such a temperature that the glass cracks. What explains why the glass cracks? Intuitively, we say things like "that glass cracked because of the temperature," or "that glass cracked because the water was boiling," or something similar along these lines. The temperature of the water or the fact that the water was boiling is causally relevant to the cracking of the glass.

However, as Jackson and Pettit explain:

---

[17] The following example is from Jackson and Pettit (2004)

"Why did it crack? First answer: because of the temperature of the water. Second answer, in simplified form: because of the momentum of such and such a molecule (group of molecules) in striking such and such a molecular bond in the container surface. The temperature property was efficacious only if the momentum property was efficacious: hence 3(i). But the temperature of the water – an aggregate statistic[18] – did not help to produce the momentum of the molecule in the way in which it, if efficacious, helped to produce the cracking: hence 3(ii). And neither did the temperature combine with the momentum to help in the same sense to produce the cracking: one could have predicted the cracking just from full information about the molecule and the relevant laws. Hence 3(iii). (Jackson and Pettit, 2004, 110)

So according to assumption (3), the temperature of the water is not causally efficacious in the cracking of the glass, and from assumption (4) the temperature of the water is not causally relevant to the cracking of the glass. This conclusion is extremely counter-intuitive.

The example of temperature and the cracking of the glass feeds into a more general point about assumption (4). Assumption (4) has two devastating consequences for causal explanations from common sense and the special sciences (i.e. chemistry, biology, psychology and sociology.) Firstly, it means all true, proper and real causal explanations of something are to be found at the level of fundamental physics. Since the ultimate base properties of the supervenience relation are microphysical properties[19] [20], and all the special sciences ultimately supervene on microphysical

---

[18] By aggregate statistic here, take it to mean something like average momentum of constituent water molecules.

[19] In Jackson and Pettit's paper they assume that the causally efficacious properties are the physical (or microphysical) properties. It is neither implicit or explicit in the text whether this takes into account quantumn mechanics. Whether it does or not isn't a worry for program explanation since all it requires is that you believe there is a lower-order level at which there are causally efficacious properties in the production of the relevant effect occurring.

[20] What happens if the lower-level property mentioned is not itself the efficacious property in the causal production of the effect? For instance, what if the lower-level property mentioned is the microphysical

properties, - and given assumption (3) about the role of properties in causal

explanations - then in all causal explanations of something the microphysical

properties are the causally efficacious properties in the causal production of an effect.

Quite clearly, if a causal explanation of something can only cite the properties of

physics, then we must deny that the special sciences and common sense are able to

give explanations in terms of causally relevant properties. Causal explanations from

the special sciences and common sense cite causally inefficacious and thus causally

irrelevant properties in causal explanations.

Secondly, a consequence of the first point means that if we have a higher-

order explanation and a lower-order explanation of the same phenomena, then we

must replace the higher-order explanation with the lower-order explanation. So, for

example, if we ask why did the glass break? The causal explanation we should prefer

is not in terms of the higher-order property of temperature, but rather in terms of

lower order momentum properties of certain molecules because the lower-level

explanation directs us to causally efficacious properties in the production of the

relevant effect[21]. Intuitively such a view is unattractive, it goes against common sense.

Jackson and Pettit's solution is to reject assumption (4) by arguing that

properties can be causally relevant without being causally efficacious. How so? A

---

properties but they are not causally efficacious because actually quantum mechanics is? Then, as
Jackson and Pettit point out, "the story will go a level or more deeper until we find an efficacious
property for which the original higher-order property programs indirectly, via the programming of the
intermediate features." (Jackson and Pettit, 2004, 120)

[21] An interesting consequence of this, as Jackson and Pettit explain, is that it is not enough to cite (for
example) that some (indeterminate) molecules are striking a molecular bond because "explanation
involving existential quantification – the reference to an indeterminate some – cannot be a proper
explanation: it does not invoke an efficacious property and so does not invoke a property that is
relevant." (2004, 124-125) For lower-order properties to be relevant in causal explanations, the lower-
order properties cited must be the exact causally efficacious properties. Most of the time we do not
know the exact causally efficacious properties in a causal explanation of something, so some of us may
go through our lives never offering a relevant causal explanation of anything.

property can be causally relevant without being causally efficacious because the instantiation of a certain property (higher-level properties) ensures the instantiation of properties (distinct lower-level properties) that are causally efficacious in the causal explanation. Although the higher-level property in question is causally inefficacious and causally inert, its realization *programs* for the instantiation of lower-level properties that are causally efficacious in the causal production of the effect occurring. In this sense the higher level property is causally relevant in causal explanations. This is known as program explanation. As Jackson and Pettit explain it, it is called *program* explanation because of the *programming* metaphor:

"A useful metaphor for describing the role of the property is to say that its realization programs for the appearance of the productive property and, under a certain description, for the event produced. The analogy is with a computer program which ensures that certain things will happen – things satisfying certain descriptions – though all the work of producing those things goes on at a lower, mechanical level". (Jackson and Pettit, 2004, 127)

So, for example, using program explanation, in the context of the temperature of the water and the glass cracking case, the temperature of water is in fact causally relevant without being causally efficacious. The realization of the boiling temperature (a higher-level property) programs for the instantiation of such-and-such momentum properties of the molecules (the lower-level properties).

So by the use of program explanation the temperature of the water is causally relevant to the cracking of the glass. But what reason is there to choose (in some cases) the higher-level explanation instead of the lower-level explanation? That is, if we have lower-level explanations available should we not just ditch the higher-level explanation in preference to the lower-level explanations every time?

No, there are times when it is preferable to use a higher-level explanation because a program explanation "may provide information which the corresponding process explanation [lower-level explanation] does not supply. Thus, it may be an explanation which the process explanation does not supersede." (Jackson and Pettit, 2004, 129)

There are two ways in which the higher-level explanation, the program explanation, can provide information not known at the corresponding process explanation in the context of the temperature of the water and the glass cracking case. Firstly, it is possible that a molecule can strike the container with sufficient momentum at the correct place and time *without* the water being at boiling temperature. So the explanation in terms of the boiling temperature conveys information that the corresponding process explanation doesn't (namely, that the water is at boiling temperature). Secondly, which is an elaboration of the first point, program explanation imparts modal information that is not conveyed by the corresponding process explanation. (Jackson and Pettit, 2004, 129) So to have the explanation in terms of the water's being boiling tells us that in any relevantly similar possible world, that there will be molecules with sufficient momentum, place and time to crack the glass container. That is, to know that the water is boiling means that in the actual world it was these molecules (for example, molecules x,y and z) that crack the glass, but had *those* molecules not cracked the glass, then in other relevantly similar possible worlds other molecules will crack the glass (for example, molecules f, g and h).

Contrast these higher-level explanations, program explanations, with lower-level explanations, process explanations. The process explanations only tells us which momentum properties of the molecules are causally efficacious in cracking the glass,

but it tells us nothing about whether the water is at boiling temperature or not. Moreover, the process explanations only tell us about the *actual world* and the original situation itself, without telling is what would happen in the *relevantly similar possible worlds*.

As should be evident now, clearly program explanation fits into strong supervenience and multiple realizability. Temperature properties strongly supervene on the momentum properties of groups of molecules. If a sample of water is boiling, then there is some momentum property of its constituent molecules such that, necessarily, any sample of water whose constituent molecules have the same momentum properties is also boiling.

The supervenience relation in this case is not reductive**:** for any given temperature property there are going to be an indefinite number of ways in which it can be realized at the level of momentum properties of constituent molecules.

Now the obvious link between program explanation, strong supervenience and multiple realizability becomes clear. *Whenever we have strong supervenience and multiple realization we have the ingredients for informative program explanation*. The program explanation, in the case of the boiling water and cracking glass, selects one combination out of the multitude possible combinations of the momentum properties of the molecules to be causally efficacious in the causal production of the glass cracking. If it had not been that particular combination selected, it would have been another possible combination of the momentum property of the molecules that were causally efficacious in the causal production of the glass cracking, and this is only possible as the boiling water can be multiply realized by an almost infinite number of possible combinations of the molecules, momentum, time and place that are causally efficacious in the causal production of the glass cracking. This can only

happen though because the relationship between the momentum property of the molecules and the boiling water holds as it is one of strong supervenience, enabling the boiling water to be able to program for the instantiation of one out of a multitude of possible combinations of the momentum property of the molecules that are causally efficacious in the glass cracking.

A final note on program explanation: as program explanation conveys information not available at the level of the corresponding process explanations, the unavailability of the program explanation would constitute an explanatory loss, so program explanation can count as best.

Program explanation, Sturgeon and Harman

The important question now is: can Sturgeon respond to Harman's objection? It is not enough for Sturgeon to argue that moral properties are causally efficacious. That, say, when a wrong act has been committed morons emanate from a wrong action and affect the observer into thinking such actions are morally wrong. For instance, Anu observes Godzilla kidnapping and eating a maiden, and from Godzilla's actions emanate some "morons" that are causally efficacious in some way to make Anu believe that what Godzilla was doing is wrong. This is not a plausible story of how moral properties play a role in explaining people's moral beliefs and judgements.

Miller (2003) has argued that Sturgeon can respond to Harman by arguing that although moral properties are not causally efficacious in the production of moral beliefs, by the use of program explanation moral properties are causally relevant.

Moral properties can "program" for the instantiation of the sufficient non-moral properties that are causally efficacious in the production of moral beliefs[22]?

To see this, consider Harman's example of Godzilla and Anu. Anu sees Godzilla kidnap and eat a maiden and forms the belief that what Godzilla is doing is wrong. We can give an explanation of Anu forming her belief by citing the non-moral properties, such as kidnapping the maiden, the pain of being eaten, and these non-moral properties are causally efficacious in producing Anu's belief that what Godzilla is doing is wrong.

We can say that the explanation of Anu's forming that belief that Godzilla's act is wrong in terms of its wrongness is a program explanation. The wrongness of Godzilla's act is causally inefficacious in the production of Anu forming the belief that what Godzilla is doing is wrong. Godzilla's act had certain natural features, such as kidnapping the maiden and putting her into his mouth to eat her, which made it true that Godzilla's act was wrong and which were causally efficacious in producing Anu's belief that what she saw was wrong. So the wrongness of Godzilla's act is only efficacious in Anu's belief if the non-moral properties are causally efficacious, hence (3)(i). The wrongness and the non-moral natural properties are neither sequential nor causal co-ordinate factors in the production of Anu's belief, hence (3)(ii) and (iii). So the moral property is inefficacious. However, the (higher-level property) wrongness ensures the instantiation of the (lower-level properties) non-moral properties that are causally efficacious in the production of Anu forming the belief that what Godzilla was doing is wrong. So we can give a program explanation of Anu's forming the relevant moral belief.

---

[22] Although Miller offers the argument on behalf of the Cornell Realist, Miller ultimately denies that the argument is successful as a response to Harman's objection. For Miller's own words check (2003, 150 – 174) I will discuss this fully in the next chapter anyway.

The program explanation counts as best because it imparts information that is not available in the corresponding lower-level explanations in two ways. Firstly, kidnapping the maiden and eating her can be causally efficacious in producing Anu to form the belief that what has just seen is wrong, even though it is not: we can think of a far-fetched scenario in which the only way to save the maiden from an even worse fate is by eating her alive. Secondly, program explanation imparts modal information that is not found at the corresponding lower-level explanation: consider another relevantly similar possible world in which we substitute the non-moral properties of eating her to throwing her off the empire state building to her death, the non-moral properties would of still have been causally efficacious in the production of Jane forming the belief that what Godzilla did was wrong.

So, we can give a program explanation of the formation of Anu's belief in terms of the wrongness of Godzilla's actions. The unavailability of program explanation would lead to an explanatory impoverishment, impoverishment of the relevant information imparted by the program explanation. (Miller, 2003, 153 - 154) So the program explanation that cites the moral features of things can, in the relevant sense, count as best. So, moral properties are a real properties because, due to premise (1) of Cornell realism, "P is a real property if and only if P figures ineliminably in the best explanation of experience." And so we can earn ontological rights for moral properties without bringing in morons.

Chapter 4

To recap, the issue at hand is whether the use of moral program explanation will earn

ontological rights for moral properties. To do so, moral program explanations must

figure ineliminably in the best explanation of experience and it must also be the case

that the unavailability of moral program explanations will result in an explanatory

loss. In this chapter I look at four arguments ((Miller (2003), Nelson (2006), Miller

(2009), Bloomfield (2009)) that are split into alternating sections of arguing that

program explanation does *not* earn ontological rights for moral properties and that

program explanation does earn ontological rights for moral properties. It so turns out

that the arguments are in historical order, so it will be possible to follow the historical

order of the discussion of whether program explanation earns ontological rights for

moral properties for the Cornell realist variety of moral realism.

Before I discuss the selected four arguments, it is important just to note that

the case of the temperature, water and the cracking of the glass container is used as a

model for the application of program explanation to the moral case in all the

arguments discussed.

Does program explanation successfully defend Sturgeon against Harman's objection?

Alex Miller (2003) does not think that program explanation can earn ontological

rights for moral properties. Program explanation figures ineliminably in the best

explanation of our experience because program explanations being ""best" – such that

the lack thereof [of program explanation] results in an explanatory loss – is a direct

consequence of some *epistemic limitation* of ours." (Miller, 2003, 173 [original

italics]) So, according to Miller, the unavailability of program explanation leads to an explanatory loss due to our epistemic limitations: once these limitations are lifted, in the way they have to be when we are considering the question of ontological rights, the unavailability of program explanation no longer results in an explanatory loss.

What does Miller mean by an epistemic limitation? In his initial attempt at the argument in Miller (2003) he is not explicit about what an epistemic limitation is. In Miller (2009) he is: by epistemic limitation Miller means epistemically limited vis-à-vis lower-level properties and the process explanations in which they figure[23]. Call an agent that is epistemically unlimited "the lower-level computer" (the L-L CPU), the most perfectly designed and built computer in the universe. It is *vitally important* to note that the introduction of the L-L CPU is intended merely as a heuristic device rather than as the introduction of an actual agent[24].

Recall that program explanations conveyed information not available at the level of the corresponding process explanations via the grasp of an epistemically limited agent in two ways. By conveying the additional information: (i) that the water is at boiling temperature, and (ii) that the glass would crack in relevantly similar possible worlds. Miller's argument is that program explanation conveys distinctive information, information of types (i) and (ii), not conveyed in process explanations because of our epistemic limitations. Thus, if we had no epistemic limitations,

---

[23] Henceforth within the main body of the text, I will simply refer to the epistemically unlimited agent vis-a-vis lower-level properties and the process explanations in which they figur as an epistemically unlimited agent. However, I will sometimes use the full title of the epistemic agent to highlight its function as an heuristic device. Likewise, "epistemic limitations" should always be read as 'epistemic limitations vis a vis the lower facts and process explanations in which they figure".

[24] I have changed Miller's example of God being an epistemically unlimited agent vis-à-vis lower-level properties and the process explanations in which they figure as a heuristic device due to the potential ambiguity with the omniscient agent, God. Hopefully by using the "L-L CPU" terminology it will be clearer that the epistemically unlimited agent vis-à-vis lower-level properties and the process explanations in which they figure is simply a heuristic device. Furthermore, and lastly, in all examples and quotes in which God has been used as a heuristic device I have replaced God with the L-L CPU

program explanations would convey the same distinctive information conveyed in the process explanations.

Remember that program explanations can only confer ontological rights for moral properties if information of type (i) and (ii) is not conveyed in the corresponding process explanations. The relevant question is: would a program explanation convey distinctive information of types (i) and (ii) to an agent epistemically unlimited vis-à-vis lower-level properties and the process explanations in which they figure? I will investigate Miller's responses to (i) and (ii) in turn.

Miller's response to (i) is that the distinctive information conveyed by program explanation of type (i) does not confer ontological rights for moral properties because "(i) is actually not relevant in the cases where we are concerned with whether program explanation can confer ontological rights." He goes onto say, "[i]f we *assume* that there is a fact about the temperature of the molecules, knowledge of which would be lost given the unavailability of a program explanation, then we are already assuming what we set out to prove: namely, that there are higher-order facts and properties and properties of the relevant kind." (Miller, 2003, 173 [original italics]). So according to Miller the issue purely concerns the capacity of program explanations to convey information of type (ii).

Miller's response to (ii) is that when considering ontological rights for higher-level properties we are wondering whether the higher-level properties would feature in the perspective an epistemically unlimited agent, the L-L CPU. So in the perspective of the L-L CPU is the program explanation still "best" in that its unavailability it would result in an explanatory loss?

Miller argues that in the perspective of the L-L CPU the unavailability of the program explanations would *not* result in an explanatory loss. As Miller says:

"Given that [the L-L CPU] knows all of the facts about the relevant process explanations and the modal information about how things would go in relevantly similar possible worlds, it is clear that he would not. How, then, could program explanation earn ontological rights for higher-order properties?" (Miller, 2003, 174)

Simply put, Miller is arguing that by removing the shackles of epistemic limitation by looking at the situation from the perspective of an agent who knows all there is to know about lower-level properties, the process explanations in which they figure, and how things would go in different possible worlds, we can see that no additional distinctive information is conveyed by the use of program explanations. So there is no need to add higher-level properties into our ontology.

Miller points out that the same applies to higher-level moral properties:

"Likewise, if we assume that [the L-L CPU] knows all of the naturalistic explanations of moral beliefs, as well as modal information about how things would go in relevantly similar possible worlds it is clear that the unavailability of program explanations would not result in his suffering an explanatory loss. Thus, program explanations which invoke moral facts and properties do not count as best for [the L-L CPU]; thus, they do not count as best in the right sort of way for their availability to earn ontological rights for the higher-order properties in which they trade." (Miller, 2003, 173-174)

Miller's argument, then, is as follows:

"(M1) Higher-level properties are causally relevant in a way that can earn them ontological rights only if they figure in the best explanation of experience, considered

from the point of view of a subject who suffers no epistemic limitations *vis-à-vis* lower-level properties and the process explanations in which they appear.

(M2) It is not the case that higher-level properties figure in the best explanation of experience, considered from the point of view of a subject who suffers no epistemic limitations *vis-à-vis* lower-level properties and the process explanations in which they appear

Therefore,

(M3) It is not the case that higher-level properties are causally relevant in a way that can earn them ontological rights." (Miller, 2009, 337 [original italics])

Mark Nelson (2006) denies (M2) by arguing: *it is the case* that higher-level properties figure in the best explanation of experience, considered from the point of view of a subject who suffers no epistemic limitations vis-à-vis lower-level properties and the process explanations in which they appear[25]. So, according to Nelson, the unavailability of program explanation would lead to an explanatory impoverishment even from the perspective of the L-L CPU.

---

[25] Interestingly in his paper Nelson (2006, 423) states his second premise as "[i]t is not the case that higher-level properties figure in the best explanation of phenomena, considered from the point of view of a subject who suffers no epistemic limitations." Compare this to Miller's second premise in the text above in which Nelson makes no reference to the agent being epistemically unlimited with reference to the lower-level properties. This is important because it seems that Nelson *may have* misinterpreted what Miler meant by **"God"**. Miller does not mean God as an agent *per se* as an omnipotent agent, but a heuristic device of an epistemically unlimited agent in regards to the lower-level properties and the process explanations in which they figure. Such a view is plausibly backed up by the fact that Miller's argument is saying with the shackles of epistemic limitation taken off, program explanations convey no additional information not already conveyed by process explanations. Moreover, such a view is backed up by Miller in (2009) emphasises the **"vis-à-vis"** rendering of his premises in the argument. However, whilst saying all of this, Miller is probably to blame for using the term "God" without adequate explanation in the original presentation of the argument.

Nelson agrees with Miller on two points, in context of the example of the boiling water cracking the glass. Firstly, that the relevant process explanations explain why effect *e*, the cracking of the glass occurs. Secondly, that the relevant process explanations also say how things do go in particular possible worlds: for example, in one particular relevantly similar possible world, molecules f, g, and h crack the glass. But according to Nelson the fact that the relevant process explanations explain what goes on in particular possible worlds it is not enough to explain why the counterfactual (e*) is true.

Counterfactual (e*):

(e*)"if molecules x, y and z had not struck the molecular bonds at place p_n at time t_n with momentum m_n, the glass still would have cracked, due to the striking of other molecules, at other times, or other places, or with slightly different momenta." (Nelson, 2006, 424)

As Nelson explains why:

"But none of this is sufficient to explain why (e*) is true … [the lower-level properties] about the microphysical state and the history of the world right up until the occurrence of e, and of all the relevant causal laws can't explain it; they can explain only how subsequent history did in fact turn out. *The true counterfactuals can't explain it, either. They say only how things do go in particular possible worlds, but they don't explain why it is true that some one of a limited range of possible worlds, out of limitless panoply of worlds that are possible will be actualized. Miller stipulates that [the L-L CPU], our epistemically unlimited subject, knows that (e*) is true, but it one thing to know that a proposition is true; it is another thing to have an explanation for it.*" (Nelson, 2006, 424 – 425 [emphasis added])

Nelson denies that knowing how the relevant process explanation go in particular possible worlds explains why (e*) is true. According to Nelson, knowing which exact molecules crack the glass in the relevant particular possible worlds is not the same as explaining the *general claim* as to why the glass cracks in relevantly similar possible worlds. We can see this as Nelson (2006, 425) says "Miller stipulates that [the L-L CPU] our epistemically unlimited subject [vis-à-vis the lower level properties and the process explanations in which they figure] knows that (e*) is true, but it is one thing to know that a proposition is true; it is another thing to have an explanation for it." And Nelson then goes onto say "Moreover, even if these materials enabled [the L-L CPU] to see that a particular possible world *would* be actualised in an alternative process-scenario to [the striking of molecules x,y, and z on the molecular bonds at place p_n at time t_n with momentum m_n] (because, for example, he could that molecules u,v and w were "next in queue to crack the glass," so to speak), this would not be an *explanation* of the general claim (e*), that *some* such world would be actualised. This is so because *ex hypothesi* (e*) still would have been true if still *other* molecules had been next in queue, so the particular claim about u,v, and w cannot be the explanation of the general claim." (ibid.)

What does Nelson mean by the general claim (e*)? Nelson seems to mean that if molecules x,y and z do not crack the glass, then there will always be molecules with sufficient momenta, position and time to crack the glass in other relevantly similar possible worlds, as the general claim (e*) guarantees this. Nelson does not think it is enough to cite the relevant process explanations in the possible worlds in order to explain why (e*) is true because (e*) would still be true had *other* molecules cracked

the glass in the relevantly similar possible worlds, and thus the true counterfactuals cannot be an explanation of the general claim.

Nelson thinks that only program explanation can explain why the general claim (e*) is true and that the other relevantly similar possible worlds where the glass cracks will be actualised if molecules x, y and z do not strike the molecular bonds with sufficient momentum etc. Recall that the glass of water being at boiling temperature is multiply realizable, it does not matter which molecules strike the molecular bonds, all that matters is that the molecules that do strike the molecular bond have sufficient kinetic energy to break it. The fact that the water is at boiling temperature means that some molecules are doing enough of what they have to do to crack the glass, the boiling temperature of water guarantees that other molecules will be available with sufficient momenta, at the relevant time and place to crack the glass, if molecules x, y, and z do not.

Notice that both sides of the argumentative debate both make the jump between the fact that sufficient molecules will do enough to crack the glass in the actual world and the fact that other sufficient molecules will do enough to crack the glass in other possible worlds. For the sake of argument both sides assume that this is true, in other words, both sides assume that counterfactual (e*) is true.

Nelson's argument then is that the program explanation conveys distinctive information not available via an epistemically unlimited agent's grasp of a process explanation because only program explanation can be invoked in order to explain why (e*) is true. So, applying this to the moral case, the unavailability of the relevant program explanations will result in an explanatory loss for the L-L CPU. So program explanations that invoke moral facts and properties do count as best for the L-L CPU,

and thus count as best in the right sort of way for their availability to earn ontological rights for moral facts and properties.

It is interesting to note what has happened now. Previously, program explanation featured ineliminably in the best explanation of our experience due to our being epistemically limited since it offered modal information, information of type (ii), not available via an epistemically limited agent's grasp of a process explanation. Now, with an epistemically unlimited agent, according to Nelson, program explanation is still "best" because it explains *why* the counterfactual (e*) is true, even though an epistemically unlimited agent of the relevant sort would know that the counterfactual is true. In attempting to deny this line of thought the relevant question to ask is whether process explanations explain why the counterfactual (e*) is true? But before we look at Miller's (2009) response which denies Nelson's line of thought by arguing that process explanations can explain why the counterfactual (e*) is true, lets assess Nelson's argument.

Nelson's argument is unconvincing. To see why take the most important section of Nelson's argument:

"The true counterfactuals can't explain it, either. They say only how things do go in particular possible worlds, but they don't explain why it is true that some one of a limited range of possible worlds, out of limitless panoply of worlds that are possible, will be actualized."

As already discussed, Nelson is arguing that the true counterfactuals show that a limited range of possible worlds are actualised where the water cracks the glass out of a unlimited panoply of possible worlds, but the true counterfactuals do not explain *why* a limited range of possible worlds will be actualised where the water cracks the glass.

The problem for Nelson is that this requirement of asking for an explanation of why the counterfactual (e*) is true, other than by knowing that the glass will crack in other possible worlds, is unmotivated. When Nelson asks for an explanation of *e*, the cracking of the glass in the actual world, Nelson feels it is sufficient to cite only the molecules involved in the process explanations. As Nelson says:

"If [the L-L CPU] traced the whole chain of process explanations of G1 [(the striking of molecules x, y, z on the molecular bonds at place p_n at time t_n with momentum m_n)] back to the beginning of time, he would need to refer only to G-type explanations, never to F-type explanations [(the water's having reached boiling temperature.)]" (Nelson, 2006, 424)

Now, clearly when we move onto possible worlds, Nelson agrees that the L-L CPU knows the relevant modal information and how things do go in possible worlds. The L-L CPU can trace the chain of process explanations and molecules back to the beginning of time. As Nelson, once again, says,

"[The L-L CPU] is left with knowledge of the microphysical state and history of the world right up until the occurrence of e, knowledge of all the relevant causal laws, and knowledge of the relevant true counterfactuals … The true counterfactuals can't explain it, either. They [the true counterfactuals] say only how things *do* go in particular possible worlds." (ibid. [original italics])

If we are following the blueprint of what counts as an explanation *why* in the actual world e, the cracking of the glass in the actual world occurs, then this should count as an explanation *why* the limited range of possible worlds will be actualised, namely, how the molecules and process explanations do go in the relevant particular possible

worlds. Nelson's additional constraint, then, as to what is required to explain why the counterfactual (e*) is true seems unmotivated.

Why then does Nelson hold this additional constraint? The only explanation available is his reading of what the counterfactual (e*) entails. Nelson's reading of the counterfactual (e*) entails that any explanation of why (e*) is true has to explain the general claim and this can not be explained by reference only to the true counterfactuals, as Nelson has already explained. However, if Nelson's reading of the counterfactual (e*) is wrong, and thus the explanation of the general claim is not needed to explain why (e*) is true, then his additional constraint is unmotivated.

Miller's reply to Nelson

Miller (2009) offers a reply to Nelson (2006) arguing that the L-L CPU would *not* need program explanation to know why the counterfactual (e*) is true. Miller is using the fact that counterfactuals are context-sensitive in order to undermine Nelson's argument.

It is widely acknowledged and accepted that counterfactuals are context sensitive in the sense that "which propositions an utterance of a counterfactual expresses is determined by the context of an utterance." (Miller, 2009, 338–339) As John Divers explains.

"What is special about counterfactual conditionals is that, even for different tokens of the same sentence type, the set of worlds that is relevant to their truth is something that varies with context of utterance." (Divers, 2002, 11)

And Jonathon Lowe:

"Quite generally, a sentence (type) is "context-sensitive" if the proposition expressed by (a token of) it is partly determined not merely by the standard meanings of its constituent words but also by its circumstances of utterance." (Lowe, 1995, 52)

"A speaker's intentions in asserting a counterfactual can help to determine the propositional content of his assertion by fixing an appropriate measure of similarity across possible worlds for the proper evaluation of the truth or falsity of what he asserts." (ibid. 55)

To illustrate what Divers and Lowe mean by counterfactuals being context-sensitive, consider this example.

Suppose two people Michael and David are outside on a freezing cold winter's day and Michael and David know that the cycle path, which is situated near them, is extremely icy and really slippery. From a short distance away Michael and David can see their friend Charlie, who is an extremely cautious individual. If Charlie wants to walk directly closer towards Michael and David, then he will step onto the cycle path.

Consider:

(1) "If Charlie walks any further, then he will fall over" as uttered by Michael who wants to convey information about what happens when someone walks on really slippery ice.

(2) "If Charlie walks any further, then he will *not* fall over" as uttered by David who wants to convey information about how cautious Charlie is when it comes to walking on unsuitable terrain.

The counterfactual's context-sensitivity shows that (1) and (2) can be true at the same time. (1) conveys information about what would happen to Charlie if he walked onto the cycle path due to the really slippery nature of the cycle path, (2) conveys information about Charlie's cautious character, as he would find another non-slippery route to walk.

How do we test the truth conditions of counterfactuals? "A counterfactual conditional is true if and only if in the closest possible world in which its antecedent is true its consequent is also true[26]." (Miller, 2009, 339) To evaluate Michael's utterance (1), the closest possible world in which Charlie steps onto the cycle path is one where his psychology is different. To evaluate David's utterance (2), the closest possible world is one where Charlie's psychology remains the same so that he walks further towards Michael and David only *after* he has found an alternative non-slippery route to the cycle path. Michael's utterance is true since in the relevant world Charlie falls over. David's utterance is true since in the relevant world Charlie does *not* fall over.

How, then, does the fact that counterfactuals are context-sensitive undermine Nelson's objection? According to the Stalkner-Lewis account (e*) is true ""if and only if in the closest possible world to the actual world in which molecules x, y, and z do not strike the molecular bonds at place p_n at time t_n with momentum m_n, some other molecules strike molecular bonds with sufficient momentum to break them." (Miller, 2009, 340) So in order to explain why (e*) is true all the L-L CPU would need to know, according to Miller, is which possible world is closest to the actual world and what goes on at the level of groups of molecules in that world. The important question is: would the L-L CPU have to call upon higher-level temperature properties in order

---

[26] This isn't just a test, its also an account of the truth-conditions of counterfactuals. Moreover, to keep things simple we are assuming the Stalnaker-Lewis account of counterfactuals

to see which possible world is closest to the actual world and thereby to see why (e*) is true?

Miller argues that we do not have to call upon higher-level temperature properties as the L-L CPU (as Miller and Nelson both agree) knows all the lower level facts (in this case, all of the facts that can be characterised without using temperature concepts) about the universe, "*including the facts about which proposition any given utterance of (e*) expresses.*" (Miller, 2009, 341 [original italics]) The L-L CPU already knows which is the closest possible world to the actual world since he knows which proposition any given utterance of (e*) expresses. So, clearly the L-L CPU does not need to call upon higher-level temperature properties to know *why* (e*) is true.

Recall earlier I remarked that Nelson's reply to Miller was unconvincing because Nelson accepted it was sufficient to cite facts about the momentum properties of molecules since the beginning of time to explain why *e* occurs, the cracking of the glass in the actual world. Yet Nelson denies that facts about the momentum properties of the molecules since the beginning of time, combined with the true counterfactuals, is sufficient in explaining why (e*) is true, and this I argued was unmotivated. This point feeds into Miller's argument here in a way that undermines Nelson's argument: by knowing which proposition any utterance of (e*) expresses and by knowing which possible world is closest to the actual world then we can see that knowledge of process explanations at the level of groups of molecules is sufficient to explain why (e*) is true. That is, the L-L CPU can see why (e*) is true in virtue of information at the level of process information in the closest possible world to the actual world.

Applying this to the moral case, as the L-L CPU is assumed to have all the process explanations of moral belief in terms of lower-level non-moral properties and

modal information about how things go in other possible worlds, following Miller's (2009) argument, it is clear that the unavailability of program explanation for the L-L CPU would *not* result in an explanatory loss. As the L-L CPU knows which proposition any given utterance of (e*) expresses, and thus knows the closest possible world to the actual world, the L-L CPU will know why (e*) is true.

Bloomfield reply to Miller

Bloomfield (2009) denies that the fact that counterfactuals are context-sensitive undermines Nelson's argument, and his reason for this is as follows. Bloomfield describes for us three possible worlds in which molecules break the glass if the molecules x, y and z do not break the glass: (i) "w, where the water is boiling and molecules p, q and r break the glass, (ii) w*, where the water is boiling and molecules f, g and h break the glass and, (iii) w** where the water isn't boiling but an angel speeds up only molecules p, q and r in such a way as to break the glass." (Bloomfield, 2009, 343) Bloomfield argues that context sensitivity, which should be more accurately characterized as conversational implicature, will only allow the L-L CPU to rule out w**[27]

Bloomfield goes onto argue that (e*) does not distinguish between w and w* as the possible world closest to the actual. It cannot distinguish between them because the consequent of the counterfactual conditional (e*) does not say which molecules

---

[27] Bloomfield thinks conversational implicature, rather than context-sensitivity, is involved because we are discussing the semantic content of a particular counterfactual and that certain interpretations of it would be ruled out by any evaluator of the counterfactual as "out of bounds". As we are trying to understand the content of (e*), the Gricean rules of interpretation alone (or so the suggestion goes) will rule out consideration of those worlds in which the water boils because the molecules are moved around by angels (or some such); we don't need to appeal to some fancy theory of contextualism to rule them out. (Bloomfield in conversation). I will ignore these complexities in what follows.

will crack the glass if molecules x, y and z do not. According to Bloomfield, then, Miller's (2009) argument does not undermine Nelson's (2006) argument as the L-L CPU will not know which possible world is closest to the actual world because he knows which proposition any utterance of (e*) expresses, since any utterance of (e*) does not distinguish between whether w or w* is the closest possible world to the actual world.

Bloomfield, however, even seems to think that knowing which possible world is closest to the actual world and knowing which molecules break the glass if molecules x, y and z do not is irrelevant to explaining why (e*) is true, since: "being able to say what the truth-conditions for a proposition are does not, by itself, entail being able to explain why those are the conditions in which the proposition comes out true." (Bloomfield, 2009, 344) Here Bloomfield endorses Nelson's reading of the counterfactual (e*) and the need to explain the general claim in order to explain why (e*) is true. Bloomfield, like Nelson, thinks that the L-L CPU will have to use program explanation in order to see why (e*) is true. To illustrate why Bloomfield thinks this, Bloomfield says that all (e*) says is simply, if molecules x, y and z do not break the glass, some other molecules will. Being more specific, the consequent of the counterfactual conditional (e*) does not specify which molecules will crack the glass, but just that as (e*) is true, it is guaranteed that in the relevant situations sufficient molecules will be actualised to crack the glass. As (e*) does not say which specific molecules will crack the glass, all (e*) says is that there is a group of possible worlds in which the water is boiling and then picks out one of these possible worlds from the group and says that this particular possible world is closest to the actual world.

According to Bloomfield in order to explain why (e*) is true, the L-L CPU will have to use program explanation to pick out a group of possible worlds where the

water is at boiling temperature. From the group selected by program explanation, we have to make sure that the closest possible world to the actual world is a member of this group.

Applying this to the moral case, although the L-L CPU is assumed to have all the process explanations of moral belief in terms of lower-level non-moral properties, modal information about how things do go in other possible worlds, according to Bloomfield's (2009) argument, the unavailability of program explanation will result in an explanatory loss. Contrary to Miller (2009), Bloomfield denies that the L-L CPU knows which possible world is closest to the actual world, since any proposition any utterance of (e*) expresses does not distinguish between (analogues of) w and w*. So with only reference to process explanations the L-L CPU cannot explain why (e*) is true. Recall that according to the Stalkner-Lewis account of counterfactuals, (e*) is true if and only if in the possible world that is closest to the actual world in which molecules x,y and z do not strike the molecular bonds at place pn at time tn with momentum mn, some other molecules strike the molecular bonds with sufficient momentum to break them. Armed with only process explanations the L-L CPU will not be able to see that the consequent is true because, as already remarked, the L-L CPU does not know the closest possible world to the actual world.

Chapter 5

In this chapter I offer an argument against Bloomfield. The argument against Bloomfield if successful will show that the fact that counterfactuals are context-sensitive undermines his and Nelson's argument that program explanation invokes distinctive information not available via an epistemically unlimited agent's grasp of a process explanation.

Recall that in the previous chapter I argued that Nelson's position of asking for an additional explanation of why the counterfactual (e*) is true, other than the explanation of how the molecules do go in the possible worlds, is unmotivated. One reason that Nelson may argue that such an additional explanation of why the counterfactual (e*) is true is required since any explanation why (e*) is true has to explain the general claim. So Nelson's and Bloomfield's argument's about the need for the general claim of (e*) has to succeed or their position is unmotivated. Furthermore, Bloomfield claims that the context-sensitivity of counterfactuals does not adequately explain why the counterfactual conditional (e*) is true. So the only explanation of why (e*) is true will have to explain *the general claim* that can only be explained by the use of program explanation. I will argue that in fact the context-sensitivity of counterfactuals do in fact undermine Nelson's and Blackburn's argument for the need of program explanation in explaining why (e*) is true.

Argument against Bloomfield

Consider a context-sensitive statement:

"I love Sophie" is true as uttered by U at time T if and only if (iff) U loves Sophie at

T

Now as the sentence "I love Sophie" currently stands, we do not know which proposition it currently expresses, so we are unable to assign it a truth-value.

To find out the proposition that the sentence expresses, we have to add in facts about the context of utterance (the contextual features) and the character to determine the content. For example, the contextual features are that the sentence is uttered by Andrew at 3pm, so that "U" refers to Andrew and "T" refers to 3pm. The character tells us to look for the speaker of the utterance, the time and hearer of the utterance, and this tells us what has to obtain in a possible world for that sentence to be true in that world. Once we have the context in addition to character, then we have the content of the statement and know the proposition the sentence expresses.[28]

Now, consider a context-sensitive counterfactual conditional:

"A $\Box\rightarrow$ B" is true iff in the closest possible world in which A is true, B is true.

Once again, as the counterfactual conditional currently stands, it expresses no proposition and thus the counterfactual conditional has no truth value.

Again we have to add the context of the utterance in addition to the character to determine the content. As we saw in our earlier example concerning stepping on

---

[28] Here I am using the David Kaplan (1978) character-content distinction, and Chapter 11 of William G. Lycan's (2000) "Philosophy of Language: a contemporary introduction."

the icy path, the relevant contextual features relate to speaker intentions, and these in turn determine what counts as the closest possible world for the purposes of evaluating the truth of the counterfactual. Once we've added the context and the character, then we get the content, and we are in a position to know which proposition the context-sensitive counterfactual conditional expresses.

So now we get:

"A $\Box\!\!\rightarrow$ B" is true iff B is true in Φ

Turning our attention back to Bloomfield's argument, Bloomfield thinks that the counterfactual conditional (e*) does not express a proposition telling us which particular molecules will break the glass if molecules x, y and z do not, but only a proposition telling us that if molecules x, y and z do not break the glass, then some other unspecified molecules will. Recall that "A $\Box\!\!\rightarrow$ B" is true iff in the closest possible world in which A is true, B is true, so in order to test the truth of a counterfactual conditional we need to know what happens in the closest possible world to the actual world in which the antecedent obtains**.** This is how Bloomfield thinks that (e*) currently stands, and as it currently stands, without the relevant contextual facts specified the counterfactual conditional expresses no proposition. To bring in the relevant contextual information required in order for a truth-value to be determined, Bloomfield thinks that we have to bring in the kinds of higher-order properties that figure in the explanandum in program explanations.

We can see this as Bloomfield says:

"So, in order to evaluate (e*), [the L-L CPU] will have to pick out the set of worlds in which

the water is boiling, that is, the set picked out by the program explanation, and then make sure that the nearest world to the actual world is a member of this set. Picking out the relevant set of worlds will require [the L-L CPU] to appeal to the property of the water that is its *boiling* and thereby program explanation." (Bloomfield, 2009, 344 [original italics])

The problem for Bloomfield is that the L-L CPU does not need to use program explanations, or appeal to the higher-order properties that figure in their explanandums, to get the relevant contextual information. The L-L CPU already has the contextual information relevant to determining the truth-value of the counterfactual conditional (e*) in virtue of the fact that the L-L CPU, as defined earlier in the paper, knows everything there is to know about the speakers communicative intentions and so on, and therefore knows the proposition that any given utterance of (e*) expresses. So the L-L CPU knows which possible world is closest to the actual world, and is therefore in a position to determine the truth-value of any given utterance of (e*) solely on the basis of his knowledge of facts about lower-level properties and the process explanations in which they appear.

Going back to Miller (2009) when he writes:

"[The L-L CPU] needn't to do anything to determine the answer to (a) [which possible world is closest to the actual world]: *he already knows the answer to (a) in virtue of the fact that he knows which proposition any given utterance of (e*) expresses*. So, a fortiori, in order to see why (e*) is true God needn't do anything that requires reference to higher-level temperature properties." (Miller, 2009, 341[emphasis added])

The L-L CPU already knows which possible world is closest to the actual world in virtue of the fact that he knows which proposition any given utterance of (e*) expresses, because the L-L CPU already has (and knows) the contextual features relevant to determining the truth-value of the counterfactual conditional (e*), such as, the speakers communicative intentions and so on. So it is the case, contrary to Bloomfield, that the L-L CPU can tell which of world w [where the water is boiling and p,q and r break the glass] and w* [where the water is boiling and f,g and h break the glass] is closest to the actual world, and thus by knowing which possible world is closest to the actual world can explain why the counterfactual conditional (e*) is true. Moreover, then, it is *not* the case that the unavailability of the relevant program explanations will result in an explanatory loss for the L-L CPU. So the program explanations that invoke moral facts and properties do not count as best for the L-L CPU and thus, they do not count as best in the right sort of way for their availability to earn ontological rights for moral facts and properties.

An additional objection to Nelson and Bloomfield is as follows. In general, if you're asked to explain why a sentence is true, you take its truth conditions for granted, and then explain why they obtain. Bloomfield and Nelson seem to require that we explain why it *means* what it means, then *also* explain why that truth-condition is satisfied. As Nelson and Bloomfield say:

"Being able to say what the truth conditions for a proposition are does not, by itself, entail being able to explain why those are the conditions in which the proposition comes out true."

This is wrong since to know whether a sentence is true you need to need to know what it means, but you don't need to know *why* it means what it means. For example,

take the sentence: "snow is white". Now to know whether the sentence is true, as long as we know what it means, we take the truth-conditions for granted and we then explain why those truth-conditions obtain: for example, the sentence "snow is white" is true because snow is in fact white. It seems absurd to ask in addition, "why it is that "snow is white" means that snow is white".

Now take the counterfactual conditional sentence (e*): "if molecules x,y and z had not struck the molecular bonds at place p_n at time t_n with momentum m_n, the glass still would have cracked, owing to the striking of other molecules, at other, or other places, or with slightly different momenta." We know what this sentence means, and we know that it is true because we can explain why those truth conditions obtain: for instance, because in the closest possible world to the actual world, we can see that the group of molecules f,g and h crack the glass if molecules x,y and z do not. This is all we need for an explanation of why the sentence is true: we do not need, in addition, an explanation of why it means what it means.

Moreover, even if this objection is incorrect and it is well motivated to ask the additional question "why is it that so-and-so means so-and-so?", then it still does not mean that program explanation is needed in order to explain why the counterfactual conditional (e*) is true. Since the L-L CPU knows all the relevant contextual features, the lower-level properties and the process explanations in which they figure, the L-L CPU will know why the sentence means what it means, even if we don't.

Lastly, it seems that Bloomfield may have misconstrued what Miller meant by an epistemically unlimited agent vis-à-vis lower level properties and the process explanations in which they figure. Bloomfield seems to think that the epistemically unlimited agent is a genuine agent, like God, rather than a heuristic device, as Bloomfield writes:

"Admittedly, these are tricky issues. At the very least, trying to know what an epistemically unconstrained agent would or would not have to know in order to explain something is fairly mind-boggling by itself. (Explain it to whom?)" (Bloomfield, 2009, 344)

This may be the reason that Bloomfield thinks that in order to bring in the relevant contextual information required in order for the counterfactual conditional (e*)'s truth-value to be determined, Bloomfield thinks that we have to bring in the kinds of higher-order properties that figure in the explanans of a program explanation.

Conclusion

Towards the end of his 2006 paper Sturgeon starts to advocate a kind of program explanation. Sturgeon writes:

"How evaluative explanations can be not just credible, but in some cases illuminating. A good causal explanation needs to describe the cause at the right level of abstraction, including relevant detail while excluding the irrelevant. That is why, even granting that slavery was unjust, it can be reasonable to ask whether the American anti-slavery movement is best described as responding to slavery's injustice or, instead, only to some specific way or ways that it was unjust. The answer presumably depends largely on how it did, or would have, responded to different forms of injustice … The answer might be that a more specific explanation is better, or it might be that the explanation that appeals to the injustice of slavery is better. And in *either* case, as we have seen, it might well be that the only terminology we have for characterizing the cause at the right level is evaluative. (Sturgeon, 2006, 251 – 2)

However, although Sturgeon thinks that "many evaluative explanations look promising in this regard", program explanation, as this thesis has demonstrated, will not confer ontological rights for moral facts and properties. (Sturgeon, 2006, 252). Despite its promise then, Cornell Realism appears incapable of yielding a plausible alternative to the metaethical views we surveyed in chapter one.

Bibliography

Ayer, A.J [1936] (1946) *Language, Truth, and Logic*, 2nd edn. London: Gollancz

Blackburn, S. (1984) *Spreading The Word*. Oxford: Oxford University Press

Bloomfield, P. (2009) "Moral realism and program explanation: a very short symposium 2: reply to Miller", *Australasian Journal of Philosophy* 87: 343-344

Boyd, R. (1988) "How to be a moral realist," in G. Sayre-McCord (ed.), *Essays on Moral Realism*. Ithaca, NY: Cornell University Press, 181 – 228

Brink, D. (1989) *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press

Darwall, S., Gibbard, A., and Railton, P. (1992) "Toward *fin de siecle* ethics: some trends." *Philosophical Review* 101, 115 – 189

Divers, J. (2002) *Possible Worlds*, London: Routledge

Dworkin, R. (1996) "Objectivity and truth: you'd better believe it." Philosophy and Public Affairs 25, 87 - 139

Frankena, W. (1938) "The naturalistic fallacy." *Mind* 48, 464 – 477

Geach, P. (1960) "Ascriptivism." *Philosophical Review* 69, 221 – 225

Geach, P. (1965) "Assertion." *Philosophical Review* 74, 449 – 465

Gibbard, A. (1990) *Wise Choices, Apt Feelings*. Oxford: Clarendon Press

Harman, G. (1977) *The Nature of Morality*. Oxford: Oxford University Press

Harman, G. (1986) "Moral explanations of natural facts – can moral claims be tested against reality." *Southern Journal of Philosophy* 24, suppl., 57 – 68

Harman, G. (1999) "Moral philosophy meets social psychology: virtue ethics and the fundamental attribution error." *Proceedings of the Aristotelian Society* 99, 315 – 331

Harman, G. (2000) "The nonexistence of character traits." *Proceedings of the Aristotelian Society* 100, 223 – 226

Jackson, F., and Pettit, P. (1995) "Moral functionalism and moral motivation." *Philosophical Quartely* 45, 20 -40

Jackson, F, and Pettit, P. (2004) "Program explanation: a general perspective", in

Jackson, F, Pettit, P, and Smith, M (ed.) *Mind, Morality and Explanation*, Oxford: Oxford University Press, 119 – 130

Kaplan, D. (1978) "Dthat", in P. Cole (ed.), *Syntax and Semantics, vol. 9: Pragmatics, New York*: Academic Press; reprinted in P.French, T. Uehling, and H.Wettstein, 221-243

Kim, J. (1984) "Concepts of supervenience." *Philosophy and Phenomenological Research* 45, 153 – 176

Kim, J. (1989) "The myth of nonreductive physicalism." *Proceedings and Addresses of the American Philosophical Association* 63, 31 – 47

Landau, R.S (2003) *Moral Realism: A Defence*. Oxford: Oxford University Press

Lowe, E.J. (1995) "The truth about counterfactuals", *Philosophical Quarterly* 45: 41-59

Lycan, W.G. (2000) *Philosophy of Language: A Contemporary Introduction*. London: Routledge

McDowell, J. (1988) *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press

McNaughton, D. (1988) *Moral Vision*. Oxford: Blackwell

Mackie, J. (1977) *Ethics: Inventing Right and Wrong*. New York: Penguin

Miller, A. (2003) *An Introduction to Contemporary Metaethics*. Cambridge: Polity

Miller, A. (2009) "Moral realism and program explanation: a very short symposium 1: reply to Nelson", *Australasian Journal of Philosophy* 87: 337 – 341

Moore, G.E. [1903] (1993) *Principia Ethica*, rev. edn. Cambridge: Cambridge University Press

Nelson, M. (2006) "Moral realism and program explanation", *Australasian Journal of Philosophy* 84: 417 - 28

Putnam, H. (1967) "Psychological predicates," in W.H Captian and D.D Merrill (eds.) *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press, 37 - 48

Railton, P. (1986a) "Moral realism." *Philosophical Review* 95, 163 - 207

Railton, P. (1986b) "Facts and values." *Philosophical Topics* 14 (2), 5 - 31

Smith, M. (1994) *The Moral Problem*. Oxford: Blackwell

Sturgeon, N. (1986) "Harman on moral explanations of natural facts." Southern Journal of Philosophy 24, suppl., 69–78

Sturgeon, N. (2006) "Moral explanations defended," in Dreier, J (ed.) *Contemporary Debates in Moral Theory*. Oxford: Blackwell, 241 – 262

Wiggins, D. (1987) "A sensible subjectivism," in *Needs, Values, Truth*. Oxford: Blackwell, 185 – 214

Wright, C. (1988a) "Moral values, projection, and secondary qualities." *Proceedings of the Aristotelian Society*, suppl. Vol., 1 -26

Wright, C. (1988b) "Realism, antirealism, irrealism, quasi-realism." *Midwest Studies in Philosophy* 12, 25 – 49

Wright, C. (1989) "Wittgenstein's rule following considerations and the central project of theoretical linguistics," in A. George (ed.) *Reflections on Chomsky*. Oxford: Blackwell, 233 – 264

Wright, C. (1992) *Truth and Objectivity*. Cambridge, MA: Harvard University Press

Contents

Can Program Explanation Confer Ontological Rights for the Cornell Realist variety of Moral Realism?

Introduction

This thesis is primarily concerned with the question as to whether program explanation can confer ontological rights for the Cornell Realist variety of moral realism? The aim of this thesis is to argue and defend the claim that program explanation *cannot* confer ontological rights for moral facts and properties. For a bulk of the time in the thesis in Chapter 4, in order to answer the question above, I will discuss and examine the papers or book excerpts of Alex Miller (2003),(2009), Mark Nelson (2006) and Paul Bloomfield (2009). Specifically, in the thesis I end up arguing against Bloomfield and Nelson and defend the arguments of Alex Miller. My argument is that program explanation is not needed to explain why a certain counterfactual conditional, which we discuss in the thesis, is true because the context-sensitivity of counterfactuals undermines the need for program explanation to explain this. Moreover, Bloomfield's argument against the fact that the context-sensitivity of counterfactuals undermines the need for program explanation to explain why the counterfactual conditional is true, does not work because the epistemically unlimited agent vis-à-vis the lower level properties and the process explanations in which they figure already knows all the contextual features needed for the context-sensitivity of counterfactuals to undermine Bloomfield and Nelson's arguments.

In Chapter 1, I present an historical overview of metaethics and examine the main metaethical positions and objections. Towards the end of the chapter I claim that Cornell Realism is not prone to the objections just discussed, and Cornell Realism

comes out best. Chapter 2 examines Cornell Realism at length and focuses on the 1980's debate between Gilbert Harman and Nicholas Sturgeon, towards the end of the chapter I argue that Harman wins the debate between the two and that Sturgeon needs an alternative reply to Harman other than the one that he gives. Chapter 3 is devoted to Sturgeon's reply to Harman that comes in the form of a program explanation; I start off by discussing what program explanation is, what its benefits are and finally how Sturgeon can use it to reply to Harman's objection[1]. Chapter 4 discusses the responses as to whether Sturgeon's use of program explanation is successful in its reply to Harman, and in doing so, confers ontological rights for moral facts and properties. In Chapter 5, I develop an objection to Bloomfield and finish the thesis with a conclusion.

---

[1] Though it should be noted that this is not a reply that Sturgeon himself makes explicitly: it is investigated on his behalf in Miller (2003).

<u>Chapter 1</u>

This chapter will present a highly selective overview of Meta-ethics in historical order. The aim of this chapter is to move from the Open Question Argument to Cornell Realism and to make plain what justifies interest in Cornell Realism through the failings of other meta-ethical theories. Hopefully, it will also trace the movement of meta-ethical theory through the failings of previous meta-ethical theories. So throughout this chapter I will make clear the position of major meta-ethical theories and the influential objections against each theory.

The meta-ethical theories and major objections that I discuss are: the Open Question Argument, Non-Naturalism, The Frege-Geach objection, Quasi-Realism, Error Theory, and then finally in the next chapter, Cornell Realism.

Cognitivism and Non- Cognitivism

Consider a typical moral judgement, such as the moral judgement that murder is wrong. *Cognitivists* think that a moral judgement expresses a belief where beliefs can either be true or false. For instance, take the belief that the tea is hot – this belief is true if and only if the tea is hot. *Non-Cognitivists*, on the other hand, think that a moral judgement expresses something like a desire or emotion. Desires and emotions cannot be either true or false. So according to the non-cognitivist, a moral judgement such as the judgement that murder is wrong is neither true nor false.

The Open Question Argument (OQA) and Moral Non-Naturalism

Meta-ethics begins with G.E. Moore. Moore in his *Principia Ethica* [1903] (1993) produces the OQA in which he argues that "good" cannot be defined in terms of natural properties, such as, pleasure or desire. By natural properties I mean properties that are either causal or detectable by the senses. By this characterization, natural properties are the properties dealt with in the natural sciences or psychology. This is called *definitional naturalism*: moral properties, by definition, are identical or can be reduced in terms of naturalistic properties. To try and define moral properties in terms of naturalistic properties commits, what Moore calls, the "naturalistic fallacy", the fallacy of trying to define "good" in naturalistic terms.

The OQA is as follows. Lets assume that the predicate "good" is synonymous (or analytically equivalent) to a naturalistic predicate N, say, pleasurable. So it is part of the meaning of "x is pleasurable" that "x is good." So if someone asked if that which is pleasurable is also good? they would betray conceptual confusion; they would appear not to understand the concept that the predicate "pleasurable" and the predicate "good" are synonymous (or analytically equivalent). Just in the same way, if someone asked is an "unmarried man" a "bachelor"? we would think that they didn't understand the concept that a "bachelor" is synonymous (or analytically equivalent) to "unmarried man." Now this next step is why the argument is called the "open question argument." However, it is always a *significant* question whether something that is pleasurable is good (*unlike* whether asking someone who is unmarried is a bachelor), thus someone asking this question actually betrays no conceptual confusion. So it is not true that the predicate "pleasurable" is synonymous or analytically equivalent to the predicate "good."

The OQA though is not as strong as it seems, at least in its original formulation. The main objection and problem of the OQA was offered by William Frankena (1938), as Frankena says, quite concisely:

"[T]he charge of committing the naturalistic fallacy can be made, if it all, only as a conclusion from the discussion and not as an instrument of deciding it." (Frankena, 1938, 465).

Basically, Frankena's objection is that the OQA begs the question against the moral definitional naturalist. Moore can only appeal to the OQ if the conviction that the question "is an x which is N, good?" is open is well substantiated. The problem is that the OQ is only well substantiated if it is the case that moral and naturalistic predicates are not synonymous or analytically equivalent. However, the definitional naturalist will say that that conviction that the question is open is *not* well substantiated. So, according to the moral definitional naturalist, the question "is an x which is N, good?" is a closed question, and Moore's argument against naturalism cannot run.

Since it roughly took 35 years – Moore's OQA in the *Principia Ethica* [1903] and Frankena's objection (1938) – for a convincing response against the OQA on behalf of the moral definitional naturalist, as this chapter is working historically through the metaethical views and movements, the OQA will be assumed successful against the moral definitional naturalist for our purposes[2].

Moore's favoured brand of cognitivism was moral non-naturalism. Non-naturalism argues that moral properties are non-natural, irreducible and sui generis[3].

---

[2] The two main contemporary versions of moral reductionism are: Analytic Moral Functionalism and its leading proponents are Frank Jackson and Philip Pettit (1995), and Peter Railton's Moral Reductionism (1986a), (1986b). I will not have space to discuss these here.

[3] Contemporary versions of non-naturalism are advocated by David Wiggins (1987), John McDowell (1988) and more recently Russ Shafer Landau (2003). Again, I cannot discuss these here.

Moore's non-naturalism claims that moral judgements are truth apt: moral judgements are either true or false. According to Moore what makes a moral judgement true is the instantiation of a non-natural, unanalysable and simple property of moral goodness. Moore thinks, then, that we access this non-natural, unanalysable, simple property via the operation of a special faculty of "intuition" (hence Moore's position is sometimes called "intuitionism"). So take the moral judgement that giving to the poor is good, it is true because of the instantiation of a simple, sui generis property moral goodness by acts of giving to the poor.

The main problem for non-naturalism that makes it an unattractive position is that it has no plausible epistemological account as to how we cognitively access moral properties. If by "intuition", as we will assume, Moore means a cognitive faculty that works as a kind of perception, similar to sense-perception, then such a view makes no sense. It amounts to a form of perception that apparently is not, ex hypothesi, causal and this is difficult to understand. Moreover, non-naturalism is viewed within widespread suspicion amongst philosophers and later in this chapter in the section "error theory" we will see that John Mackie expounds on the same general problem of intuitionism.

Assuming for the time being the success of the OQA and the standard rejection of non-naturalism the prospects of cognitivism look bleak; moral properties are neither identical nor reducible to naturalistic properties or non-natural properties themselves. For these various reasons many philosophers turned towards trying to find a viable non-cognitivist theory. It is now that we look at Quasi-Realism.

Quasi-Realism (QR)

Before the in-depth discussion of Quasi Realism, I will focus briefly on Emotivism and its main objection. The main objection is the Frege-Geach objection which threatens not only Emotivism but all non-cognitivist theories. I am dealing with Emotivism first because Quasi Realism was set up to try and solve the objections that defeated Emotivism, and Emotivism and Quasi Realism both share the same underlying expressive semantics.

A.J. Ayer's Emotivism (1936) denies that moral judgements express propositions. According to Emotivism, in the standard moral judgement that murder is wrong, I am not saying that murder is wrong or anything else, but rather I am expressing my disapproval. Just as I would cry !%*$ in anguish when I burn my hand: crying !%*$ does not express the proposition that I am in pain, it is an expression of anguish.

The central objection to emotivism and non-cognitivism is the Frege-Geach objection (Geach 1960; 1965). The problem for the non-cognitivist is how to account for the appearance of a moral judgement in an unasserted context in a way that does not render intuitively valid arguments guilty of a fallacy of equivocation. Take an intuitively valid moral modus ponens argument. For instance,

(1) Murder is wrong

(2) If murder is wrong, then getting your little brother to murder is wrong

(3) Therefore, getting your little brother to murder is wrong.

In order for the argument to be valid the occurrence of "murder is wrong" in (1) has to have the same semantic function as "murder is wrong" occurs in (2). The argument is *not* valid if semantic function in (1) is different than how it occurs in (2). This is the equivocation fallacy.

The equivocation fallacy is committed when the same word or phrase is used in an inference but each token of the word or phrase has a different meaning. For instance,

(4) My clock has hands

(5) If something has a hand, then it must have fingers and a thumb

(6) Therefore, my clock has fingers and a thumb.

As you can see here the semantic function of the word "hands" within (4) is different from how it occurs within (5). In (4) "hands" means (usually) black instruments that point to numbers on a clock, but in (5) "hands" means a part of the human anatomy.

Now the problem for the emotivist and non-cognitivism is that it renders intuitively valid arguments guilty of equivocation. To see this, take the valid moral modus ponens argument but with an underlying expressive semantics:

(1) Murder is wrong

(2) If murder is wrong, then getting your little brother to murder is wrong

(3) Therefore, getting your little brother to murder is wrong.

The semantic function of "murder is wrong" as it occurs within an *asserted* context in (1) is different from its semantic function as it occurs within an *unasserted* context in

(2). This commits the fallacy of equivocation and the argument is not valid unless the semantic function is the same as it occurs within (1) and (2). In (1) "murder is wrong" is asserted to express disapproval of murder, but in (2) "if murder is wrong, then getting your little brother to murder is wrong," "murder is wrong" is being used as an antecedent of the conditional. It is *not* being used to express disapproval of murder. The argument thus appears to turn out invalid.

Quasi-Realism is developed in order to defuse this worry**.** What is QR? As Simon Blackburn puts it:

"Quasi-Realism is the enterprise of explaining why our discourse has the shape it does, in particular by way of treating evaluative predicates like others, if projectivism is true. It thus seeks to explain, and justify, the realistic seeming nature of our talk of evaluations – the way we think we think we can be wrong about them, that there is a truth to be found, and so on." (Blackburn, 1984, 180).

It is useful to note that there are two parts to the QR enterprise: *modest QR* and *ambitious QR*. Blackburn's above quote concerns modest QR. Modest QR attempts to explain why we talk as if it is true (or at least truth-apt) that, for example, murder is wrong and giving money to the poor is good even though it is assumed that such moral predicates do not refer to actual moral properties. That is, even though we judge and express the belief that murder is wrong, moral predicates do not refer to actual moral properties, but rather serve to express feelings and sentiments. So for instance, when we say that "murder is wrong" we are talking as if we are expressing the proposition that murder is wrong, even though we are expressing no proposition at all but actually an expression of disapproval.

How then does such a position work? Lets take a closer look at modest QR. The QR believes that there is a distinction between the *surface form* of a region of discourse and the *deep form* of a region of discourse. The surface form of a discourse is usually suggested by the syntax. The surface form of a moral discourse is *cognitive* or *propositional*: "murder is wrong" and "giving money to the poor is right", for example are *declarative sentences*, suggesting that moral sentences represent states of affairs, and the *predicates* "wrong," "good" and "permissible", for example suggest that predicates denote properties. Furthermore, "Mark believes that murder is wrong" and "Patricia believes that giving money to the poor is good" are syntactically well-formed, suggesting that moral judgements express beliefs. However, the projectivist wants to deny that the surface form of a moral discourse is an accurate guide to its deep form, although moral statements appear propositional or cognitive on the surface, their essential role is expressive. (Miller, 2003, 60)

Ambitious QR, on the other hand, argues that moral judgements, such as, the judgement that murder is wrong or the judgement that giving money to the poor is right, have truth conditions, but this is done solely on a projectivist basis. So it is either true or false that murder is wrong or giving money to the poor is right, but a moral truth is constructed by our expressions of attitudes and moral sensibilities. The shift from modest QR to ambitious QR is the shift from talking "as if" there is moral truth when really there is not, to allowing that there is such a thing as moral truth but viewing it as constructed solely on a projectivist basis[4].

Since QR was designed to meet the problems that emotivism faced, how then does Blackburn try and find a solution on behalf of the QR against the Frege-Geach

---

[4] Since I'm concentrating on modest QR due to its application to the Frege-Geach problem, the discussion of ambitous QR is far more brief and is intended merely to show what the other part of the QR enterprise is.

problem**?** In what follows I borrow heavily from Miller's (2003) exposition on Blackburn's attempted solution to the Frege-Geach problem.

Blackburn's solution consists of three sections. Firstly, consider a simpler connective than "if … then", such as "and". A conjunction is true only when both of its conjuncts are true, and false otherwise. We use "and" to join commitments: for instance "murder is wrong and giving money to the poor is good." What can a projectivist say about such a conjunction since he does not want to assess the conjunction in terms of truth and falsity? Blackburn answers:

"[we should] expand the way we think of "and". We have to do this anyway, for it can link utterances when they certainly do not express beliefs which are genuinely susceptible of truth-value e.g. commands: "hump that barge and tote that bale." We would instead say something like this: "and" links commitments to give an overall commitment which is accepted only if each component is accepted." (Blackburn, 1984, 191-192)

So a projectivist account of "murder is wrong and giving money to the poor is good", is as follows: this conjunctive sentence serves to express my disapproval of murder and my approval of giving money to the poor.

Before I move on, secondly, I will quickly explain the notion of a moral sensibility. A moral sensibility is a disposition to respond to various situations with different attitudes. For instance, we may be pleased by selflessness, inspired by ambition, repulsed by megalomania and indolent about dictatorships. The complete set of dispositions is a person's moral sensibility. We can also take attitudes towards a person's moral sensibility: for example, we can regard it as sensitive or insensitive, fickle or stubborn and admirable or abhorrent. (McNaughton, 1988, 183)

Projectivists can do something similar for conditionals, such as, "if murder is wrong, then getting your little brother to murder is wrong." In uttering a conditional like the one we are concerned with you are expressing an attitude towards a moral sensibility. So in the conditional "if murder is wrong, then getting your little brother to murder is wrong" I am expressing my attitude of approval towards moral sensibilities which combine disapproval of murder with disapproval of getting little brother to murder people.

Finally, in order to see how Blackburn's attempted solution works consider a language in which its surface form is expressive (call such a language Eex) instead of its surface being propositional.

"It might contain a "hooray!" operator and a "boo!" operator (H!, B!) which attach to descriptions of things to result in expressions of attitude. H!(the playing of Tottenham Hotspur) would express the attitude towards the playing, B!(lying) would express the contrary attitude towards lying, and so on". (Blackburn, 1984, 193)

In order to talk about an attitude of approval or disapproval we use square brackets. So [H!(the drinking of a cup of tea)] refers to the sentiment of approval of drinking a cup of tea. Moreover, to indicate two attitudes together we place a semi-colon between them: so [[H!(drinking of a cup of tea]); [B!(eating crisps)]] refers to the joining of attitude of approval towards the drinking of cups of tea *with* the attitude of disapproval towards the eating of crisps.

Now what would the conditional (2) in the original argument look like in Eex? The projectivist interprets (2) in terms of the expression of approval of moral sensibilities which combine disapproval of murder with disapproval of getting little brother to murder. In Eex (2) would be represented as: H![[B!(murder);[B!(getting

little brother to murder)]]. So the argument of (1), (2) and (3) comes out as:

(1ex) B!(murder)

(2ex) H![[B!(murder);[B!(getting little brother to murder)]].

- Therefore –

(3ex) B!(getting little brother to murder)

The argument then appears to be valid. Somebody who was committed to the premises but who rejected the conclusion would have, as Blackburn puts it, a "clash of attitudes". He would fail to have the combination of disapproval towards murder and disapproval towards getting little brother to murder, whilst approving of that combination.

However, Blackburn's attempted solution is not successful. Crispin Wright raises the following worry:

"Anything worth calling the validity of an inference has to reside in the inconsistency of accepting its premises but denying its conclusion. Blackburn does indeed speak of the "clash of attitudes" involved in endorsing the premises of the modus ponens example, construed as he construes it, but in failing to endorse the conclusion. But nothing worth regarding as inconsistency seems to be involved. Those who do that merely fail to have every combination of attitudes of which they themselves approve. That is a *moral* failing, not a logical one. (Wright, 1988, 33 [original italics])

In essence Wright's problem is that the "clash of attitudes" shows a moral failing not a logical failing. And for the validity of the inference to be accepted, being committed to the premises and not the conclusion must be a logical failing.

Despite Blackburn's attempts the Frege-Geach problem is a strong and powerful objection against QR and non-cognitivist theories. There has yet to be a convincing solution to the problem and the worry that non-cognitivism cannot account for the validity of a valid modus ponens argument still persists; if we cannot find a convincing solution to the Frege-Geach objection, then all the worse for QR and other non-cognitivist positions[5].

Error theories are supposed to be an alternative to both Moorean non-naturalism and non-cognitivism so I am now going to move onto John Mackie's Error-Theory.

Error Theory

John Mackie's *Error-Theory* (1977) argues that moral judgements express beliefs, but that atomic positive moral judgements are systematically and uniformly false. So, for instance, when we assert the moral judgement that murder is wrong this moral judgement comes out as false: it is not true that murder is wrong.

How then does Mackie argue for such a radical claim? By the conjunction of a conceptual claim with an ontological claim. I'll discuss the former in detail first and then move onto the latter. Mackie's conceptual claim is that our concept of a moral requirement is a concept of an *objectively categorical prescriptive* fact. To say that moral requirements are *prescriptive* is to say that it tells us how we ought to act, or

---

[5] For instance, Allan Gibbard's *Norm Expressivism* (1990) which is the theory that moral judgements express our acceptance of norms.

that it gives us reasons to act. For instance, if something is morally good, the prescriptive requirement of moral properties entails, it is something that we ought to act towards, or have reasons to act towards. And conversely, if something is morally bad it is something that we ought *not* to act towards, or have reasons *not* to act towards. Moreover, to say that moral judgements are *categorical* is to say that the reasons for actions are not dependent on our desires, wills or inclinations. If something is morally good, then irrespective of our desires or inclinations we ought to act towards it. Or if something is morally bad, no matter our desires, wills, inclinations, then we ought *not* to act towards it. *"Objective"* here means - for simplicity's sake for nothing will turn on it in this discussion, as Mackie says a lot of different things – is a property that is independent of our choices and preferences, and it is part of the fabric of the world.

The ontological claim is simply that there are no categorical objectively prescriptive facts in the world; there are no moral facts in the world. Both of these claims together (the conceptual and ontological) entail a cognitivist error theory view of morals[6].

Mackie's main argument for the ontological claim is the argument from queerness. Mackie has two reasons as to why there exist no moral properties, which must fit the objective and categorically prescriptive requirement: metaphysical and epistemological. I will look at each one in turn respectively.

Firstly, the metaphysical argument: if such moral properties were to exist they would be metaphysically queer, moral properties would be utterly and completely different to the properties that we are familiar with in the universe as described in science and common-sense. As Mackie explains:

---

[6] Which, obviously, due to the ontological claim is not a version of moral realism.

"An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. Similarly, if there were objective principles of right and wrong, any wrong (possible) course of action would have not-to-be-doneness somehow built into it." (Mackie, 1977, 40)

And, according to Mackie, no instantiation of moral state of affairs exist in the world. Secondly, the epistemological argument (which expounds on the problem of "intuitionism"), backs up the metaphysical argument:

" If we were aware [of objective values], it would have to be by some special faculty of moral perception or intuition, utterly different from our ways of knowing everything else. These points were recognised by Moore when he spoke of non-natural qualities, and by the intuitionists in their talk about a "faculty of moral intuition." Intuitionism has long been out of favour, and is indeed easy to point out its implausibilities. What is not so often stressed, but is more important, is that the central thesis of intuitionism is one to which any objectivist view of values is in the end committed: intuitionism merely makes unpalatably plain what other forms of objectivism wrap up." (Mackie, 1977, 38)

The only way in which we can come into contact into such metaphysically queer moral state of affairs is to include unexplanatory and unpalatable conceptions of how we come into cognitive content with such moral state of affairs, such as that exhibited by moral intuitionism

Wright's objection is that error theory's story about the point of moral discourse is inherently unstable given its moral scepticism about moral truth. Wright

and Mackie both realize that the error-theorist need a subsidiary norm which statements of moral discourse aim at and can satisfy besides moral truth. (Wright, 1996, 2) The point of moral discourse, according to Mackie, is to secure the benefits of social co-operation[7]. (Mackie, 1977, chapter 5) Suppose there's a plausible account of the subsidiary norm taken from the story of moral discourse – to secure benefits of social co-operation – that governs the practice of forming moral judgements, then moral judgements will be aimed at satisfying this subsidiary norm, whatever it is. Not all moral judgements will satisfy the subsidiary norm in equal manner, for example, the moral judgement "murder is good" will frustrate the subsidiary norm, whilst the moral judgement "murder is bad" will facilitate the subsidiary norm. The problem that Wright raises for Mackie's error theory concerns whether it can plausibly combine a story about the benefits of practicing moral judgements with the central negative claim of the error theory. Wright thinks not:

" [I]f, among the welter of falsehoods which we enunciate in moral discourse, there is a good distinction to be drawn between those which are acceptable in the light of some subsidiary norm and those which are not – a distinction which actually informs ordinary discussion and criticism of moral claims – then why insist on construing *truth* for moral discourse in terms which motivate a charge of global error, rather than explicate it in terms of the satisfaction of the putative subsidiary norm, whatever it is. The question may have a good answer. The error-theorist may be able to argue that the superstition that he finds in ordinary moral thought goes too deep to permit any construction of moral truth which avoids it to be acceptable as an account of *moral* truth. But I do not know of promising argument in that direction." (Wright, 1996, 3 [original italics])

---

[7] This is a simplification, but the details need not concern us here, nothing turns on it.

Wright's problem for Mackie is that if there is such a thing as a subsidiary norm, then why not construct moral truth in terms of the satisfaction of the subsidiary norm, instead of constructing truth for moral discourse in terms of that motivates a global error-theory? According to Wright there is no plausible reason or good argument to do so, in which case, Mackie might as we construe moral truth in terms of the satisfaction of the subsidiary norm which will enable him to bypass the global error-theory of the practice of moral judgements.

Another problem for error-theory is that opposing metaethical views can simply reject the metaphysical aspect of Mackie's argument, denying his formulation of moral properties is the correct version. To reject the metaphysical aspect of Mackie's argument from queerness, we need to find a plausible account of moral properties that does not see moral properties as being categorical objective prescriptive requirements[8] [9].

Cornell Realism

Finally we come to Cornell Realism. In the next chapter I will discuss Cornell Realism more fully, but for now let's see how Cornell Realism manoeuvres past the objections and problems that beset the other metaethical theories.

Cornell Realism is a cognitivist theory that argues that moral facts and properties are irreducible (non-reductive) sui generis natural properties. By claiming that moral properties are irreducible properties this enables Cornell Realism to

---

[8] One option is Crispin Wright's Judgement Dependent Account (1988; 1989; 1992) which argues "that our best judgement about morals determine the extension of moral predicates" (Miller, 2003, 6) This is a dispositionalist account of moral values.

[9] Throughout this paper I am quiet on whether certain meta-ethical views can offer a plausible epistemological account of morals to diffuse the epistemological aspect of Mackie's argument from queerness.

sidestep the OQA and the naturalistic fallacy because Cornell Realism does not claim that moral predicates are analytically equivalent or synonymous to other natural predicates. Also by claiming that moral properties are natural properties the Cornell Realist avoids being caught up in moral non-naturalism and the bizarre faculty of "intuitionism" in cognitively accessing moral facts or properties. The fact that Cornell Realism is a cognitivist theory means obviously that it does not face the same problem as the non-cognitivists. The Cornell Realist escapes the Frege-Geach problem because they can explain the semantic function of "murder is wrong" in an unasserted context without making intuitively valid argument guilty of equivocation. Finally, Cornell Realism does not fall into a global error theory since it does not accuse moral judgements of being systematically and uniformly false. Cornell Realism is a type of moral realism in that it allows that moral judgements are objectively true. However, is Cornell Realism as attractive as it initially appears? The rest of this paper is devoted to looking at one recently discussed aspect of this question.

Chapter 2

In this chapter we turn to Cornell Realism and discuss whether it is a viable meta-ethical theory. Starting off by discussing the Cornell Realist position in depth, I move onto an influential objection by Gilbert Harman against the Cornell Realist. I will then give an overview of the debate discussing Nicholas Sturgeon's replies and counter replies[10] to Harman's objections and counter-objections. This chapter will provide background for the next chapter when I discuss a recent argument in response to Harman's objection, which is the central topic of this paper

Detailed account of Cornell Realism

According to Cornell Realism moral properties are natural but irreducible properties. Moral properties supervene on other natural (non-moral) properties, but moral properties themselves are irreducible and cannot be reduced to a particular or complex set of non-moral natural properties. Firstly, in this discussion of Cornell Realism, I will solely discuss what it means for moral properties to supervene on natural properties. Cornell Realists prefer a strong supervenience claim, which is as Jaegwon Kim explains:

"A strongly supervenes on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and necessarily if any y has G, it has F." (Kim, 1984, 165)

---

[10] Throughout this chapter we mainly concerned with Sturgeon's 1980's presentations.

To say that moral properties strongly supervene on non-moral natural properties is so to say that if an object has a moral property, such as being good, then there is some non-moral natural property, say being N, such that, necessarily, if some object has N, then that object is good. For example, suppose that giving aid to the hungry is good. Then according to strong supervenience, acts of that type have some non-moral natural property N such that necessarily, any act which has N is also good. Moreover, this means that in any possible world, suppose someone gives aid to the poor, then acts of this type have the property N, and if some object has N then that object is good.

Strong supervenience is usually distinguished from weak supervenience: to say that moral properties *weakly* supervene on non-moral natural properties is only to say that within any particular possible world if two objects have the same non-moral properties, then they have the same moral properties. For example, suppose that two actions are instances of giving aid to the hungry and are identical in all other natural respects. Then, within the possible world in question, either both are good or neither are.

Strong supervenience is consistent with multiple realization: a good action A might be good in virtue of having N_1, while another good action B might be good in virtue of having N_2. Note that multiple realization is usually held to rule out reduction.

Later in the paper the terms "higher-level property" and "lower-level property" will show up. I'm not going to give a precise definition of what higher and lower level properties are since the best way to understand them is through illustration.

In any supervenience relation between the base properties and the supervening property, the base properties are the lower-level properties and the supervening

property is the higher-level property. For instance, the moral property supervening onto the non-moral base properties is the higher-level property and the non-moral natural properties are the lower-level properties. To take another example, chemical properties are widely held to supervene onto microphysical properties, in this case the supervening property, the higher-level property, is the chemical properties and the base properties, the lower-level properties, are the microphysical properties. And as a last example, to show, for instance, that the same property will not always be the higher-level supervening property, take the supervenience relationship between biological and chemical properties. The biological properties supervene onto the chemical base properties, where the biological properties, in this case, are the higher-level property and, in this case, the chemical properties are the lower-level properties.

Now I will approach the second part of the Cornell Realist argument: that moral properties are *irreducible* (or non-reductive) to a particular or complex set of non-moral natural properties.

Cornell Realists hold the view that moral properties can be multiple realized by other natural properties. Contrary to moral reductionists, Cornell Realists hold that moral properties are not reducible to other natural properties. For instance, to take an archaic reductionists position, Cornell Realists deny that "good" is analytically equivalent or synonymous to "pleasure".

So what does multiple realizability mean? To help explain this I will first illustrate the famous example of multiple realizability in philosophy of mind. Take a mental type, pain (for example): many mind-brain identity theorists held that pain was reducible to c-fibre stimulation. Just as, analogously, moral reductionists hold that a moral kind was reducible to a particular or complex non-moral natural kinds.

But as Hilary Putnam (1967) pointed out: many life-forms – terrestrial, extraterrestrial or robotic – could feel (or eventually could be known to feel) pain. Now, for this to be compatible with the mind-brain identity theorist, then they would have to argue that the particular or complex physical kind was common to all the life-forms that could feel pain. This seems absurd because the mental kind, pain, could plausibly be realized by many *distinct* neurophysiological types, such as, green slime in aliens, silicon in cyborgs, or an electronical state in a supercomputer. Multiple realizability about the mental, then, is that given a mental kind, pain, this can be realized by many distinct neurophysiological types: c-fibre stimulation in humans, green slime in aliens and silicon in cyborgs etc. So, it is not just organisms that have c-fibre stimulation that can feel pain, being in pain is a state which is multiply realizable by neurophysiological types and cannot be reduced to a specific neurophysiological type. Regardless of the philosophy of mind example above, multiple realizability is commonly held to block reduction separately[11]

Analogously Cornell Realists block reduction because they argue moral kinds, wrongness (for example), can be multiply realized by distinct non-moral natural kinds. For example wrongness can be realized by unprovoked physical violence, stealing money, unjustifiably mocking someone, deceiving and lying etc, but in each of these examples, wrongness, cannot be reduced to a particular or complex non-moral natural kind. This is still consistent with strong supervenience, as Cornell Realists are prepared to say that necessarily, anything else which is an instance of unprovoked physical violence, for example, is wrong.

---

[11] Not all philosophers accept this view, for example, a notable and influential philosopher, such as, Jaegwon Kim. Kim, briefly, believes that all multiple realizablity does is make reduction "messy" (cf. Jaegwon Kim (1989) "The myth of nonreductive materialism.")

How, then, do the Cornell Realists justify the existence of irreducible natural moral properties:

"[The Cornell Realists] have pursued analogies with natural and social science to argue that moral properties might be both irreducible and explanatorily efficacious. One might, for example, argue that various chemical or biological "natural kinds" – acid, catalyst, gene, organism – are not obviously type reducible to the natural kinds of physics, and yet play a good role in good scientific explanation." (Darwall, Gibbard and Railton, 1992, 169-170)

So, as we can see from the above quote, analogously the Cornell Realists argue that although moral properties are not type reducible to other natural properties, moral properties still play a distinctive role in the best moral explanations.

So the Cornell Realist, then, argues as follows:

(1) "P is a real property if and only if P figures ineliminably in the best explanation of experience.

(2) Moral properties figure ineliminably in the best explanation of experience. Therefore:

(3) Moral properties are real properties." (Miller, 2003, 140-141)

The Cornell Realist argues that moral properties and facts can justifiably be included in our ontology of the natural world in the same way that physical, chemical, biological, psychological and sociological properties and facts are, in that they pull their weight in our best overall explanatory picture of the world. Just as we can explain why a physicist forms the belief that there is a proton in the cloud chamber in

virtue of their being a proton in the cloud chamber, moral properties can contribute to the explanation of moral belief.

Harman's objection to Cornell Realism

Gilbert Harman (1977) argues that whilst physical properties figure ineliminably in the best explanation of experience, as demonstrated by the case of the physicist's belief in the proton, he denies that moral properties figure ineliminably in the best explanation of experience. Harman's motivation for the argument is to show how moral explanations are problematic in a way that scientific explanations are not.

To see why consider this example. Godzilla kidnaps a young maiden, takes the maiden on top of the empire state building (the maiden is scared of heights) and eats her, and Anu in seeing this immediately forms the belief that what Godzilla is doing is wrong. But do we need to cite wrongness as explanatorily relevant as to why Anu thinks what Godzilla is doing is wrong? According to Harman, we do not: we only need to cite the natural non-moral facts, such as, the unwanted kidnapping, the extreme pain, the suffering the maiden experiences, and facts about our psychology (specifically our character and upbringing).[12]

The difference between the moral and the physical case is as follows. In the physical case we have to assume the existence of the proton to explain why the physicist forms the belief that a proton is in the cloud chamber. In the moral case, however, we do not need to assume the existence of wrongness to explain why Anu forms the belief that Godzilla's action of kidnapping and eating the maiden is wrong.

---

[12] In other works, Harman (1999; 2000) denies that there is such a thing as character! This is an odd fact since he believes, above, that you can explain why something is wrong, in part, due to our character.

As Harman explains:

"You need to make assumptions about certain physical facts to explain the occurrence of the observations that support a scientific theory, but you do not seem to need to make assumptions about any moral facts to explain the occurrence of the so-called moral observations I have been talking about. In the moral case, it would seem that you need only make assumptions about the psychology or moral sensibility of the person making the moral observation[13]." (Harman, 1977, 6)

So according to Harman, we can deny premise (2) of the Cornell Realist argument by arguing that: (2a) Moral "properties" do not figure ineliminably in the best explanation of experience. By dropping reference to moral facts in the explanation of moral beliefs we suffer *no* explanatory loss, the unavailability of moral facts does *not* entail explanatory impoverishment. The use of moral facts in explanation does not, in the relevant sense, count as best, and so we do not earn ontological rights for moral "properties."

Sturgeon's reply to Harman

Nicholas Sturgeon (1986) argues, in defence of Cornell Realism[14], that wrongness is explanatorily relevant in the forming of Anu's belief that what Godzilla was doing is wrong. Sturgeon's reply is that if wrongness passes the *counterfactual test*, a way to test claims about explanatory relevance, then wrongness *is* explanatorily relevant to Anu forming her belief

---

[13] "moral observations" here are really just moral beliefs
[14] Other leading proponents of the Cornell Realist position that I will not discuss in this paper are Richard Boyd (1988) and David Brink (1989)

The counterfactual test is as follows. If (if a had not been F, then b would not have been G) then a's being F is explanatorily relevant to b's being G. To illustrate this, consider this example: Godzilla's ferocious roar is explanatorily relevant to the town's citizens screaming in terror because if Godzilla had not ferociously roared, the town's citizens would not have screamed in terror.

So the pertinent question now is: - when we apply the counterfactual test to Godzilla and the maiden – if Godzilla had not done something wrong, then Anu would not have believed that Godzilla did something wrong? Sturgeon points out firstly that when reading this counterfactual conditional that we have a normative moral theory that conveys the relationship between non-moral facts and moral facts. According to this normative moral theory, an act that possesses the non-moral properties of unwarranted kidnapping and suffering needless extreme pain, or some other non-moral properties upon which moral properties supervene, also possess the moral property of wrongness. Suppose this normative moral theory is correct, then for the action not to be wrong, it would have to lack the non-moral properties of needless extreme pain and unwarranted kidnapping, or some other non-moral properties upon which moral properties supervene according to our correct normative moral theory. Now it is incorrect to assert that Anu would have still thought that Godzilla's actions were wrong even though the act lacked the non-moral properties of needless extreme pain and unwarranted kidnapping and so on. So, holding the assumption that our normative moral theory is correct, due to the counterfactual test, it seems true that the wrongness of the action is explanatorily relevant to Anu's belief that Godzilla's actions were wrong.

Sturgeon is well aware that moral sceptics will deny that he is entitled to assume a correct (or roughly correct) normative moral theory, but according to

Sturgeon this move of denying the normative moral theory does not help Harman. For, if Harman denies the assumption that the normative moral theory is correct, then we can derive a sceptical conclusion about the explanatory relevance about *physical properties*. Consider the case of the physicist and the proton in the cloud chamber. The physicist has a theory about what phenomena signal the existence and presence of the proton, namely, the vapour trail in the cloud chamber. Suppose this theory is correct, if there hadn't been a proton present there wouldn't have been a vapour trail, the physicist would not have formed the belief that a proton was present. So by this counterfactual test, assuming the truth of the physical theory, the proton is relevant to the physicist's belief that a proton was present. This is exactly the same as the moral case and the normative moral theory.

Now suppose that the physicist's theory is *incorrect*, that it is false that a vapour trail signifies the presence of a proton. That is, even without the presence of a proton there still would have been a vapour trail and due to the physicist's belief in a mistaken physical theory the physicist would have still believed that a proton was present even though it was not. So, on the assumption that the physicist's theory is incorrect we can derive the conclusion that if a vapour trail is present, then the physicist would believe that a proton was present when in fact it is not. If we assume that the physicist's theory is mistaken, then the presence of the proton is explanatorily irrelevant, by the counterfactual test, to the explanation of the physicist's belief that a proton is present[15].

---

[15] If some cannot get their head around the fact that a modern physical theory could be mistaken, there are two additional points that can be made. Firstly, an example of a now defunct scientific theory can illustrate this, such as, phlogiston. Not going comprehensively into phlogiston theory, scientists saw certain phenomena as signals that phlogiston was present. But since phlogiston theory is infact incorrect and phlogiston does not exist, the scientists who believed in phlogiston theory would still have believed that phlogiston was present due to certain signalling phenomena, when in fact it was not. Secondly, scientific hypotheses' are continually being up-dated and shown to be false, before a new scientific theory and hypothesis takes its place.

Harman has the option of "biting the bullet" and accepting that science and morality are on a par**.** So by "biting the bullet" Harman would be showing that moral and scientific theory are problematic in the same way, i.e. that they can be used to derive sceptical conclusions. However, Harman's motivation for his argument was to show that moral theory is problematic in ways that scientific theory is not. So instead of "biting the bullet" Harman needs a counter reply.

Harman's counter reply

Harman (1986) denies Sturgeon's claim that wrongness is explanatorily relevant to the formation of Anu's belief that Godzilla's action of kidnapping and eating the maiden is wrong. The truth of the counterfactual conditional that appears in the counterfactual test is *not* sufficient to yield genuine explanatory relevance. As Harman says:

"What is needed is some account of how the actual wrongness of [Godzilla's] action could help to explain [Anu's] disapproval of it. And we have to be able to believe in this account. We cannot just make something up, saying, for example, that the wrongness of the act affects the quality of the light reflected into [Anu's] eyes, causing her to react negatively. That would be an example of wrongness manifesting itself in the world in a way that could serve as evidence for and against moral claims, but it is not something we can believe in." (Harman, 1986, 63)

Harman's thought is that the counterfactual test is not sufficient for genuine explanatory relevance because for wrongness to be explanatorily relevant what we need is an account of the mechanisms that underlie such counterfactual dependence.

That is, we need to know what makes moral properties causally efficacious in the production of our experience and it cannot be the idea that "wrongness of the act affect the quality of light reflected into Anu's eyes" as Harman, above suggested. Or that moral protons, "morons" as Ronald Dworkin (1996) calls them, emanate from wrong actions, such as Godzilla kidnapping a maiden and eating her, and affect the observer in such a way that they end up thinking that what Godzilla did was wrong. Since Sturgeon cannot use the above examples and has no alternative plausible story about how irreducible moral properties play a genuine explanatory role in people's moral beliefs and judgements or how wrongness is causally efficacious to the production of our experience, it is acceptable to conclude that moral properties do not play a genuine role in explaining the formation of moral beliefs.

Harman re-affirms this point when he argues that the counterfactual dependence of Sturgeon's purported moral properties cannot justify the explanatory relevance of moral properties because a moral *epiphenomenalist*- someone who accepts the explanatory impotence of moral properties – can accept that the relevant counterfactual conditional is true. The moral epiphenomenalist accepts that moral properties supervene onto the non-moral properties (for instance, that wrongness supervenes onto unprovoked kidnapping, torture, abuse etc.), but that moral properties have no explanatory relevance themselves. That is, the counterfactual dependence of moral belief upon moral properties is consistent with the denial that moral properties are explanatorily relevant. So Sturgeon's claim that moral properties are explanatorily relevant in explaining why Anu formed her belief that what Godzilla was doing is wrong is unconvincing. The non-moral properties of the Godzilla situation were causally efficacious in the production of Anu's belief that what Godzilla was doing is wrong, but wrongness itself was not causally efficacious in producing Anu to form

her belief that what Godzilla was doing was wrong. So although a moral epiphenomenalist can accept that the relevant counterfactual conditional is true, this does not show that moral properties are genuinely explanatorily relevant in production of Anu forming her belief that what Godzilla was doing is wrong.

Sturgeon's response against Harman's counter-reply

Sturgeon (1986) agrees that a moral epiphenomenalist would pass the counterfactual test, but argues that we need to find an independently plausible reason to accept moral epiphenomenalism, otherwise moral epiphenomenalism does not undermine the intuition we have that moral properties are explanatorily relevant. The intuition remains because all supervening facts would be epiphenomenal since the only lower-level properties that have a causally efficacious mechanism are microphysical properties. The properties of the special sciences (chemical, biological, psychological, and social properties) will all be rendered epiphenomenal since they supervene onto the microphysical properties and are causally inefficacious and explanatorily impotent. According to Sturgeon this is implausible.

Furthermore, Sturgeon (2006) claims that this epiphenomenalist argument against the supervenience thesis proves too much. Physicalists take psychological properties to supervene on physical properties, that the base physical properties are sufficient for psychological states, for example, pain. However, and importantly, for Sturgeon, they resist the conclusion that psychological states are epiphenomenal or that psychological states are causally inefficacious in the production of our

experience[16]. Why should we do so purely in the moral case and not in the other cases? Bringing both responses together, Sturgeon is arguing that although moral epiphenomenalism is a logical possibility since it passes the counterfactual test, by itself this does not damage the claim that moral properties are explanatorily relevant. Epiphenomenalism is not accepted in other areas of philosophy; we need an independent reason to accept moral epiphenomenalism in this case.

Miller (2003) finds Sturgeon's reply here to Harman implausible for another reason. Miller points out that Sturgeon's original claim was that the counterfactual dependence of "b's being G upon a's being F is *sufficient* to justify the idea that the property F is explanatorily relevant to b's being G: this is what allows him to argue that moral properties are explanatorily relevant." (Miller, 2003, 149 [original italics]) So the fact that moral epiphenomenalism is logically possible does undermine Sturgeon's sufficiency claim: the counterfactual test is *not* sufficient to show the explanatory relevance of moral properties because moral epiphenomenalism asserts that moral properties are explanatorily impotent at the same time as accepting the claim about counterfactual dependence. So, this means that Sturgeon needs to find some alternative way of responding to Harman. In the next chapter we will look at program explanation as a way of responding to Harman on behalf of Sturgeon.

---

[16] This is just one example out of many that Sturgeon can run – for instance, he could use the example of chemical properties supervening on physical properties, or social properties supervening on psychological properties amongst others. In those examples, the supervening property will not be seen as epiphenomenal or as causally inefficacious to the production of our experience.

Chapter 3

In this chapter I will suggest a possible solution for Sturgeon in response to Harman's objection by the use of program explanation (this is derived from Miller 2003). Firstly, the need for program explanation will be motivated by examining the role of properties in causal explanations, which eventually yields a counter-intuitive conclusion that can only be assuaged by the use of program explanation. Secondly, what exactly program explanation is and the importance of program explanation will be discussed by showing that it imparts information not imparted by the relevant process explanations. Finally, it will be shown how and why program explanation can be used as an attempted solution to Harman's objection, showing that a program explanation can be run to show that the unavailability of moral properties will lead to an explanatory impoverishment.

Program explanation

To explain what program explanation is we start off with Frank Jackson and Philip Pettit. In Jackson and Pettit (2004) they identify three common and plausible background assumptions about the role of properties in causal explanations. They then suggest that if we add a fourth assumption we get a problem. The notion of program explanation emerges via the rejection of the fourth assumption.

(1) A causal explanation of something must direct us to a *causally relevant* property as opposed to a *causally irrelevant* property of the factor it identifies as explanatory: a property relevant to the causal production of the effect explained.

For example: a causal explanation of why a keycard unlocked a secured door must direct us to the property of having a suitable electromagnetic strip rather than the colour of the keycard, the electromagnetic strip is relevant but the colour is not.

(2) One way in which properties are causally relevant is by being causally efficacious. A causally efficacious property in virtue of whose instantiation, at least in part, the effect occurs; the instance of the property helps to produce the effect and does so because it is an instance of that property.

One way to be causally relevant is to be causally efficacious. For example: the property of having a suitable electromagnetic strip is causally efficacious because it is in virtue of the keycard's having the strip that the door unlocked. The intuition is that if the property is causally efficacious the effect occurs, at least in part, because of the instantiation of that property (in which case the property is obviously causally relevant).

(3) A property F is not causally efficacious in the production of an effect e if these three conditions are fulfilled together.

Intuitively, if conditions (3i) (ii) and (iii) are satisfied, then property F is not causally efficacious in the causal production of the effect occurring**,** since all the causal work will be done by the relevant G property.

(i) there is a distinct property G such that F is efficacious in the production of e only if G is efficacious in its production;

(ii)   the F-instance does not help to produce the G-instance in the sense in which the G-instance, if G is efficacious, helps to produce e; they are not sequential causal factors

(iii)   the F-instance does not combine with the G-instance, directly or via further effects, to help in the same sense to produce e (nor of course, vice versa): they are not coordinate causal factors. (Jackson and Pettit, 2004, 120)

We now add a fourth assumption:

(4) The only way for a property to be causally relevant to the production of a certain effect is by being causally efficacious in the process of production (Jackson and Pettit, 2004, 124).

Taken together these four assumptions are incompatible with strong intuitions about the role of certain properties in explanation. To illustrate this point, consider the following example[17]: suppose that we boil water in a closed glass container, and the water reaches such a temperature that the glass cracks. What explains why the glass cracks? Intuitively, we say things like "that glass cracked because of the temperature," or "that glass cracked because the water was boiling," or something similar along these lines. The temperature of the water or the fact that the water was boiling is causally relevant to the cracking of the glass.

However, as Jackson and Pettit explain:

---

[17] The following example is from Jackson and Pettit (2004)

"Why did it crack? First answer: because of the temperature of the water. Second answer, in simplified form: because of the momentum of such and such a molecule (group of molecules) in striking such and such a molecular bond in the container surface. The temperature property was efficacious only if the momentum property was efficacious: hence 3(i). But the temperature of the water – an aggregate statistic[18] – did not help to produce the momentum of the molecule in the way in which it, if efficacious, helped to produce the cracking: hence 3(ii). And neither did the temperature combine with the momentum to help in the same sense to produce the cracking: one could have predicted the cracking just from full information about the molecule and the relevant laws. Hence 3(iii). (Jackson and Pettit, 2004, 110)

So according to assumption (3), the temperature of the water is not causally efficacious in the cracking of the glass, and from assumption (4) the temperature of the water is not causally relevant to the cracking of the glass. This conclusion is extremely counter-intuitive.

The example of temperature and the cracking of the glass feeds into a more general point about assumption (4). Assumption (4) has two devastating consequences for causal explanations from common sense and the special sciences (i.e. chemistry, biology, psychology and sociology.) Firstly, it means all true, proper and real causal explanations of something are to be found at the level of fundamental physics. Since the ultimate base properties of the supervenience relation are microphysical properties[19] [20], and all the special sciences ultimately supervene on microphysical

---

[18] By aggregate statistic here, take it to mean something like average momentum of constituent water molecules.

[19] In Jackson and Pettit's paper they assume that the causally efficacious properties are the physical (or microphysical) properties. It is neither implicit or explicit in the text whether this takes into account quantumn mechanics. Whether it does or not isn't a worry for program explanation since all it requires is that you believe there is a lower-order level at which there are causally efficacious properties in the production of the relevant effect occurring.

[20] What happens if the lower-level property mentioned is not itself the efficacious property in the causal production of the effect? For instance, what if the lower-level property mentioned is the microphysical

properties, - and given assumption (3) about the role of properties in causal explanations - then in all causal explanations of something the microphysical properties are the causally efficacious properties in the causal production of an effect. Quite clearly, if a causal explanation of something can only cite the properties of physics, then we must deny that the special sciences and common sense are able to give explanations in terms of causally relevant properties. Causal explanations from the special sciences and common sense cite causally inefficacious and thus causally irrelevant properties in causal explanations.

Secondly, a consequence of the first point means that if we have a higher-order explanation and a lower-order explanation of the same phenomena, then we must replace the higher-order explanation with the lower-order explanation. So, for example, if we ask why did the glass break? The causal explanation we should prefer is not in terms of the higher-order property of temperature, but rather in terms of lower order momentum properties of certain molecules because the lower-level explanation directs us to causally efficacious properties in the production of the relevant effect[21]. Intuitively such a view is unattractive, it goes against common sense.

Jackson and Pettit's solution is to reject assumption (4) by arguing that properties can be causally relevant without being causally efficacious. How so? A

---

properties but they are not causally efficacious because actually quantum mechanics is? Then, as Jackson and Pettit point out, "the story will go a level or more deeper until we find an efficacious property for which the original higher-order property programs indirectly, via the programming of the intermediate features." (Jackson and Pettit, 2004, 120)

[21] An interesting consequence of this, as Jackson and Pettit explain, is that it is not enough to cite (for example) that some (indeterminate) molecules are striking a molecular bond because "explanation involving existential quantification – the reference to an indeterminate some – cannot be a proper explanation: it does not invoke an efficacious property and so does not invoke a property that is relevant." (2004, 124-125) For lower-order properties to be relevant in causal explanations, the lower-order properties cited must be the exact causally efficacious properties. Most of the time we do not know the exact causally efficacious properties in a causal explanation of something, so some of us may go through our lives never offering a relevant causal explanation of anything.

property can be causally relevant without being causally efficacious because the instantiation of a certain property (higher-level properties) ensures the instantiation of properties (distinct lower-level properties) that are causally efficacious in the causal explanation. Although the higher-level property in question is causally inefficacious and causally inert, its realization *programs* for the instantiation of lower-level properties that are causally efficacious in the causal production of the effect occurring. In this sense the higher level property is causally relevant in causal explanations. This is known as program explanation. As Jackson and Pettit explain it, it is called *program* explanation because of the *programming* metaphor:

"A useful metaphor for describing the role of the property is to say that its realization programs for the appearance of the productive property and, under a certain description, for the event produced. The analogy is with a computer program which ensures that certain things will happen – things satisfying certain descriptions – though all the work of producing those things goes on at a lower, mechanical level". (Jackson and Pettit, 2004, 127)

So, for example, using program explanation, in the context of the temperature of the water and the glass cracking case, the temperature of water is in fact causally relevant without being causally efficacious. The realization of the boiling temperature (a higher-level property) programs for the instantiation of such-and-such momentum properties of the molecules (the lower-level properties).

So by the use of program explanation the temperature of the water is causally relevant to the cracking of the glass. But what reason is there to choose (in some cases) the higher-level explanation instead of the lower-level explanation? That is, if we have lower-level explanations available should we not just ditch the higher-level explanation in preference to the lower-level explanations every time?

No, there are times when it is preferable to use a higher-level explanation because a program explanation "may provide information which the corresponding process explanation [lower-level explanation] does not supply. Thus, it may be an explanation which the process explanation does not supersede." (Jackson and Pettit, 2004, 129)

There are two ways in which the higher-level explanation, the program explanation, can provide information not known at the corresponding process explanation in the context of the temperature of the water and the glass cracking case. Firstly, it is possible that a molecule can strike the container with sufficient momentum at the correct place and time *without* the water being at boiling temperature. So the explanation in terms of the boiling temperature conveys information that the corresponding process explanation doesn't (namely, that the water is at boiling temperature). Secondly, which is an elaboration of the first point, program explanation imparts modal information that is not conveyed by the corresponding process explanation. (Jackson and Pettit, 2004, 129) So to have the explanation in terms of the water's being boiling tells us that in any relevantly similar possible world, that there will be molecules with sufficient momentum, place and time to crack the glass container. That is, to know that the water is boiling means that in the actual world it was these molecules (for example, molecules x,y and z) that crack the glass, but had *those* molecules not cracked the glass, then in other relevantly similar possible worlds other molecules will crack the glass (for example, molecules f, g and h).

Contrast these higher-level explanations, program explanations, with lower-level explanations, process explanations. The process explanations only tells us which momentum properties of the molecules are causally efficacious in cracking the glass,

but it tells us nothing about whether the water is at boiling temperature or not. Moreover, the process explanations only tell us about the *actual world* and the original situation itself, without telling is what would happen in the *relevantly similar possible worlds*.

As should be evident now, clearly program explanation fits into strong supervenience and multiple realizability. Temperature properties strongly supervene on the momentum properties of groups of molecules. If a sample of water is boiling, then there is some momentum property of its constituent molecules such that, necessarily, any sample of water whose constituent molecules have the same momentum properties is also boiling.

The supervenience relation in this case is not reductive**:** for any given temperature property there are going to be an indefinite number of ways in which it can be realized at the level of momentum properties of constituent molecules.

Now the obvious link between program explanation, strong supervenience and multiple realizability becomes clear. *Whenever we have strong supervenience and multiple realization we have the ingredients for informative program explanation.* The program explanation, in the case of the boiling water and cracking glass, selects one combination out of the multitude possible combinations of the momentum properties of the molecules to be causally efficacious in the causal production of the glass cracking. If it had not been that particular combination selected, it would have been another possible combination of the momentum property of the molecules that were causally efficacious in the causal production of the glass cracking, and this is only possible as the boiling water can be multiply realized by an almost infinite number of possible combinations of the molecules, momentum, time and place that are causally efficacious in the causal production of the glass cracking. This can only

happen though because the relationship between the momentum property of the molecules and the boiling water holds as it is one of strong supervenience, enabling the boiling water to be able to program for the instantiation of one out of a multitude of possible combinations of the momentum property of the molecules that are causally efficacious in the glass cracking.

A final note on program explanation: as program explanation conveys information not available at the level of the corresponding process explanations, the unavailability of the program explanation would constitute an explanatory loss, so program explanation can count as best.

Program explanation, Sturgeon and Harman

The important question now is: can Sturgeon respond to Harman's objection? It is not enough for Sturgeon to argue that moral properties are causally efficacious. That, say, when a wrong act has been committed morons emanate from a wrong action and affect the observer into thinking such actions are morally wrong. For instance, Anu observes Godzilla kidnapping and eating a maiden, and from Godzilla's actions emanate some "morons" that are causally efficacious in some way to make Anu believe that what Godzilla was doing is wrong. This is not a plausible story of how moral properties play a role in explaining people's moral beliefs and judgements.

Miller (2003) has argued that Sturgeon can respond to Harman by arguing that although moral properties are not causally efficacious in the production of moral beliefs, by the use of program explanation moral properties are causally relevant.

Moral properties can "program" for the instantiation of the sufficient non-moral properties that are causally efficacious in the production of moral beliefs[22]?

To see this, consider Harman's example of Godzilla and Anu. Anu sees Godzilla kidnap and eat a maiden and forms the belief that what Godzilla is doing is wrong. We can give an explanation of Anu forming her belief by citing the non-moral properties, such as kidnapping the maiden, the pain of being eaten, and these non-moral properties are causally efficacious in producing Anu's belief that what Godzilla is doing is wrong.

We can say that the explanation of Anu's forming that belief that Godzilla's act is wrong in terms of its wrongness is a program explanation. The wrongness of Godzilla's act is causally inefficacious in the production of Anu forming the belief that what Godzilla is doing is wrong. Godzilla's act had certain natural features, such as kidnapping the maiden and putting her into his mouth to eat her, which made it true that Godzilla's act was wrong and which were causally efficacious in producing Anu's belief that what she saw was wrong. So the wrongness of Godzilla's act is only efficacious in Anu's belief if the non-moral properties are causally efficacious, hence (3)(i). The wrongness and the non-moral natural properties are neither sequential nor causal co-ordinate factors in the production of Anu's belief, hence (3)(ii) and (iii). So the moral property is inefficacious. However**,** the (higher-level property) wrongness ensures the instantiation of the (lower-level properties) non-moral properties that are causally efficacious in the production of Anu forming the belief that what Godzilla was doing is wrong. So we can give a program explanation of Anu's forming the relevant moral belief.

---

[22] Although Miller offers the argument on behalf of the Cornell Realist, Miller ultimately denies that the argument is successful as a response to Harman's objection. For Miller's own words check (2003, 150 – 174) I will discuss this fully in the next chapter anyway.

The program explanation counts as best because it imparts information that is not available in the corresponding lower-level explanations in two ways. Firstly, kidnapping the maiden and eating her can be causally efficacious in producing Anu to form the belief that what has just seen is wrong, even though it is not: we can think of a far-fetched scenario in which the only way to save the maiden from an even worse fate is by eating her alive. Secondly, program explanation imparts modal information that is not found at the corresponding lower-level explanation: consider another relevantly similar possible world in which we substitute the non-moral properties of eating her to throwing her off the empire state building to her death, the non-moral properties would of still have been causally efficacious in the production of Jane forming the belief that what Godzilla did was wrong.

So, we can give a program explanation of the formation of Anu's belief in terms of the wrongness of Godzilla's actions. The unavailability of program explanation would lead to an explanatory impoverishment, impoverishment of the relevant information imparted by the program explanation. (Miller, 2003, 153 - 154) So the program explanation that cites the moral features of things can, in the relevant sense, count as best. So, moral properties are a real properties because, due to premise (1) of Cornell realism, "P is a real property if and only if P figures ineliminably in the best explanation of experience." And so we can earn ontological rights for moral properties without bringing in morons.

Chapter 4

To recap, the issue at hand is whether the use of moral program explanation will earn ontological rights for moral properties. To do so, moral program explanations must figure ineliminably in the best explanation of experience and it must also be the case that the unavailability of moral program explanations will result in an explanatory loss. In this chapter I look at four arguments ((Miller (2003), Nelson (2006), Miller (2009), Bloomfield (2009)) that are split into alternating sections of arguing that program explanation does *not* earn ontological rights for moral properties and that program explanation does earn ontological rights for moral properties. It so turns out that the arguments are in historical order, so it will be possible to follow the historical order of the discussion of whether program explanation earns ontological rights for moral properties for the Cornell realist variety of moral realism.

Before I discuss the selected four arguments, it is important just to note that the case of the temperature, water and the cracking of the glass container is used as a model for the application of program explanation to the moral case in all the arguments discussed.

Does program explanation successfully defend Sturgeon against Harman's objection?

Alex Miller (2003) does not think that program explanation can earn ontological rights for moral properties. Program explanation figures ineliminably in the best explanation of our experience because program explanations being ""best" – such that the lack thereof [of program explanation] results in an explanatory loss – is a direct consequence of some *epistemic limitation* of ours." (Miller, 2003, 173 [original

italics]) So, according to Miller, the unavailability of program explanation leads to an explanatory loss due to our epistemic limitations: once these limitations are lifted, in the way they have to be when we are considering the question of ontological rights, the unavailability of program explanation no longer results in an explanatory loss.

What does Miller mean by an epistemic limitation? In his initial attempt at the argument in Miller (2003) he is not explicit about what an epistemic limitation is. In Miller (2009) he is: by epistemic limitation Miller means epistemically limited vis-à-vis lower-level properties and the process explanations in which they figure[23]. Call an agent that is epistemically unlimited "the lower-level computer" (the L-L CPU), the most perfectly designed and built computer in the universe. It is *vitally important* to note that the introduction of the L-L CPU is intended merely as a heuristic device rather than as the introduction of an actual agent[24].

Recall that program explanations conveyed information not available at the level of the corresponding process explanations via the grasp of an epistemically limited agent in two ways. By conveying the additional information: (i) that the water is at boiling temperature, and (ii) that the glass would crack in relevantly similar possible worlds. Miller's argument is that program explanation conveys distinctive information, information of types (i) and (ii), not conveyed in process explanations because of our epistemic limitations. Thus, if we had no epistemic limitations,

---

[23] Henceforth within the main body of the text, I will simply refer to the epistemically unlimited agent vis-a-vis lower-level properties and the process explanations in which they figur as an epistemically unlimited agent. However, I will sometimes use the full title of the epistemic agent to highlight its function as an heuristic device. Likewise, "epistemic limitations" should always be read as 'epistemic limitations vis a vis the lower facts and process explanations in which they figure".

[24] I have changed Miller's example of God being an epistemically unlimited agent vis-à-vis lower-level properties and the process explanations in which they figure as a heuristic device due to the potential ambiguity with the omniscient agent, God. Hopefully by using the "L-L CPU" terminology it will be clearer that the epistemically unlimited agent vis-à-vis lower-level properties and the process explanations in which they figure is simply a heuristic device. Furthermore, and lastly, in all examples and quotes in which God has been used as a heuristic device I have replaced God with the L-L CPU

program explanations would convey the same distinctive information conveyed in the process explanations.

Remember that program explanations can only confer ontological rights for moral properties if information of type (i) and (ii) is not conveyed in the corresponding process explanations. The relevant question is: would a program explanation convey distinctive information of types (i) and (ii) to an agent epistemically unlimited vis-à-vis lower-level properties and the process explanations in which they figure? I will investigate Miller's responses to (i) and (ii) in turn.

Miller's response to (i) is that the distinctive information conveyed by program explanation of type (i) does not confer ontological rights for moral properties because "(i) is actually not relevant in the cases where we are concerned with whether program explanation can confer ontological rights." He goes onto say, "[i]f we *assume* that there is a fact about the temperature of the molecules, knowledge of which would be lost given the unavailability of a program explanation, then we are already assuming what we set out to prove: namely, that there are higher-order facts and properties and properties of the relevant kind." (Miller, 2003, 173 [original italics]). So according to Miller the issue purely concerns the capacity of program explanations to convey information of type (ii).

Miller's response to (ii) is that when considering ontological rights for higher-level properties we are wondering whether the higher-level properties would feature in the perspective an epistemically unlimited agent, the L-L CPU. So in the perspective of the L-L CPU is the program explanation still "best" in that its unavailability it would result in an explanatory loss?

Miller argues that in the perspective of the L-L CPU the unavailability of the program explanations would *not* result in an explanatory loss. As Miller says:

"Given that [the L-L CPU] knows all of the facts about the relevant process explanations and the modal information about how things would go in relevantly similar possible worlds, it is clear that he would not. How, then, could program explanation earn ontological rights for higher-order properties?" (Miller, 2003, 174)

Simply put, Miller is arguing that by removing the shackles of epistemic limitation by looking at the situation from the perspective of an agent who knows all there is to know about lower-level properties, the process explanations in which they figure, and how things would go in different possible worlds, we can see that no additional distinctive information is conveyed by the use of program explanations. So there is no need to add higher-level properties into our ontology.

Miller points out that the same applies to higher-level moral properties:

"Likewise, if we assume that [the L-L CPU] knows all of the naturalistic explanations of moral beliefs, as well as modal information about how things would go in relevantly similar possible worlds it is clear that the unavailability of program explanations would not result in his suffering an explanatory loss. Thus, program explanations which invoke moral facts and properties do not count as best for [the L-L CPU]; thus, they do not count as best in the right sort of way for their availability to earn ontological rights for the higher-order properties in which they trade." (Miller, 2003, 173-174)

Miller's argument, then, is as follows:

"(M1) Higher-level properties are causally relevant in a way that can earn them ontological rights only if they figure in the best explanation of experience, considered

from the point of view of a subject who suffers no epistemic limitations *vis-à-vis*

lower-level properties and the process explanations in which they appear.

(M2) It is not the case that higher-level properties figure in the best explanation of

experience, considered from the point of view of a subject who suffers no epistemic

limitations *vis-à-vis* lower-level properties and the process explanations in which they

appear

Therefore,

(M3) It is not the case that higher-level properties are causally relevant in a way that

can earn them ontological rights." (Miller, 2009, 337 [original italics])

Mark Nelson (2006) denies (M2) by arguing: *it is the case* that higher-level

properties figure in the best explanation of experience, considered from the point of

view of a subject who suffers no epistemic limitations vis-à-vis lower-level properties

and the process explanations in which they appear[25]. So, according to Nelson, the

unavailability of program explanation would lead to an explanatory impoverishment

even from the perspective of the L-L CPU.

---

[25] Interestingly in his paper Nelson (2006, 423) states his second premise as "[i]t is not the case that higher-level properties figure in the best explanation of phenomena, considered from the point of view of a subject who suffers no epistemic limitations." Compare this to Miller's second premise in the text above in which Nelson makes no reference to the agent being epistemically unlimited with reference to the lower-level properties. This is important because it seems that Nelson *may have* misinterpreted what Miler meant by **"God"**. Miller does not mean God as an agent *per se* as an omnipotent agent, but a heuristic device of an epistemically unlimited agent in regards to the lower-level properties and the process explanations in which they figure. Such a view is plausibly backed up by the fact that Miller's argument is saying with the shackles of epistemic limitation taken off, program explanations convey no additional information not already conveyed by process explanations. Moreover, such a view is backed up by Miller in (2009) emphasises the **"vis-à-vis"** rendering of his premises in the argument. However, whilst saying all of this, Miller is probably to blame for using the term "God" without adequate explanation in the original presentation of the argument.

Nelson agrees with Miller on two points, in context of the example of the boiling water cracking the glass. Firstly, that the relevant process explanations explain why effect *e*, the cracking of the glass occurs. Secondly, that the relevant process explanations also say how things do go in particular possible worlds: for example, in one particular relevantly similar possible world, molecules f, g, and h crack the glass. But according to Nelson the fact that the relevant process explanations explain what goes on in particular possible worlds it is not enough to explain why the counterfactual (e*) is true.

Counterfactual (e*):

(e*)"if molecules x, y and z had not struck the molecular bonds at place p_n at time t_n with momentum m_n, the glass still would have cracked, due to the striking of other molecules, at other times, or other places, or with slightly different momenta." (Nelson, 2006, 424)

As Nelson explains why:

"But none of this is sufficient to explain why (e*) is true … [the lower-level properties] about the microphysical state and the history of the world right up until the occurrence of e, and of all the relevant causal laws can't explain it; they can explain only how subsequent history did in fact turn out. *The true counterfactuals can't explain it, either. They say only how things do go in particular possible worlds, but they don't explain why it is true that some one of a limited range of possible worlds, out of limitless panoply of worlds that are possible will be actualized. Miller stipulates that [the L-L CPU], our epistemically unlimited subject, knows that (e*) is true, but it one thing to know that a proposition is true; it is another thing to have an explanation for it.*" (Nelson, 2006, 424 – 425 [emphasis added])

Nelson denies that knowing how the relevant process explanation go in particular possible worlds explains why (e*) is true. According to Nelson, knowing which exact molecules crack the glass in the relevant particular possible worlds is not the same as explaining the *general claim* as to why the glass cracks in relevantly similar possible worlds. We can see this as Nelson (2006, 425) says "Miller stipulates that [the L-L CPU] our epistemically unlimited subject [vis-à-vis the lower level properties and the process explanations in which they figure] knows that (e*) is true, but it is one thing to know that a proposition is true; it is another thing to have an explanation for it." And Nelson then goes onto say "Moreover, even if these materials enabled [the L-L CPU] to see that a particular possible world *would* be actualised in an alternative process-scenario to [the striking of molecules x,y, and z on the molecular bonds at place p_n at time t_n with momentum m_n] (because, for example, he could that molecules u,v and w were "next in queue to crack the glass," so to speak), this would not be an *explanation* of the general claim (e*), that *some* such world would be actualised. This is so because *ex hypothesi* (e*) still would have been true if still *other* molecules had been next in queue, so the particular claim about u,v, and w cannot be the explanation of the general claim." (ibid.)

What does Nelson mean by the general claim (e*)? Nelson seems to mean that if molecules x,y and z do not crack the glass, then there will always be molecules with sufficient momenta, position and time to crack the glass in other relevantly similar possible worlds, as the general claim (e*) guarantees this. Nelson does not think it is enough to cite the relevant process explanations in the possible worlds in order to explain why (e*) is true because (e*) would still be true had *other* molecules cracked

the glass in the relevantly similar possible worlds, and thus the true counterfactuals cannot be an explanation of the general claim.

Nelson thinks that only program explanation can explain why the general claim (e*) is true and that the other relevantly similar possible worlds where the glass cracks will be actualised if molecules x, y and z do not strike the molecular bonds with sufficient momentum etc. Recall that the glass of water being at boiling temperature is multiply realizable, it does not matter which molecules strike the molecular bonds, all that matters is that the molecules that do strike the molecular bond have sufficient kinetic energy to break it. The fact that the water is at boiling temperature means that some molecules are doing enough of what they have to do to crack the glass, the boiling temperature of water guarantees that other molecules will be available with sufficient momenta, at the relevant time and place to crack the glass, if molecules x, y, and z do not.

Notice that both sides of the argumentative debate both make the jump between the fact that sufficient molecules will do enough to crack the glass in the actual world and the fact that other sufficient molecules will do enough to crack the glass in other possible worlds. For the sake of argument both sides assume that this is true, in other words, both sides assume that counterfactual (e*) is true.

Nelson's argument then is that the program explanation conveys distinctive information not available via an epistemically unlimited agent's grasp of a process explanation because only program explanation can be invoked in order to explain why (e*) is true. So, applying this to the moral case, the unavailability of the relevant program explanations will result in an explanatory loss for the L-L CPU. So program explanations that invoke moral facts and properties do count as best for the L-L CPU,

and thus count as best in the right sort of way for their availability to earn ontological rights for moral facts and properties.

It is interesting to note what has happened now. Previously, program explanation featured ineliminably in the best explanation of our experience due to our being epistemically limited since it offered modal information, information of type (ii), not available via an epistemically limited agent's grasp of a process explanation. Now, with an epistemically unlimited agent, according to Nelson, program explanation is still "best" because it explains *why* the counterfactual (e*) is true, even though an epistemically unlimited agent of the relevant sort would know that the counterfactual is true. In attempting to deny this line of thought the relevant question to ask is whether process explanations explain why the counterfactual (e*) is true? But before we look at Miller's (2009) response which denies Nelson's line of thought by arguing that process explanations can explain why the counterfactual (e*) is true, lets assess Nelson's argument.

Nelson's argument is unconvincing. To see why take the most important section of Nelson's argument:

"The true counterfactuals can't explain it, either. They say only how things do go in particular possible worlds, but they don't explain why it is true that some one of a limited range of possible worlds, out of limitless panoply of worlds that are possible, will be actualized."

As already discussed, Nelson is arguing that the true counterfactuals show that a limited range of possible worlds are actualised where the water cracks the glass out of a unlimited panoply of possible worlds, but the true counterfactuals do not explain *why* a limited range of possible worlds will be actualised where the water cracks the glass.

The problem for Nelson is that this requirement of asking for an explanation of why the counterfactual (e*) is true, other than by knowing that the glass will crack in other possible worlds, is unmotivated. When Nelson asks for an explanation of *e*, the cracking of the glass in the actual world, Nelson feels it is sufficient to cite only the molecules involved in the process explanations. As Nelson says:

"If [the L-L CPU] traced the whole chain of process explanations of G1 [(the striking of molecules x, y, z on the molecular bonds at place p_n at time t_n with momentum m_n)] back to the beginning of time, he would need to refer only to G-type explanations, never to F-type explanations [(the water's having reached boiling temperature.)]" (Nelson, 2006, 424)

Now, clearly when we move onto possible worlds, Nelson agrees that the L-L CPU knows the relevant modal information and how things do go in possible worlds. The L-L CPU can trace the chain of process explanations and molecules back to the beginning of time. As Nelson, once again, says,

"[The L-L CPU] is left with knowledge of the microphysical state and history of the world right up until the occurrence of e, knowledge of all the relevant causal laws, and knowledge of the relevant true counterfactuals … The true counterfactuals can't explain it, either. They [the true counterfactuals] say only how things *do* go in particular possible worlds." (ibid. [original italics])

If we are following the blueprint of what counts as an explanation *why* in the actual world e, the cracking of the glass in the actual world occurs, then this should count as an explanation *why* the limited range of possible worlds will be actualised, namely, how the molecules and process explanations do go in the relevant particular possible

worlds. Nelson's additional constraint, then, as to what is required to explain why the counterfactual (e*) is true seems unmotivated.

Why then does Nelson hold this additional constraint? The only explanation available is his reading of what the counterfactual (e*) entails. Nelson's reading of the counterfactual (e*) entails that any explanation of why (e*) is true has to explain the general claim and this can not be explained by reference only to the true counterfactuals, as Nelson has already explained. However, if Nelson's reading of the counterfactual (e*) is wrong, and thus the explanation of the general claim is not needed to explain why (e*) is true, then his additional constraint is unmotivated.

Miller's reply to Nelson

Miller (2009) offers a reply to Nelson (2006) arguing that the L-L CPU would *not* need program explanation to know why the counterfactual (e*) is true. Miller is using the fact that counterfactuals are context-sensitive in order to undermine Nelson's argument.

It is widely acknowledged and accepted that counterfactuals are context sensitive in the sense that "which propositions an utterance of a counterfactual expresses is determined by the context of an utterance." (Miller, 2009, 338–339) As John Divers explains.

"What is special about counterfactual conditionals is that, even for different tokens of the same sentence type, the set of worlds that is relevant to their truth is something that varies with context of utterance." (Divers, 2002, 11)

And Jonathon Lowe:

"Quite generally, a sentence (type) is "context-sensitive" if the proposition expressed by (a token of) it is partly determined not merely by the standard meanings of its constituent words but also by its circumstances of utterance." (Lowe, 1995, 52)

"A speaker's intentions in asserting a counterfactual can help to determine the propositional content of his assertion by fixing an appropriate measure of similarity across possible worlds for the proper evaluation of the truth or falsity of what he asserts." (ibid. 55)

To illustrate what Divers and Lowe mean by counterfactuals being context-sensitive, consider this example.

Suppose two people Michael and David are outside on a freezing cold winter's day and Michael and David know that the cycle path, which is situated near them, is extremely icy and really slippery. From a short distance away Michael and David can see their friend Charlie, who is an extremely cautious individual. If Charlie wants to walk directly closer towards Michael and David, then he will step onto the cycle path.

Consider:

(1) "If Charlie walks any further, then he will fall over" as uttered by Michael who wants to convey information about what happens when someone walks on really slippery ice.

(2) "If Charlie walks any further, then he will *not* fall over" as uttered by David who wants to convey information about how cautious Charlie is when it comes to walking on unsuitable terrain.

The counterfactual's context-sensitivity shows that (1) and (2) can be true at the same time. (1) conveys information about what would happen to Charlie if he walked onto the cycle path due to the really slippery nature of the cycle path, (2) conveys information about Charlie's cautious character, as he would find another non-slippery route to walk.

How do we test the truth conditions of counterfactuals? "A counterfactual conditional is true if and only if in the closest possible world in which its antecedent is true its consequent is also true[26]." (Miller, 2009, 339) To evaluate Michael's utterance (1), the closest possible world in which Charlie steps onto the cycle path is one where his psychology is different. To evaluate David's utterance (2), the closest possible world is one where Charlie's psychology remains the same so that he walks further towards Michael and David only *after* he has found an alternative non-slippery route to the cycle path. Michael's utterance is true since in the relevant world Charlie falls over. David's utterance is true since in the relevant world Charlie does *not* fall over.

How, then, does the fact that counterfactuals are context-sensitive undermine Nelson's objection? According to the Stalkner-Lewis account (e*) is true ""if and only if in the closest possible world to the actual world in which molecules x, y, and z do not strike the molecular bonds at place p_n at time t_n with momentum m_n, some other molecules strike molecular bonds with sufficient momentum to break them." (Miller, 2009, 340) So in order to explain why (e*) is true all the L-L CPU would need to know, according to Miller, is which possible world is closest to the actual world and what goes on at the level of groups of molecules in that world. The important question is: would the L-L CPU have to call upon higher-level temperature properties in order

---

[26] This isn't just a test, its also an account of the truth-conditions of counterfactuals. Moreover, to keep things simple we are assuming the Stalnaker-Lewis account of counterfactuals

to see which possible world is closest to the actual world and thereby to see why (e*) is true?

Miller argues that we do not have to call upon higher-level temperature properties as the L-L CPU (as Miller and Nelson both agree) knows all the lower level facts (in this case, all of the facts that can be characterised without using temperature concepts) about the universe, "*including the facts about which proposition any given utterance of (e*) expresses.*" (Miller, 2009, 341 [original italics]) The L-L CPU already knows which is the closest possible world to the actual world since he knows which proposition any given utterance of (e*) expresses. So, clearly the L-L CPU does not need to call upon higher-level temperature properties to know *why* (e*) is true.

Recall earlier I remarked that Nelson's reply to Miller was unconvincing because Nelson accepted it was sufficient to cite facts about the momentum properties of molecules since the beginning of time to explain why *e* occurs, the cracking of the glass in the actual world. Yet Nelson denies that facts about the momentum properties of the molecules since the beginning of time, combined with the true counterfactuals, is sufficient in explaining why (e*) is true, and this I argued was unmotivated. This point feeds into Miller's argument here in a way that undermines Nelson's argument: by knowing which proposition any utterance of (e*) expresses and by knowing which possible world is closest to the actual world then we can see that knowledge of process explanations at the level of groups of molecules is sufficient to explain why (e*) is true. That is, the L-L CPU can see why (e*) is true in virtue of information at the level of process information in the closest possible world to the actual world.

Applying this to the moral case, as the L-L CPU is assumed to have all the process explanations of moral belief in terms of lower-level non-moral properties and

modal information about how things go in other possible worlds, following Miller's (2009) argument, it is clear that the unavailability of program explanation for the L-L CPU would *not* result in an explanatory loss. As the L-L CPU knows which proposition any given utterance of (e*) expresses, and thus knows the closest possible world to the actual world, the L-L CPU will know why (e*) is true.

Bloomfield reply to Miller

Bloomfield (2009) denies that the fact that counterfactuals are context-sensitive undermines Nelson's argument, and his reason for this is as follows. Bloomfield describes for us three possible worlds in which molecules break the glass if the molecules x, y and z do not break the glass: (i) "w, where the water is boiling and molecules p, q and r break the glass, (ii) w*, where the water is boiling and molecules f, g and h break the glass and, (iii) w** where the water isn't boiling but an angel speeds up only molecules p, q and r in such a way as to break the glass." (Bloomfield, 2009, 343) Bloomfield argues that context sensitivity, which should be more accurately characterized as conversational implicature, will only allow the L-L CPU to rule out w**[27]

Bloomfield goes onto argue that (e*) does not distinguish between w and w* as the possible world closest to the actual. It cannot distinguish between them because the consequent of the counterfactual conditional (e*) does not say which molecules

---

[27] Bloomfield thinks conversational implicature, rather than context-sensitivity, is involved because we are discussing the semantic content of a particular counterfactual and that certain interpretations of it would be ruled out by any evaluator of the counterfactual as "out of bounds". As we are trying to understand the content of (e*), the Gricean rules of interpretation alone (or so the suggestion goes) will rule out consideration of those worlds in which the water boils because the molecules are moved around by angels (or some such); we don't need to appeal to some fancy theory of contextualism to rule them out. (Bloomfield in conversation). I will ignore these complexities in what follows.

will crack the glass if molecules x, y and z do not. According to Bloomfield, then, Miller's (2009) argument does not undermine Nelson's (2006) argument as the L-L CPU will not know which possible world is closest to the actual world because he knows which proposition any utterance of (e*) expresses, since any utterance of (e*) does not distinguish between whether w or w* is the closest possible world to the actual world.

Bloomfield, however, even seems to think that knowing which possible world is closest to the actual world and knowing which molecules break the glass if molecules x, y and z do not is irrelevant to explaining why (e*) is true, since: "being able to say what the truth-conditions for a proposition are does not, by itself, entail being able to explain why those are the conditions in which the proposition comes out true." (Bloomfield, 2009, 344) Here Bloomfield endorses Nelson's reading of the counterfactual (e*) and the need to explain the general claim in order to explain why (e*) is true. Bloomfield, like Nelson, thinks that the L-L CPU will have to use program explanation in order to see why (e*) is true. To illustrate why Bloomfield thinks this, Bloomfield says that all (e*) says is simply, if molecules x, y and z do not break the glass, some other molecules will. Being more specific, the consequent of the counterfactual conditional (e*) does not specify which molecules will crack the glass, but just that as (e*) is true, it is guaranteed that in the relevant situations sufficient molecules will be actualised to crack the glass. As (e*) does not say which specific molecules will crack the glass, all (e*) says is that there is a group of possible worlds in which the water is boiling and then picks out one of these possible worlds from the group and says that this particular possible world is closest to the actual world.

According to Bloomfield in order to explain why (e*) is true, the L-L CPU will have to use program explanation to pick out a group of possible worlds where the

water is at boiling temperature. From the group selected by program explanation, we have to make sure that the closest possible world to the actual world is a member of this group.

Applying this to the moral case, although the L-L CPU is assumed to have all the process explanations of moral belief in terms of lower-level non-moral properties, modal information about how things do go in other possible worlds, according to Bloomfield's (2009) argument, the unavailability of program explanation will result in an explanatory loss. Contrary to Miller (2009), Bloomfield denies that the L-L CPU knows which possible world is closest to the actual world, since any proposition any utterance of (e*) expresses does not distinguish between (analogues of) w and w*. So with only reference to process explanations the L-L CPU cannot explain why (e*) is true. Recall that according to the Stalkner-Lewis account of counterfactuals, (e*) is true if and only if in the possible world that is closest to the actual world in which molecules x,y and z do not strike the molecular bonds at place pn at time tn with momentum mn, some other molecules strike the molecular bonds with sufficient momentum to break them. Armed with only process explanations the L-L CPU will not be able to see that the consequent is true because, as already remarked, the L-L CPU does not know the closest possible world to the actual world.

Chapter 5

In this chapter I offer an argument against Bloomfield. The argument against Bloomfield if successful will show that the fact that counterfactuals are context-sensitive undermines his and Nelson's argument that program explanation invokes distinctive information not available via an epistemically unlimited agent's grasp of a process explanation.

Recall that in the previous chapter I argued that Nelson's position of asking for an additional explanation of why the counterfactual (e*) is true, other than the explanation of how the molecules do go in the possible worlds, is unmotivated. One reason that Nelson may argue that such an additional explanation of why the counterfactual (e*) is true is required since any explanation why (e*) is true has to explain the general claim. So Nelson's and Bloomfield's argument's about the need for the general claim of (e*) has to succeed or their position is unmotivated. Furthermore, Bloomfield claims that the context-sensitivity of counterfactuals does not adequately explain why the counterfactual conditional (e*) is true. So the only explanation of why (e*) is true will have to explain *the general claim* that can only be explained by the use of program explanation. I will argue that in fact the context-sensitivity of counterfactuals do in fact undermine Nelson's and Blackburn's argument for the need of program explanation in explaining why (e*) is true.

Argument against Bloomfield

Consider a context-sensitive statement:

"I love Sophie" is true as uttered by U at time T if and only if (iff) U loves Sophie at T

Now as the sentence "I love Sophie" currently stands, we do not know which proposition it currently expresses, so we are unable to assign it a truth-value.

To find out the proposition that the sentence expresses, we have to add in facts about the context of utterance (the contextual features) and the character to determine the content. For example, the contextual features are that the sentence is uttered by Andrew at 3pm, so that "U" refers to Andrew and "T" refers to 3pm. The character tells us to look for the speaker of the utterance, the time and hearer of the utterance, and this tells us what has to obtain in a possible world for that sentence to be true in that world. Once we have the context in addition to character, then we have the content of the statement and know the proposition the sentence expresses.[28]

Now, consider a context-sensitive counterfactual conditional:

"A $\square\rightarrow$ B" is true iff in the closest possible world in which A is true, B is true.

Once again, as the counterfactual conditional currently stands, it expresses no proposition and thus the counterfactual conditional has no truth value.

Again we have to add the context of the utterance in addition to the character to determine the content. As we saw in our earlier example concerning stepping on

---

[28] Here I am using the David Kaplan (1978) character-content distinction, and Chapter 11 of William G. Lycan's (2000) "Philosophy of Language: a contemporary introduction."

the icy path, the relevant contextual features relate to speaker intentions, and these in turn determine what counts as the closest possible world for the purposes of evaluating the truth of the counterfactual. Once we've added the context and the character, then we get the content, and we are in a position to know which proposition the context-sensitive counterfactual conditional expresses.

So now we get:

"A $\Box\rightarrow$ B" is true iff B is true in Φ

Turning our attention back to Bloomfield's argument, Bloomfield thinks that the counterfactual conditional (e*) does not express a proposition telling us which particular molecules will break the glass if molecules x, y and z do not, but only a proposition telling us that if molecules x, y and z do not break the glass, then some other unspecified molecules will. Recall that "A $\Box\rightarrow$ B" is true iff in the closest possible world in which A is true, B is true, so in order to test the truth of a counterfactual conditional we need to know what happens in the closest possible world to the actual world in which the antecedent obtains. This is how Bloomfield thinks that (e*) currently stands, and as it currently stands, without the relevant contextual facts specified the counterfactual conditional expresses no proposition. To bring in the relevant contextual information required in order for a truth-value to be determined, Bloomfield thinks that we have to bring in the kinds of higher-order properties that figure in the explanandum in program explanations.

We can see this as Bloomfield says:

"So, in order to evaluate (e*), [the L-L CPU] will have to pick out the set of worlds in which

the water is boiling, that is, the set picked out by the program explanation, and then make sure

that the nearest world to the actual world is a member of this set. Picking out the relevant set

of worlds will require [the L-L CPU] to appeal to the property of the water that is its *boiling*

and thereby program explanation." (Bloomfield, 2009, 344 [original italics])

The problem for Bloomfield is that the L-L CPU does not need to use program

explanations, or appeal to the higher-order properties that figure in their

explanandums, to get the relevant contextual information. The L-L CPU already has

the contextual information relevant to determining the truth-value of the

counterfactual conditional (e*) in virtue of the fact that the L-L CPU, as defined

earlier in the paper, knows everything there is to know about the speakers

communicative intentions and so on, and therefore knows the proposition that any

given utterance of (e*) expresses. So the L-L CPU knows which possible world is

closest to the actual world, and is therefore in a position to determine the truth-value

of any given utterance of (e*) solely on the basis of his knowledge of facts about

lower-level properties and the process explanations in which they appear.

Going back to Miller (2009) when he writes:

"[The L-L CPU] needn't to do anything to determine the answer to (a) [which possible world

is closest to the actual world]: *he already knows the answer to (a) in virtue of the fact that he*

knows which proposition any given utterance of (e) expresses*. So, a fortiori, in order to see

why (e*) is true God needn't do anything that requires reference to higher-level temperature

properties." (Miller, 2009, 341[emphasis added])

The L-L CPU already knows which possible world is closest to the actual world in virtue of the fact that he knows which proposition any given utterance of (e*) expresses, because the L-L CPU already has (and knows) the contextual features relevant to determining the truth-value of the counterfactual conditional (e*), such as, the speakers communicative intentions and so on. So it is the case, contrary to Bloomfield, that the L-L CPU can tell which of world w [where the water is boiling and p,q and r break the glass] and w* [where the water is boiling and f,g and h break the glass] is closest to the actual world, and thus by knowing which possible world is closest to the actual world can explain why the counterfactual conditional (e*) is true. Moreover, then, it is *not* the case that the unavailability of the relevant program explanations will result in an explanatory loss for the L-L CPU. So the program explanations that invoke moral facts and properties do not count as best for the L-L CPU and thus, they do not count as best in the right sort of way for their availability to earn ontological rights for moral facts and properties.

An additional objection to Nelson and Bloomfield is as follows. In general, if you're asked to explain why a sentence is true, you take its truth conditions for granted, and then explain why they obtain. Bloomfield and Nelson seem to require that we explain why it *means* what it means, then *also* explain why that truth-condition is satisfied. As Nelson and Bloomfield say:

"Being able to say what the truth conditions for a proposition are does not, by itself, entail being able to explain why those are the conditions in which the proposition comes out true."

This is wrong since to know whether a sentence is true you need to need to know what it means, but you don't need to know *why* it means what it means. For example,

take the sentence: "snow is white". Now to know whether the sentence is true, as long as we know what it means, we take the truth-conditions for granted and we then explain why those truth-conditions obtain: for example, the sentence "snow is white" is true because snow is in fact white. It seems absurd to ask in addition, "why it is that "snow is white" means that snow is white".

Now take the counterfactual conditional sentence (e*): "if molecules x,y and z had not struck the molecular bonds at place p_n at time t_n with momentum m_n, the glass still would have cracked, owing to the striking of other molecules, at other, or other places, or with slightly different momenta." We know what this sentence means, and we know that it is true because we can explain why those truth conditions obtain: for instance, because in the closest possible world to the actual world, we can see that the group of molecules f,g and h crack the glass if molecules x,y and z do not. This is all we need for an explanation of why the sentence is true: we do not need, in addition, an explanation of why it means what it means.

Moreover, even if this objection is incorrect and it is well motivated to ask the additional question "why is it that so-and-so means so-and-so?", then it still does not mean that program explanation is needed in order to explain why the counterfactual conditional (e*) is true. Since the L-L CPU knows all the relevant contextual features, the lower-level properties and the process explanations in which they figure, the L-L CPU will know why the sentence means what it means, even if we don't.

Lastly, it seems that Bloomfield may have misconstrued what Miller meant by an epistemically unlimited agent vis-à-vis lower level properties and the process explanations in which they figure. Bloomfield seems to think that the epistemically unlimited agent is a genuine agent, like God, rather than a heuristic device, as Bloomfield writes:

"Admittedly, these are tricky issues. At the very least, trying to know what an epistemically unconstrained agent would or would not have to know in order to explain something is fairly mind-boggling by itself. (Explain it to whom?)" (Bloomfield, 2009, 344)

This may be the reason that Bloomfield thinks that in order to bring in the relevant contextual information required in order for the counterfactual conditional (e*)'s truth-value to be determined, Bloomfield thinks that we have to bring in the kinds of higher-order properties that figure in the explanans of a program explanation.

Conclusion

Towards the end of his 2006 paper Sturgeon starts to advocate a kind of program explanation. Sturgeon writes:

"How evaluative explanations can be not just credible, but in some cases illuminating. A good causal explanation needs to describe the cause at the right level of abstraction, including relevant detail while excluding the irrelevant. That is why, even granting that slavery was unjust, it can be reasonable to ask whether the American anti-slavery movement is best described as responding to slavery's injustice or, instead, only to some specific way or ways that it was unjust. The answer presumably depends largely on how it did, or would have, responded to different forms of injustice … The answer might be that a more specific explanation is better, or it might be that the explanation that appeals to the injustice of slavery is better. And in *either* case, as we have seen, it might well be that the only terminology we have for characterizing the cause at the right level is evaluative. (Sturgeon, 2006, 251 – 2)

However, although Sturgeon thinks that "many evaluative explanations look promising in this regard", program explanation, as this thesis has demonstrated, will not confer ontological rights for moral facts and properties. (Sturgeon, 2006, 252). Despite its promise then, Cornell Realism appears incapable of yielding a plausible alternative to the metaethical views we surveyed in chapter one.

Bibliography

Ayer, A.J [1936] (1946) *Language, Truth, and Logic*, 2nd edn. London: Gollancz

Blackburn, S. (1984) *Spreading The Word*. Oxford: Oxford University Press

Bloomfield, P. (2009) "Moral realism and program explanation: a very short symposium 2: reply to Miller", *Australasian Journal of Philosophy* 87: 343-344

Boyd, R. (1988) "How to be a moral realist," in G. Sayre-McCord (ed.), *Essays on Moral Realism*. Ithaca, NY: Cornell University Press, 181 – 228

Brink, D. (1989) *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press

Darwall, S., Gibbard, A., and Railton, P. (1992) "Toward *fin de siecle* ethics: some trends." *Philosophical Review* 101, 115 – 189

Divers, J. (2002) *Possible Worlds*, London: Routledge

Dworkin, R. (1996) "Objectivity and truth: you'd better believe it." Philosophy and Public Affairs 25, 87 - 139

Frankena, W. (1938) "The naturalistic fallacy." *Mind* 48, 464 – 477

Geach, P. (1960) "Ascriptivism." *Philosophical Review* 69, 221 – 225

Geach, P. (1965) "Assertion." *Philosophical Review* 74, 449 – 465

Gibbard, A. (1990) *Wise Choices, Apt Feelings*. Oxford: Clarendon Press

Harman, G. (1977) *The Nature of Morality*. Oxford: Oxford University Press

Harman, G. (1986) "Moral explanations of natural facts – can moral claims be tested against reality." *Southern Journal of Philosophy* 24, suppl., 57 – 68

Harman, G. (1999) "Moral philosophy meets social psychology: virtue ethics and the fundamental attribution error." *Proceedings of the Aristotelian Society* 99, 315 – 331

Harman, G. (2000) "The nonexistence of character traits." *Proceedings of the Aristotelian Society* 100, 223 – 226

Jackson, F., and Pettit, P. (1995) "Moral functionalism and moral motivation." *Philosophical Quartely* 45, 20 -40

Jackson, F, and Pettit, P. (2004) "Program explanation: a general perspective", in

Jackson, F, Pettit, P, and Smith, M (ed.) *Mind, Morality and Explanation*, Oxford: Oxford University Press, 119 – 130

Kaplan, D. (1978) "Dthat", in P. Cole (ed.), *Syntax and Semantics, vol. 9: Pragmatics, New York*: Academic Press; reprinted in P.French, T. Uehling, and H.Wettstein, 221-243

Kim, J. (1984) "Concepts of supervenience." *Philosophy and Phenomenological Research* 45, 153 – 176

Kim, J. (1989) "The myth of nonreductive physicalism." *Proceedings and Addresses of the American Philosophical Association* 63, 31 – 47

Landau, R.S (2003) *Moral Realism: A Defence*. Oxford: Oxford University Press

Lowe, E.J. (1995) "The truth about counterfactuals", *Philosophical Quarterly* 45: 41-59

Lycan, W.G. (2000) *Philosophy of Language: A Contemporary Introduction*. London: Routledge

McDowell, J. (1988) *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press

McNaughton, D. (1988) *Moral Vision*. Oxford: Blackwell

Mackie, J. (1977) *Ethics: Inventing Right and Wrong*. New York: Penguin

Miller, A. (2003) *An Introduction to Contemporary Metaethics*. Cambridge: Polity

Miller, A. (2009) "Moral realism and program explanation: a very short symposium 1: reply to Nelson", *Australasian Journal of Philosophy* 87: 337 – 341

Moore, G.E. [1903] (1993) *Principia Ethica*, rev. edn. Cambridge: Cambridge University Press

Nelson, M. (2006) "Moral realism and program explanation", *Australasian Journal of Philosophy* 84: 417 - 28

Putnam, H. (1967) "Psychological predicates," in W.H Captian and D.D Merrill (eds.) *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press, 37 - 48

Railton, P. (1986a) "Moral realism." *Philosophical Review* 95, 163 - 207

Railton, P. (1986b) "Facts and values." *Philosophical Topics* 14 (2), 5 - 31

Smith, M. (1994) *The Moral Problem*. Oxford: Blackwell

Sturgeon, N. (1986) "Harman on moral explanations of natural facts." Southern Journal of Philosophy 24, suppl., 69–78

Sturgeon, N. (2006) "Moral explanations defended," in Dreier, J (ed.) *Contemporary Debates in Moral Theory*. Oxford: Blackwell, 241 – 262

Wiggins, D. (1987) "A sensible subjectivism," in *Needs, Values, Truth*. Oxford: Blackwell, 185 – 214

Wright, C. (1988a) "Moral values, projection, and secondary qualities." *Proceedings of the Aristotelian Society*, suppl. Vol., 1 -26

Wright, C. (1988b) "Realism, antirealism, irrealism, quasi-realism." *Midwest Studies in Philosophy* 12, 25 – 49

Wright, C. (1989) "Wittgenstein's rule following considerations and the central project of theoretical linguistics," in A. George (ed.) *Reflections on Chomsky*. Oxford: Blackwell, 233 – 264

Wright, C. (1992) *Truth and Objectivity*. Cambridge, MA: Harvard University Press