

Sorites

Σωρίτης

A Digital Journal of Analytical Philosophy
Indexed and Abstracted in *THE PHILOSOPHER'S INDEX*
ISSN 1135-1349

Legal Deposit Registration: M 14867-1995

Editor: Lorenzo Peña

(CSIC [Spanish Academy of Sciences])
Logic-and-Law Research Group (JuriLog)

Board of Editorial Consultants:

Jean-Yves Béziau, Enrique Alonso, Guillermo Hurtado,
Manuel Liz, Raymundo Morado, Txetxu Ausín

Regular-Mail Address:

Prof. Lorenzo Peña
CSIC-CCHS — Department of Philosophy
Albasanz 26
E-28037 Madrid
Spain

Fax +3491 22 986 91
Voice Tph +3491 60 223 72

<<http://www.sorites.org/>>
<<http://www.sorites.info/>>
<sorites@sorites.org>

Issue #19 — December 2007

SORITES (Σωρίτης) ®

ISSN 1135-1349

Issue #19 — December 2007

Copyright © by SORITES and the authors

<<http://www.sorites.org/>>

<<http://www.sorites.info/>>

<sorites@sorites.org>

SORITES (Σωρίτης)

ISSN 1135-1349

ROLL OF REFEREES

Rainer Born Johannes-Kepler Universitaet Linz (Austria)
Amedeo Conte University of Pavia (Italy)
Newton C.A. da Costa University of São Paulo (Brazil)
Marcelo Dascal University of Tel Aviv (Israel)
Dorothy Edgington Birbeck College (London, UK)
Graeme Forbes Tulane University (New Orleans, Louisiana, USA)
Manuel García-Carpintero University of Barcelona (Spain)
Laurence Goldstein University of Hong Kong (Hong Kong)
Jorge Gracia State University of New York, Buffalo (USA)
Nicholas Griffin McMaster University (Hamilton, Ontario, Canada)
Rudolf Haller Karl-Franzens-Universitaet Graz (Austria)
Terence Horgan University of Memphis (Tennessee, USA)
Victoria Iturralde Univ. of the Basque Country (San Sebastian, Spain)
Tomis E. Kapitan Northern Illinois University (USA)
Manuel Liz University of La Laguna (Canary Islands, Spain)
Peter Menzies Australian National University (Canberra, Australia)
Carlos Moya University of Valencia (Spain)
Kevin Mulligan University of Geneva (Switzerland)
Jesús Padilla-Gálvez Johannes-Kepler Universitaet Linz (Austria)
Philip Pettit Australian National University (Canberra, Australia)
Graham Priest University of Queensland (Brisbane, Australia)
Eduardo Rabossi University of Buenos Aires (Argentina)
David-Hillel Ruben School of Oriental and African Studies, University of London
Mark Sainsbury King's College (London, UK)
Daniel Schulthess University of Neuchâtel (Switzerland)
Peter Simons University of Leeds (Leeds, UK)
Ernest Sosa Brown University (Providence, Rhode Island, USA)
Friedrich Stadler Institut «Wien Kreis» (Vienna, Austria)

SORITES (Σωρίτης)

ISSN 1135-1349

Issue #19 — December 2007

TABLE OF CONTENTS

| | |
|--|-----|
| — Abstracts of the Papers | 03 |
| — «On An Attempt to Undermine Reason-Responsive Compatibilism by Appealing to Moral Luck. Reply to Gerald K. Harrison» by Sergi Rosell | 07 |
| — «Whither Morality in a Hard Determinist World?» by Nick Trakakis | 14 |
| — «Essential Dependence and Realism» by Daniel Laurier | 41 |
| — «The Logic of ‘If’ — Or How to Philosophically Eliminate Conditional Relations» by Rani Lill Anjum | 51 |
| — «Wittgenstein and the Sorites Paradox» by David Michael Wolach . . | 58 |
| — «Saying the Unsayable: Wittgenstein’s Early Ethical Thought» by Paul Formosa | 61 |
| — «Is the Yellow Ball Green?» by Jack Lee | 74 |
| — «Incommensurability and Interpretation» by Anthony D. Baldino . . . | 79 |
| — «On the Semantic Indecision of Vague Singular Terms» by Dan López de Sa | 88 |
| — «Truthmakers for Negative Truths» by Yuki Miyoshi | 92 |
| — «Reference, Knowledge, and Scepticism about Meaning» by Elisabetta Lalumera | 108 |
| — The SORITES Charter | 123 |
| — Release Notice | 129 |

SORITES (Σωρίτης), ISSN 1135-1349
<http://www.sorites.org>
Issue #19 — December 2007. Pp. 03-06
Abstracts of the Papers
Copyright © by SORITES and the authors

ABSTRACTS OF THE PAPERS

ON AN ATTEMPT TO UNDERMINE REASON-RESPONSIVE COMPATIBILISM BY APPEALING TO MORAL LUCK. REPLY TO GERALD K. HARRISON

by **Sergi Rosell**

This is a reply to G.K. Harrison's article «Hyper Libertarianism and Moral Luck». There he argues for the advantage of hyper-libertarianism upon reason-responsive compatibilism in virtue of its integration of moral luck in a principled way. I shall try to show that his argument is unsound. Crucial to my reply will be that Harrison's idea of moral luck is an unjustifiedly narrow one. Although the aim of establishing an appropriate connection between the issues of moral luck and free will is worth pursuing, I shall argue that moral luck cannot solve the free will dispute in the way Harrison intends.

WHITHER MORALITY IN A HARD DETERMINIST WORLD?

by **Nick Trakakis**

What would the world be like if hard determinism were true, that is, if all events were determined in such a way as to render all our decisions and actions unfree? In particular, what would morality be like? Indeed, could there be anything distinctively moral in such a world, or would we be left with a moral nihilism in which nothing of moral significance remains? In this paper I explore the ethical implications of hard determinism, focusing on the consequences that our lack of free will would have for moral responsibility (and thus praise and blame), moral obligation, moral rightness and wrongness, and moral goodness. I argue that the truth of hard determinism would compel us to significantly revise our commonsensical understanding of these moral categories. I add, however, that this change in moral outlook would not have dire practical consequences, for we would retain the attitudes and emotions that are essential to forming good interpersonal relationships and to developing morally. In fact, far from being a threat to human flourishing, hard determinism offers the prospect of a life that is morally deeper and more fulfilling than in a world in which we are free.

ESSENTIAL DEPENDENCE AND REALISM

by **Daniel Laurier**

It has recently been suggested that realism about some subject matter is best construed as the claim that the facts pertaining to this subject matter are *essentially* independent from the mind, in a sense to be explained, and *not* as the admittedly weaker claim that they are *modally*

independent from the mind. In this paper, I argue that this proposal is liable to trivialize the realist's position and is biased against his irrealist opponent.

THE LOGIC OF 'IF' — OR HOW TO PHILOSOPHICALLY ELIMINATE CONDITIONAL RELATIONS

by Rani Lill Anjum

In this paper I present some of Robert N. McLaughlin's critique of a truth functional approach to conditionals as it appears in his book *On the Logic of Ordinary Conditionals*. Based on his criticism I argue that the basic principles of logic together amount to epistemological and metaphysical implications that can only be accepted from a logical atomist perspective. Attempts to account for conditional relations within this philosophical framework will necessarily fail. I thus argue that it is not truth functionality as such that is the problem, but the philosophical foundation of modern logic.

WITTGENSTEIN AND THE SORITES PARADOX

by David Michael Wolach

Any discussion regarding the famous Sorites Paradox is incomplete without considering the value of contextual logic and its meta-language of vagueness. Wittgenstein, though he did not write extensively on the Sorites Paradox in particular, is deeply concerned with its supposed implications. The later Wittgenstein's treatment of logical vagueness in natural and formal languages, and his accompanying treatment of logical soundness (necessary and sufficient conditions) as it applies to ordinary languages is thus of considerable help when thinking about the Sorites Paradox. In this paper I pair the later Wittgenstein's treatment of meaning in context with the age-old problem, manifest in the Sorites Paradox, of what happens when we apply induction ad infinitum to a seemingly stable item in a specific, meaning-bearing lexicon.

SAYING THE UNSAYABLE: WITTGENSTEIN'S EARLY ETHICAL THOUGHT

by Paul Formosa

In this paper I present an account of Wittgenstein's ethics that follows from a so-called 'metaphysical' reading of the *Tractatus*. I argue that Wittgenstein forwards two distinct theses. Negatively, he claims that there can be no ethical propositions. Positively, he claims that the ethical good, or good in-itself, is the rewarding happy life. The happy life involves living in perfect contented harmony with the world, however it is, because how the world is, is a manifestation of God's will. Given the negative thesis, the positive thesis cannot strictly speaking even be said. We can only make sense of this by assuming that Wittgenstein takes this positive thesis to be 'illuminating nonsense'.

IS THE YELLOW BALL GREEN?

by Jack Lee

It sounds contradictory that the yellow ball is green. Indeed, it is believed that yellow is not green. If so, then the yellow ball shouldn't be green, either. In this paper, however, I want to argue that «the yellow ball is green» is intelligible. To achieve this purpose, I distinguish between (and analyze) two kinds of concepts: «digital concepts», and «analog concepts». By using this pair of concepts, I propose to show that the yellow ball is in a very important sense green. To sum up, there are at least two kinds of concepts: «digital concepts», and «analog concepts». «Color» is among the analog concepts. It is argued that the analog concepts do not conform to the principle of non-contradiction. However, to be handled well, they must be digitalized first. But note that these digitalized concepts are in reality analog. Therefore, in reality, it is not contradictory that the yellow ball is green.

INCOMMENSURABILITY AND INTERPRETATION

by Anthony D. Baldino

Although the central target of Donald Davidson's influential essay 'On the Very Idea of a Conceptual Scheme' is the scheme/content distinction, Davidson also maintains that his argument undermines the thesis of incommensurability as advocated by Thomas Kuhn and Paul Feyerabend. It will be argued here that when Davidson's contentions are carefully disentangled and scrutinized, the elements needed to dismantle the scheme/content distinction are ones that do not subvert incommensurability, and the ones that are held to contravene incommensurability are implausible. Therefore, it is shown that it is possible to be an incommensurabilist without holding on to any objectionable third dogma, and that the disrepute that the thesis of incommensurability has fallen into based on claims of semantic incoherence like those offered by Davidson is undeserved. Many in the philosophy of science heralded the thesis of incommensurability as a radical, intriguing, and powerfully informative conjecture about the history of science. This essay tries to reconnect a large segment of the philosophy of language, influenced to dismiss the thesis based on semantic qualms such as Davidson offers, with that segment of philosophy of science that still considers and utilizes the thesis of incommensurability as a powerful explanatory and elucidative tool.

ON THE SEMANTIC INDECISION OF VAGUE SINGULAR TERMS

by Dan López de Sa

According to a popular, plausible, but also controversial view about the nature of vagueness, vagueness is a matter of semantic indecision. I show that, even if «I» is vague and the view of vagueness as semantic indecision is correct, I could be a material composite object all the same.

TRUTHMAKERS FOR NEGATIVE TRUTHS

by Yuki Miyoshi

Finding truthmakers for negative truths has been a problem in philosophy for a long time. I will present and discuss the solution to this problem offered by Bertrand Russell, Raphael Demos, D. M. Armstrong, and myself. I will argue that some negative truths do not require truthmakers and that truthmakers for the other negative truths are the entities that these negative truths imply exist. I will also argue that truthmakers for general truths of the form, only X, Y, Z, and etc. are F's, are the sum of each X, Y, Z, and etc.'s being F, and hence that sometimes truthmakers do not necessitate truths.

REFERENCE, KNOWLEDGE, AND SCEPTICISM ABOUT MEANING

by Elisabetta Lalumera

This paper explores the possibility of resisting meaning scepticism — the thesis that there are many alternative incompatible assignments of reference to each of our terms — by appealing to the idea that the nature of reference is to maximize knowledge. If the reference relation is a knowledge maximizing-relation, then some candidate referents are privileged among the others — i.e., those referents we are in a position to know about — and a positive reason against meaning scepticism is thus individuated. A knowledge-maximizing principle on the nature of reference was proposed by Williamson in a recent paper (Williamson 2004). According to Williamson, such a principle would count as a defeasible reason for thinking that most of our beliefs tend to be true. My paper reverses Williamson's dialectic, and argues that reference is knowledge-maximizing from the premise that most of our beliefs tend to be true. I will therefore defend such premise on different grounds than Williamson's, and precisely by revisiting a Naturalist argument he rejected, centred on the role of true beliefs in successful action. In the conclusion, an opposition to meaning-scepticism comes out as motivated by the knowledge-maximizing nature of reference, and backed by the plausibility of the claim that beliefs tend to be true.

SORITES (Σωρίτης), ISSN 1135-1349
<http://www.sorites.org>
Issue #19 — December 2007. Pp. 07-13
On An Attempt to Undermine Reason-Responsive
Compatibilism by Appealing to Moral Luck. Reply to
Gerald K. Harrison
Copyright © by Sergi Rosell and SORITES

ON AN ATTEMPT TO UNDERMINE REASON-RESPONSIVE COMPATIBILISM BY APPEALING TO MORAL LUCK. REPLY TO GERALD K. HARRISON

by **Sergi Rosell**

1. In ‘Hyper Libertarianism and Moral Luck’,¹ Gerald K. Harrison tries to put forward a ‘surprising’ view, as he says, of the prospects for libertarianism being the Principle of Alternative Possibilities (PAP, henceforth) refuted. He claims that «the kind of libertarian positions which survive post PAP are ones which have the resources to make sense of moral luck in a way unavailable to compatibilism.» (93) After recognising the historical support that PAP has provided for libertarian positions on freedom and moral responsibility, he remarks that refuting PAP does not involve refuting libertarianism; there is logical room for libertarian positions in a post PAP scenario. So, assuming PAP as refuted, the author focuses on the actual sequence and wonders whether causal determinism in the actual sequence rules out moral responsibility.

Harrison commits himself to a type of hyper libertarianism (or source incompatibilism, as it is more commonly known), that is, a libertarian position which defends that determinism undermines moral responsibility for reasons that have not to do with the availability of alternative possibilities or with the agent’s ability to do otherwise. Beyond this claim, his position is indeed very close to Kane’s (1996). He stresses that «determinism deprives one of ownership over one’s decisions»; we need to be the ultimate, or (in Kane’s words) ‘buck-stopping’ explanation of our decisions in order for them to be truly ours. That is, a requirement for ownership or ultimacy is needed; and this is a different requirement than that about control. «[I]n addition to controlling our decisions we need also to own them... it is ownership that is threatened by determinism.» (94) His position can be labelled as *Kane minus PAP*, whose key element is the demand of ownership in the actual sequence. The problematic of such an account will be seen later on in this paper.

But compatibilists have also their own interpretations of ownership: you own your decisions insofar as «the world had to go through you to get that to happen». Of course, the compatibilist conception of ownership is restricted to the actual sequence, but that is exactly the same for source incompatibilism. Furthermore, an important gain for compatibilism is John M. Fischer’s claim that it should be favoured «in so far as our basic views about ourselves — our view of ourselves as persons and as morally responsible — should not be held hostage

¹. In this journal (December 2005), 16: 93-102. Page numbers without year reference correspond to Harrison’s article.

to the discoveries of a consortium of scientists about the precise nature of the equations that describe the universe». (2003: 221)

So, what reasons do we have to favour libertarianism over compatibilism? Harrison values the previous claim for (semi-)compatibilism, but he thinks there is a counterweight favouring libertarianism: in a post PAP scenario, only a hyper-libertarian perspective can make sense of the phenomenon of moral luck; and only hyper libertarianism can achieve a principled explanation of it (97). Then, the query is about the capacity of both compatibilism and hyper-libertarianism to *make sense* of the phenomenon of moral luck, and that must be understood as being able to provide a *principled explanation* of it. In particular, Harrison affirms the incapacity of compatibilism to achieve this kind of principled explanation.

I have to say that I find very commendable his target of trying to establish a satisfactory connection between the issue of moral luck and the free will and moral responsibility debate. I am really sympathetic with his idea that not giving room to moral luck — or not having an adequate explanation of it — counts as a serious demerit for a position in the debate. Harrison's position can be described as one which makes (I hope not to introduce more confusion here) moral luck 'compatible' with freedom and moral responsibility, and in a libertarian way. However, Harrison's paper lacks an explicit characterization of what he understands by moral luck, and a clarification of the kind of moral luck addressed; a fundamental point inexcusably missed.

The rest of this response will focus on assessing Harrison's argument for the advantage of hyper libertarianism upon compatibilism concerning integration of moral luck. Furthermore, an attempt will be made to make clear what his conception of the latter is. I shall argue, on the one hand, that his argument is not convincing and, on the other, that the idea of moral luck shown in his paper is an unusual and extremely narrow one. Besides that, a direct treatment of the very topic of moral luck is first needed to be able to apply it next to the debate about free will and moral responsibility.

2. In section III Harrison introduces a Fischer-style reason-responsiveness account of control, and the problem that moral luck posits to it. According to that account, what it is for a decision or choice to have been controlled is simply for it to have been the output of a certain type of a mechanism that is sensitive or responsive to reasons to *a certain degree*. In this sense, we can speak, in technical terms, of a *moderate* reasons-responsiveness. Having the control relevant to moral responsibility means simply having some sort of mechanism sufficient to be moved by reasons; or, in other words, having capacities and dispositions to respond in certain ways to a relevant range of inputs. However, the problem is that

...if an agent has only compatibilist control over their decision and choices, then they nevertheless lack control over how their mechanism operates in the actual sequence. Given that they are, by hypothesis, morally responsible for the decisions that they make, this means that the agent will be exposed to certain kind of moral luck. It is their bad luck, for instance, that they possess a mechanism which, in these exact circumstances, will issue in this, morally reprehensible decision. For in different but relevantly similar circumstances [any possible world in which both the mechanism and the morally relevant reasons are held fixed], it would have issued in a different, blameless, or even praiseworthy decision. (95)

The point, thus, is that the control mechanism is sensitive to the morally relevant reasons but only to a limited extent. So, in different but relevantly similar circumstances the mechanism would have issued in a different decision, which can deserve a different moral assessment. Adding some example would be good here. Anyway, the point is that this type of control relies on merely possessing capacities that he is not able to exercise control over.

We can suspect at this time that the problem Harrison is pointing out is that of (a type of) control in achieving our rational capacities. But this turns out not to be the point Harrison makes. The question concerns rather our lack of control upon the (moral) reasons to which we are responsive given our moderate mechanism of reasons-responsiveness. In his opinion, the unique alternative here, to avoid the exposure to ‘the kind of moral luck in question’, is to insist that what is required for moral responsibility is *strong* reason-responsiveness. Accordingly, the decision delivered by the strongly responsive mechanism will be the same in all scenarios in which both the mechanism and the morally relevant reasons are held fixed. So, the concern is not our control in how we arrive to form our reason-responsive mechanism, but its refinement. If our mechanism is totally attuned to the reasons there are, it does not mind how we have achieved it.

But strong reason-responsiveness is far too demanding a control requirement. It would mean, as Harrison claims, that «to avoid exposure to this kind of moral luck, the reasons to which the mechanism would have to be strongly sensitive would have to be moral reasons present, which would have the upshot that no agent could ever do wrong culpably.» (96) But, while for a compatibilist an agent with a strong reason-responsive mechanism could, in principle, be morally responsible for the decisions and choices that they make; for a hyper-libertarian position (or, at least for Harrison’s type), she could not.

The alternative Harrison defends is a compatibilist *control* (as moderate reason-responsiveness) plus *ownership*, understood in terms of Kane’s choices or ‘self-forming acts’ (SFAs). Both the post-PAP hyper-libertarian and the compatibilist should agree about control (mechanism), both must affirm the reality of ‘the kind of moral luck outlined’, and both have to give a coherent account of it. But, ultimately, for Harrison, making sense of this kind of moral luck is unavailable for the compatibilist. Only with a mechanism internally deterministic in place we will be able to say that it would have delivered the same decision in all relevant possible worlds. However, if the mechanism were internally indeterministic there would always be some possible worlds in which the same reasons are present and the mechanism is held fixed, yet a different decision issues.

But this seems a little confusing. Harrison has said that there is a problem with luck for Fischer-style compatibilism and now he states that it does not have place for luck. The answer could be: for moral responsibility, moderate reasons responsiveness is not enough, while strong reason responsiveness is too much. He claims:

If internal indeterminism is required, as I maintain that it is, then something less than strong reason responsiveness is actually a requirement for moral responsibility. This means that the kind of freedom needed for moral responsibility actually exposes the agent to the moral luck in question. (97)

However, Harrison is mixing without justification considerations about allowing luck in the mechanism and considerations about the most fitting account of moral responsibility. Regarding moral luck, it seems that both moderate reason-responsiveness and hyper-libertarianism’s indeterministic decisions can indeed accommodate it. But, he still can reply that hyper-libertarianism does not only give room for ‘the moral luck in question’, it indeed gives a *principled explanation*. According to him, «if we only have compatibilist control, then it seems fundamentally unfair to hold us morally responsible for the decisions and choices that we make in the actual sequence, for we would have made different decisions in alternative, but relevantly identical sequences.» (96) But actually this is the same, or worse, for indeterministic decisions. In a post PAP scenario, the only addition hyper-libertarian ownership condition can do is introducing indeterminism in a certain point of the reason-responsiveness

process. But to appeal here, in relation to moral luck, to indeterminism is to get nothing, unless you confuse luck with indeterminism as such, as I think Harrison does. Indeed, luck is any type of an agent's lack of control, either in an indeterminist way or in a determinist one.²

3. But we must pay more attention to the very nature of the moral luck at stake here. What is this 'kind of moral luck in question'? As stated above, Harrison does not give any explicit characterization of what he understands by moral luck, or the kind of connexion he has in mind with the classical (Nagelian) types of moral luck.³ Indeed, he does not even mention any author or literature on moral luck. Remember that *constitutive luck* — that is, the lack of control in the achievement of the reason-responsive mechanism each of us has, or hasn't — has already been discarded as the kind of moral luck *in question*. It seems to me that we must understand Harrison's kind of moral luck as the luck involved in our reason responsiveness at some point (in a time/decision-slice), and which will have repercussion on the moral judgments we deserve. Specifically, luck in the output we yield according to the input we receive, given an internal indeterministic reason-responsive mechanism. It is the kind of moral luck implied in the fact that, for instance, the business woman of Kane's example finally chose either to help the unfortunate pedestrian or to go to the meeting. Kane's account (Kane 1996) gives room for luck in the sense that there is not an explanation for the fact that she tries, by means of SFAs, to be one way and not the other,⁴ and this is not to say that her decisions are accidental, insofar they are product of her efforts and willings.

However, why does the compatibilist not explain the kind of luck in question as luck due to the variation in the mechanism's responsiveness to reasons the agent has? Indeed, I think this is the most natural way to understand the point: our choices are a function (deterministic or not) of our reasons, intentions, capacities (result of an all-time forming life) and the opportunities present to us at choice's time.

But, beyond that, the competition is unfair; Harrison dismisses inexplicably the relevance of the mechanism's history for Fischer and Ravizza's theory of moral responsibility — which indeed constitute their account of the ownership condition. In such an account, an agent's mechanism is appropriately/owner reasons-responsive only if she has come to own that mechanism by means of a process whereby she *takes responsibility* for the mechanisms that give rise to her actions. In order to take responsibility for her conduct, an agent must see herself in a certain manner (1998: 207-39). Only under that condition she owns her mechanisms of decision. This is the main content of the ownership condition, and then, it is an essential part of what has to be confronted with Kane's SFAs.

On the other hand, what about the extra merit of hyper-libertarianism in achieving a principled way of explaining the phenomenon? Actually, Harrison can only state that hyper-libertarianism has a principled explanation of moral luck by means of constructing an 'ad hoc' notion of moral luck. His concept of moral luck is too narrow, and the only thing that explains such a narrow conception is that it is an 'ad hoc' notion. But moral luck is a wider

². On the other hand, Harrison also seems not to notice the distinction between (reason-responsive) *mechanisms* and *agents*. I put this issue aside. Cf. Moya (2006: 116-ff).

³. See Nagel (1979).

⁴. Although there is a reasons explanation, but neither determinist nor contrastive.

phenomenon; indeed, it is a very complex one. The phenomenon is usually motivated in terms of a collision between the belief that we cannot blame (or judge morally negatively) someone for those outcomes of her actions that are beyond her control and the fact that we judge people for such things that simply are the outcome of their actions. However, the issue extends also to the fact that an agent has to face some or other circumstances, with possible repercussion on her responsibility, or to possess a certain constitution or to have received such and such influences and not others. It is a clash between so settled an intuition as the control requirement is, and a working ordinary practice. Daniel Statman summarizes the ‘problem’ as the tension posed by the fact that it seems that moral responsibility, justification, blame, and so on, cannot be affected by luck and the possibility that luck plays an outstanding, even essential, role (1993: 2).⁵

In short, a case of moral luck will happen when we morally judge someone in a right way for something that is, in some sense, beyond her control. And what is beyond the agent’s control is not only the result of a Kanean SFA; there is much more in the agent’s rational constitution due to luck. Moreover, Fischer has emphasized recently this fact, and indeed has been typically accused of giving luck too much room in his account of moral responsibility.⁶

However, although Harrison acknowledges that what he has said only shows «how one kind of luck could be intelligible», he maintains that this kind of moral luck he is pointing out is «the most fundamental» (98). But he is not sufficiently clear here. As advanced, a clear statement of what kind of moral luck he is vindicating would be needed. He also says that a direct treatment of the moral luck problem itself will be a topic of another paper. It would be good if he tried to become clearer about the very topic of moral luck, before applying it to the free will and moral responsibility debate.

4. Furthermore, I am suspicious about the possibility of defending Kane’s position without endorsing PAP, and do not know how exactly a picture of a Kanean-minus-PAP account would look. It seems that if Harrison accepts, for instance, the business woman example, he had to endorse all of Kane’s account, including PAP.

For moral responsibility in Kane’s account, both PAP and indeterminism are necessary conditions (we need also the plurality and the will-setting conditions), although they are not needed for every action, but only for SFAs. According to Kane (1996), there are some fundamental free choices to an agent’s autonomy which result of processes of deliberation guided by efforts of the will. In such a kind of cases, a conflict arises between what an agent thinks she ought to do and her actual wants or desires. And one struggles to sort out and to establish a priority order within one’s own values. The idea of «efforts of will» is crucial to this picture: they both produce and occupy a causal gap between reasons and choices, which allows that the possible outcomes will be *undetermined*, and also *indeterminate*. As Timothy O’Connors puts it, «at each stage of the struggle, the possible outcomes have no specific objective probability of occurring.» (2005; sect. 3.2). Doubtless, in Kane’s account, this kind of indeterminacy is essential to freedom of will; but also PAP is.

It is true that Kane (Kane 2002) has lately rejected the excessive focus on the requirement

⁵. Williams (1981) and Nagel (1979) are the original sources on Moral Luck; see also the anthology by Statman (1993).

⁶. See Fischer (2006).

of alternative possibilities of recent debates about free will. Certainly, Kane states that the idea of being ultimate responsible for an action is more important, but he says explicitly that alternative possibilities are still a necessary condition for ultimate responsibility. He states:

For UR does require that we could have done otherwise with respect to *some* acts in our past life histories by which we formed our present characters. ...we formed it by other choices or actions in the past (SFAs) for which we could have done otherwise (which did satisfy AP). (2002: 408)

Indeed, it seems to me that the insight the (libertarian) idea of ultimacy or ownership wants to capture can only be satisfied if the agent could have done otherwise; that is, having alternative possibilities is what would make the agent into a *true* owner — as a condition different of rational control — of her decisions and actions.

On the other hand, Harrison seems very happy with the idea that indeterminism — although it does not preclude an agent's being reasons-responsive enough for moral responsibility — does nothing to enhance control and will in fact diminish it. He says:

In absence of any special kind of libertarian control, making an agent's reason responsive mechanism internally indeterministic will diminish the degree of control it can be said to deliver, for it will invariably make it less responsive than it would otherwise have been. (97)

In this way, Harrison acknowledges a tension between the ownership condition and the control condition — as they are conceived by him — «insofar as it [the ownership condition] can only be satisfied if the agent has less than the kind of full control that would rule out moral luck.» But why does he think that moral luck plays a role favourable to ownership, or rule out full control in the direction of ownership? And, why does moral luck do that against a compatibilist account of ownership?

I think his final point is especially controversial; he opposes the fairness-related aspect of moral responsibility to its freedom-related one. He says: «An agent is morally responsible not because it would be fair to hold them morally responsible, but because they made their decision freely.» (98) But this is a false dilemma: moral responsibility is about freedom and about fairness; an agent can be fairly held morally responsible only if he is free, in the adequate sense. To think otherwise—in the direction Harrison does — is to embrace an image of ownership dangerously close to voluntarism or freedom arbitrarism.

I agree that control, or direct control, is not all there is to moral responsibility; but the concern about moral luck does not stem, as Harrison states, from focussing only on the control condition. The ownership conditions can be also vulnerable to luck. Luck is a pervasive phenomenon. Moral luck is rather the cost of moral responsibility.

5. To conclude, Harrison's position can be described as one which makes moral luck compatible with freedom and moral responsibility. This is something that I myself support. His point is that a control mechanism is compatible with moral luck, and this does not undermine moral responsibility because there is another requirement at issue: the requirement of ultimacy, which assures our ownership over our decisions, namely, that we are the 'buck-stopping' explanation of them. But this idea of ultimacy is a very puzzling one; it has received a lot of criticism, and is one for which Harrison does not offer any extra argument beyond those given by Kane — but, at the same time, adding the incoherence of rejecting PAP. His point was that this account is superior to the compatibilist one only because it can give room to the phenomenon of moral luck and explain it in a principled way. But, clearly, Fischer-style compatibilism gives also room to moral luck, and if it cannot explain it in a principled way—as

hyper libertarianism supposedly can — it is because ‘the kind of moral luck in question’ Harrison talks about is an *ad hoc* kind, specially designed for filling the indeterministic hole of hyper libertarianism — a hole that has been traditionally seen as a weakness.

To summarize, Harrison’s argument is not convincing. The idea of moral luck he uses is extremely narrow and, so, clearly inadequate. Moral luck cannot solve the dispute in the way Harrison intends.⁷

References

- Fischer, John M. (2006) «The Cards That Are Dealt You», *The Journal of Ethics* 10: 107-129.
- (2003) «Frankfurt-style Compatibilism», in Watson, G. (ed.) *Free Will*, New York: Oxford University Press.
- & Ravizza, Mark (1998) *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.
- Harrison, Gerald K. (2005) «Hyper Libertarianism and Moral Luck», *SORITES* 16: 93-102.
- Kane, Robert (2002) «Some Neglected Pathways in the Free Will Labyrinth», in Kane, R. (ed.) *The Oxford Handbook of Free Will*, New York: Oxford University Press.
- (1996) *The significance of Free Will*, Oxford-New York: Oxford University Press.
- Moya, Carlos (2006) *Moral Responsibility: The Ways of Scepticism*, Oxford: Routledge.
- Nagel, Thomas (1979) «Moral Luck», in Nagel, T., *Mortal Questions*, New York: Cambridge University Press.
- O’Connor, Timothy (2006) «Free Will», *The Stanford Encyclopedia of Philosophy* (Spring 2006 Edition), Edward N. Zalta (ed.), URL <<http://plato.stanford.edu/archives/spr2006/entries/freewill/>>.
- Statman, Daniel (1993) «Introduction», in Statman, D. (ed.) *Moral Luck*, Albany: State University of New York Press.
- Williams, Bernard (1981) «Moral Luck», in Williams, B., *Moral Luck*, Cambridge: Cambridge University Press.

Sergi Rosell
University of Valencia, Spain
<sergi.rosell@uv.es>

⁷. I thank Carlos Moya for helpful comments on a previous draft. Research funded by project ‘Belief, Responsibility, and Action’, Spanish Ministry of Education and Science (HUM2006-04907) and University of Valencia-V Segles Fellowship.

WHITHER MORALITY IN A HARD DETERMINIST WORLD?

by Nick Trakakis

Suppose we accept the truth of hard determinism. Suppose, in other words, we come to believe that everything that happens is necessitated to happen, so that for every event that takes place there are conditions (such as antecedent physical causes plus the laws of nature, or our genetic make-up and environment, or the decrees and foreknowledge of God) whose joint occurrence is logically sufficient for the occurrence of that event. And suppose further that we come to believe that, since everything that happens is determined to happen, there is no free will, or at least no free will of the sort required for genuine responsibility, blameworthiness, or desert. What consequences (other than the repudiation of some form of moral responsibility and concomitant notions) would the adoption of such a view have for morality?

Many, of course, have thought that nothing less than moral nihilism would result: ‘If all is determined, everything is permitted’. Peter van Inwagen, for example, writes that, «If there is no free will, then morality as it is ordinarily conceived is an illusion».¹ Similarly, Ishtiyaque Haji has argued that in a deterministic world we would be bereft of any ‘moral anchors’, that is to say, moral deontic normative statuses like those of being morally right, wrong, or obligatory.² There are, however, a growing number of dissenters who maintain that the consequences for morality, given hard determinism, are not as dire as is ordinarily feared. Indeed, some have gone on to argue that hard determinism offers us a ‘purer’ or ‘higher’ form of morality than would otherwise be available.³

In looking for a way to settle this dispute, I will concentrate on four central moral notions: moral responsibility (and the allied notions of praiseworthiness and blameworthiness), moral obligation, moral rightness and wrongness, and moral goodness. The question, then, to be addressed is: Must the hard determinist adopt a ‘scorched earth’ policy, setting all such moral

¹. Peter van Inwagen, «Response to Slote,» *Social Theory and Practice* 16 (1990), p.394.

². See Haji, «Moral Anchors and Control,» *Canadian Journal of Philosophy* 29 (1999), pp.175-203, and *Deontic Morality and Control* (Cambridge: Cambridge University Press, 2002), ch. 3. I should point out, however, that, on Haji’s view, ‘deontic morality’ is undermined not so much by the absence of free will but by the truth of determinism (I thank Ish Haji for clarifying this). See also Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1996), ch. 6, where he discusses ten important goods that humans generally desire and which are often thought to be imperiled by the absence of libertarian free will.

³. For a list of some exponents and detractors of the idea that there can be no substantive morality without moral responsibility, see Bruce Waller, «Virtue Unrewarded: Morality Without Moral Responsibility,» *Philosophia* 31 (2004), pp.427-28, 442-43 (fn.5). Waller himself supports the view that «morality — morality of almost any variety one favors — can survive and flourish in the absence of moral responsibility» (p.428).

categories to the flames, or can some or even all of them be retained or at least rehabilitated?

At the outset, however, I would like to counter a possible line of response to this question. One might contend that even if there is little or no evidence that we are free and morally responsible, the consequences of denying moral responsibility are so grave that we should continue living under the (perhaps false) impression that we were free in the sense required for moral responsibility. This line of thought goes as far back as Kant, and has been vigorously defended in more recent times by Saul Smilansky.⁴ According to Smilansky, libertarian free will does not exist, but once we remove it from our picture of reality, our fundamental values, practices and attitudes — such as the belief in our potential for blameworthiness according to what we do, our appreciation of achievement in ourselves and others, the sense of value and meaning in our lives, and so on — would be placed under great risk, if not largely destroyed. Since the consequences for humanity of widespread belief that we lack libertarian free will would be utterly destructive, Smilansky exhorts us to maintain and even promote the illusion that we possess such free will. As he puts it, «In the context of free will,...we cannot live well with the full truth. So, we shall have to make do with illusion, to a significant extent.»⁵

This is an interesting response, but I think it can be easily dispensed with. Firstly, I suspect that most people would find it psychologically difficult to engage in the required sort of pretense. Smilansky, to be fair, is not suggesting that the illusory belief in libertarian free will needs to be induced in people, for as he is aware, that belief is already part of our cognitive endowment. For Smilansky, however, what is crucial is that there is a ‘double illusion’ in place: most people hold the false belief that we have responsibility-entailing free will, while being oblivious to the fact that their belief is false. There is therefore a widespread illusion about the presence of illusion.⁶ The question, however, remains as to what people with a psychology like ours would likely do once they discovered that libertarian free will is merely an illusion. Smilansky’s answer seems to be that the illusion of free will fulfills such important emotional needs in people that it is unlikely that many people would be able to

⁴. As is well known, Kant held that, despite the paucity of evidence in support of agent-causal libertarianism (which, in his view, is required for moral responsibility), we must postulate that we possess free will so as to uphold the validity of the moral law. Smilansky defends a similar position in a number of publications, but especially in *Free Will and Illusion* (Oxford: Clarendon Press, 2000).

⁵. Smilansky, *Free Will and Illusion*, p.170. Smilansky, I might add, does not think that hard determinism is the whole truth of the matter in regards to free will. Rather, he adopts the ‘dualist’ view that both the hard determinist perspective and the compatibilist perspective capture part of the truth relating to free will, with neither perspective being in principle superior to the other (see ch. 6 of his *Free Will and Illusion*). Smilansky, however, holds that just as one cannot live well with the belief that there is no libertarian free will, so too one cannot live well with the dissonance created by acknowledging the validity of both compatibilism and hard determinism. Dualism, therefore, creates a further need for illusion regarding free will (*Free Will and Illusion*, pp.149-61).

⁶. For this reason, Smilansky stresses that his position is not to be equated with Hans Vaihinger’s idea that, although we lack libertarian free will, we ought to behave as if we possessed such free will — see Vaihinger, *The Philosophy of ‘As If’: A System of the Theoretical, Practical and Religious Fictions of Mankind*, 2nd ed., trans. C.K. Ogden (London: Routledge & Kegan Paul, 1935), pp.42-48. As Smilansky states, «Vaihinger generally holds that awareness of the fictional character of an idea accompanies its use. My position is different — lack of awareness (to varying degrees) of the illusory element is often the very condition of its productive existence» (*Free Will and Illusion*, p.147).

accept the harsh reality with regard to free will.⁷ But surely some people will hit upon the truth and will be brave enough to accept it — Smilansky himself is a case in point. Would these brave souls be required to engage in some sort of self-deception in order to carry on with their daily lives? It is not clear, in other words, how someone in Smilansky's shoes, who is fully aware of the illusory character of free will, can (without discounting this awareness) continue thinking of himself and others as free and morally responsible agents.⁸

Secondly, Smilansky's line of response presupposes that much of what is important in human life (e.g., morality, interpersonal relations) would be undermined if we came to believe that hard determinism is true. Smilansky, to be sure, has his reasons for thinking that an acceptance of determinism, including hard determinism, has these damaging effects. However, if hard determinism has no such implications, then the proposed pragmatic justification for belief in free will and moral responsibility would be undercut. Although not a criticism of Smilansky, this does indicate that it is incumbent on us to determine what the moral consequences of hard determinism in fact are before even considering whether we should opt for a pragmatic defence of 'illusion'.

Thirdly, when we believe that someone is blameworthy for committing a serious moral offense, we are normally angry with them, we hold them in contempt, and we endeavour to punish them by, for example, incarcerating them. Clearly, these actions (of expressing anger and contempt towards someone, of punishing them) tend to be very harmful to the offender. But it is arguably the case, though some wish to deny this, that it is permissible to intentionally inflict harm on someone when they do not deserve to be harmed (i.e., when they are not genuinely blameworthy for their behaviour) *only if* one has very good reason to believe that by inflicting such harm they will be procuring some great benefit for the sufferer.⁹ However, in many cases of the relevant sort, we would lack any reason to think that our treatment of the offender is likely to benefit them — indeed, in some cases (e.g., those involving the death penalty) we would have good reason to think that our treatment will bring

⁷. At least, this is what is suggested by comments made by Smilansky on pp.263-64 of *Free Will and Illusion*. The difficulty here, as Smilansky has related to me in correspondence, concerns the matter of internalization: «I think that most human beings (including myself, even after all those years of dealing with the problem) have great difficulties in really internalizing the absence of libertarian free will, and what it implies about things like their moral worth» (personal correspondence, 10 April 2005).

⁸. A further unsavoury, if not paradoxical, aspect of Smilansky's 'illusionism' is that, given the grave dangers involved in people becoming aware of the illusory nature of free will, it is important to keep the truth on this matter as hidden as possible. Smilansky, for example, writes that «it is important for her role as part of a community that the philosopher does not proclaim to everyone her findings about free will» (*Free Will and Illusion*, p.274). Even if this kind of deception or dishonesty can be justified (as Smilansky thinks it can, in pp.258-68 of *Free Will and Illusion*), one wonders why Smilansky publishes papers and books proclaiming the non-reality of libertarian free will, rather than making every effort to conceal this truth through publications that defend libertarianism with rhetorically effective (albeit invalid and unsound) arguments.

⁹. A principle of this sort has been defended by many philosophers of religion in the context of the problem of evil, where it is argued that it is morally permissible for God to allow a human being S to undergo intense, involuntary and undeserved suffering only if the good secured by S's suffering is shared by S in a conscious way. See, for example, Eleonore Stump, «The Problem of Evil,» *Faith and Philosophy* 2 (1985), p.411; William Alston, «The Inductive Argument from Evil and the Human Cognitive Condition,» *Philosophical Perspectives* 5 (1991), p.48; and Marilyn Adams, *Horrendous Evils and the Goodness of God* (Melbourne: Melbourne University Press, 1999), pp.29-31. For some dissenting views, see Peter van Inwagen, «The Magnitude, Duration, and Distribution of Evil: A Theodicy,» *Philosophical Topics* 16 (1988), p.184, and C. Behan McCullagh, «Evil and the God of Love,» *Sophia* 31 (1992), p.55.

more harm than good to them.¹⁰

(An autobiographical note at this point might be appropriate before proceeding further. This investigation is intended as a prelude to a somewhat distant topic in the philosophy of religion, viz., the problem of evil. In particular, I want to see how this problem can be tackled by a theist who also happens to be a hard determinist. But since the problem of evil is at bottom a *moral* problem, it will be helpful to first ascertain what the moral world of the hard determinist — and by extension the theistic hard determinist — looks like.)

My procedure in this paper is to take the views of Derk Pereboom, one of the ablest contemporary defenders of hard determinism, as my point of departure.¹¹ Pereboom argues that some, but not all, of our central moral categories can be salvaged from the ruins of hard determinism. My view is that Pereboom does not go far enough. In particular, I will argue that all of our central moral notions, and not merely some of them, can be reconciled with the hard determinist outlook — though such reconciliation, I argue, will be possible only after the moral notions in question have, at least to some degree, been reconceived. But first it will be important to get clear about some terminology.

I. Hard Determinism and Hard Incompatibilism

Following William James¹², determinists are usually grouped into two camps. There are *soft determinists* or compatibilists, who believe that determinism does not undermine any free will or responsibility worth having, and there are *hard determinists*, who take a harsher line: determinism is true and therefore free will (or at least the kind of free will required for moral responsibility) does not exist. More perspicuously, hard determinism — at least as classically expressed by such thinkers as Spinoza and Holbach¹³ — is committed to the following three

¹⁰. The second and third criticisms made above are indebted to Derk Pereboom, *Living Without Free Will* (Cambridge: Cambridge University Press, 2001), pp.198-99.

¹¹. I will be relying in the main on Pereboom's defence of hard determinism in *Living Without Free Will*. For more succinct expressions of Pereboom's position, see his «Determinism *Al Dente*,» *Noûs* 29 (1995), pp.21-45, and «Living Without Free Will: The Case For Hard Incompatibilism,» in Robert Kane (ed.), *The Oxford Handbook of Free Will* (Oxford: Oxford University Press, 2002), ch. 21.

¹². See William James, «The Dilemma of Determinism,» in *The Will to Believe and Other Essays in Popular Philosophy* (New York: Longmans, Green, and Co., 1917), p.149.

¹³. Pereboom also cites Joseph Priestley (1733-1804) as a hard determinist (see Pereboom, *Living Without Free Will*, p.129). However, it is clear from Priestley's major writings on determinism — viz., *The Doctrine of Philosophical Necessity Illustrated* (1777), and his correspondence with Richard Price, published as *A Free Discussion of the Doctrines of Materialism and Philosophical Necessity* (1778) — that he rejected the libertarian conception of free will (or what he calls 'the doctrine of philosophical liberty'), but accepted a compatibilist understanding of free will. He notes, for example, that

I allow to man all the liberty or power that is *possible in itself*, and to which the ideas of mankind in general ever go, which is the *power of doing whatever they will, or please*, both with respect to the operations of their minds and the motions of their bodies, uncontrouled by any foreign principle or cause.

(*The Doctrine of Philosophical Necessity Illustrated*, Section I, in *The Theological and Miscellaneous Works of Joseph Priestley*, New York: Klaus Reprint, 1972, vol. 3, p.459, emphases in the original.)

Elsewhere, Priestley writes that

To me it seems sufficient that men be *voluntary agents*, or that motives, such as hopes and fears, can influence them in a certain and mechanical manner to make it in the highest degree *right* and *wise* in the Divine Being to lay such motives before them, and consequently to place them in a state of moral discipline, or a state in which rewards and punishments are distributed so as to correspond to certain characters and actions.

(*A Free Discussion of the Doctrines of Materialism and Philosophical Necessity*, Introduction, in *The Theological and Miscellaneous Works of Joseph Priestley*, New York: Klaus Reprint, 1972, vol. 4, p.11, emphases in the

theses:

- (1) Free will (in the strong sense required for moral responsibility and moral desert) is not compatible with determinism.
- (2) There is no free will (in the above sense), because:
- (3) All events are determined.

Contemporary hard determinists, however, are unlikely to accept the third of these theses, preferring instead to remain non-committal about the truth of universal determinism (particularly in the light of developments in twentieth-century physics). Derk Pereboom is a case in point. He describes himself as a ‘hard incompatibilist’ rather than a ‘hard determinist’, where hard incompatibilism involves a commitment to the following two theses:

- (4) All of our actions and choices are either alien-deterministic events (i.e., events such that there are causal factors beyond our control by virtue of which they are causally determined), or truly random events (i.e., events not produced by anything at all), or partially random events (i.e., events such that factors beyond the agent’s control contribute to their production but do not determine them, and there is nothing that supplements the contribution of these factors to produce the events).
- (5) An action is free in the sense required for moral responsibility only if the decision to perform it is not an alien-deterministic event, nor a truly random event, nor a partially random event.

The idea behind (5), or what Pereboom calls ‘the Causal History Principle’, is that an agent’s responsibility for an action is explained not by the existence of alternative possibilities open to her, but rather by the action’s having a certain causal history — viz., one which allows the agent to be the source of her action in such a way that she has control over the production of her decision to perform that action.¹⁴ But according to (4), all our actions and choices lack the requisite kind of causal history, as they are produced by (deterministic or random) processes over which we have no control. From (4) and (5) it clearly follows that we do not have free will of the sort required for moral responsibility.¹⁵

By contrast, classical hard determinism holds that we lack free will because we live in a fully deterministic world. Pereboom, however, does not want to commit himself to universal determinism. He would rather leave the matter open, arguing instead that, since our actions and choices are events that lie on the continuum from alien-deterministic through partially random to truly random events, we have no freedom in the sense required for moral responsibility. But as Pereboom is careful to point out, his position is not to be confused with the ‘no-free-will-either-way’ theory advocated by contemporary hard determinists such as Galen Strawson and Richard Double. On this view, free will of the kind required for moral

original.)

¹⁴. Pereboom concedes that Frankfurt-type cases succeed in showing that it is not the lack of alternate possibilities in a deterministic world that prevents us from holding agents in such a world morally responsible and blameworthy for their actions. See Harry G. Frankfurt, «Alternate Possibilities and Moral Responsibility,» *Journal of Philosophy* 66 (1969), pp.829-39. For Pereboom, what undermines moral responsibility in a deterministic world is not the lack of alternate possibilities, but the lack of causal histories that render agents ultimate sources of their actions.

¹⁵. See Pereboom, *Living Without Free Will*, pp.127-28.

responsibility is incompatible with both determinism and indeterminism, and so irrespective of whether the world turns out to be deterministic or indeterministic, we could not have free will of the relevant sort. Pereboom disagrees with this assessment, arguing instead that the kind of indeterminacy required for agent-causal libertarianism *is* compatible with the sort of free will required for moral responsibility. In his view, however, the evidence of our best scientific theories suggests that no such indeterminacy is to be found. This, of course, means that our lack of free will is a merely contingent fact: we might have been undetermined agent-causes, and hence free and responsible agents.¹⁶

In what follows I will speak only of ‘hard determinism’, taking thesis (2) above as the kernel of this position.¹⁷ On this reading, the hard incompatibilist and the ‘no-free-will-either-way’ theorist belong to the same camp as their classical determinist predecessors, and this is not entirely inappropriate. For what they all have in common is the belief that there is no free will, in the responsibility-entailing sense, although each group arrives at this belief from a different direction. It is this belief, furthermore, that distinguishes them from both compatibilists and libertarians.¹⁸

II. Moral Responsibility, Praise and Blame

It is often thought that the first and most obvious casualty in the moral realm, once hard determinism is adopted, is the idea that human beings are sometimes morally responsible for their decisions and actions. Indeed, this result seems to simply fall out from the account of hard determinism given above, as thesis (2) states that there is no free will *of the sort required for moral responsibility*. But what is it for a person to be morally responsible? Roughly speaking, to be morally responsible for something (e.g., an action, an omission, a character trait) is to be worthy of a particular kind of reaction — typically, praise or blame. But when are such reactions of praise and blame appropriate or called for? Answers to this question give rise to two competing conceptions of moral responsibility. According to one view, the *merit-based view*, praise or blame is an appropriate reaction toward someone if and only if they merit — in the sense of ‘deserve’ — such a reaction, given their behaviour and/or character traits. This is perhaps the commonsensical view of moral responsibility, but there is an alternative: the *consequentialist view*, according to which praise or blame is an appropriate reaction if and only if a reaction of this sort is likely to bring about a desired change in the agent and/or their behaviour.¹⁹

¹⁶. See Pereboom, *Living Without Free Will*, pp.128-33, where Pereboom compares his position with the views of other contemporary hard determinists.

¹⁷. Cf. Robert Kane, «Introduction: The Contours of Contemporary Free Will Debates,» in Kane (ed.), *The Oxford Handbook of Free Will*, p.28, where both thesis (1) and thesis (2) are taken to be the kernel of the hard determinist position.

¹⁸. It may be noted that it is entirely consistent with the hard determinist view, as expressed in (1), (2) and (3) above, that humans do possess free will, albeit not free will of the sort required for moral responsibility. That is why Bruce Waller can argue in *Freedom Without Responsibility* (Philadelphia, PA: Temple University Press, 1990) that «our contemporary, nonmiraculous, naturalist (determinist) world system leaves no room for moral responsibility — though it leaves quite adequate space for individual freedom» (p.4).

¹⁹. I take the labels ‘merit-based’ and ‘consequentialist’ from Andrew Eshleman, «Moral Responsibility,» §1, in *The Stanford Encyclopedia of Philosophy (Fall 2004 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2004/entries/moral-responsibility/>>. The ‘consequentialist’ view of moral responsibility will also be referred to as the ‘utilitarian’ or ‘pragmatic’ view, without thereby wishing to draw any

Pereboom states that hard determinists must relinquish our ordinary view of ourselves as morally responsible in the merit-based sense, that is, as deserving blame for immoral actions and as deserving praise for actions that are morally exemplary.²⁰ In this regard, I think Pereboom is entirely correct. Even so, the hard determinist need not renounce moral responsibility altogether, but may instead reconceive our commonsense notion of moral responsibility in consequentialist terms. Indeed, this is the route normally taken by determinists, both hard and soft. J.J.C. Smart, for example, states that

Threats and promises, punishments and rewards, the ascription of responsibility and the nonascription of responsibility, have...a clear pragmatic justification which is quite consistent with a wholehearted belief in metaphysical determinism.²¹

Borrowing one of Smart's examples, we can see how this pragmatic or consequentialist justification would operate in practice by supposing that a student, Tommy, fails to do his homework and as a result his teacher deems him to be morally responsible or blameworthy. On the view we are considering, the teacher's reaction would be appropriate only if her act of blaming Tommy is likely to influence Tommy's attitudes and conduct in certain desirable ways.²² On the other hand, such a reaction on the part of the teacher would not be appropriate if Tommy suffered from some illness which prevented him from doing his homework, since in that case no amount of exhortation or threats would succeed in modifying his behaviour. But if Tommy's failure to do his homework is simply due to laziness, then his behaviour may well be open to influence, and the threats or punishments of his teacher would form part of the environment that helps influence Tommy's behaviour. In such a scenario, ascriptions of blame would have a definite and valuable purpose. As Smart sees it, then, ascriptions of moral responsibility, and therefore of praise and blame, have a purely instrumentalist rationale of influencing other people's behaviour.²³

distinction between these three labels.

²⁰. See Pereboom, *Living Without Free Will*, pp.139-41.

²¹. J.J.C. Smart, «Free Will, Praise and Blame,» *Mind* 70 (1961), p.302.

²². As Joel Kupperman points out, «This is *not* to say that people, including philosophers, normally praise or blame as a device to modify behavior; to equate the function of an utterance with a speaker's intention would be like saying that people in agony normally scream in order to get help» (*Character*, New York: Oxford University Press, 1991, p.60, emphasis in the original).

To prevent any unnecessary misunderstandings, I stress that 'blameworthiness' in this context simply consists in the appropriateness of outward expressions of blame by words, gestures, or actions — there is no implication of someone's being morally worthy (or deserving) of being judged in a certain way on account of their behaviour.

²³. Apart from Smart, other proponents of this pragmatic or consequentialist account of moral responsibility have included Henry Sidgwick, *The Methods of Ethics*, 7th edition (London: Macmillan, 1907), pp.71-72; Moritz Schlick, *Problems of Ethics*, trans. David Rynin (New York: Dover Publications, 1939), ch. 7; Richard B. Brandt, «A Utilitarian Theory of Excuses,» *Philosophical Review* 78 (1969), pp.337-61; Daniel Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (Oxford: Clarendon Press, 1984), pp.163-65; Michael Slote, «Ethics Without Free Will», *Social Theory and Practice* 16 (1990), pp.369-75; and Joel J. Kupperman, *Character*, ch. 3.

In «Denying Moral Responsibility: The Difference It Makes,» *Analysis* 49 (1989), pp.44-47, Bruce Waller points out some interesting implications of such pragmatic accounts of praise and blame. Waller notes that the hard determinist who aligns his views with the findings of contemporary behavioural science (with its account of how behaviour is shaped by various contingencies and schedules of reinforcement) will respond quite differently to good and bad behaviour in comparison with a believer in (a non-utilitarian conception of) moral responsibility. In fact, such a hard determinist will dispense rewards and punishments in the *opposite* way to that in which someone who

Unfortunately, as Richard Arneson points out, this instrumentalist account is «the position everyone loves to hate.»²⁴ Various objections have been raised against this account, but here I will only address two of the more common ones.²⁵ The first of these is derived from an objection that is often raised against consequentialist ethical theories in general, viz., the objection that such theories do not give sufficient weight to the needs or rights of individuals, and as a result these theories allow or even promote the unjust or cruel treatment of people in certain circumstances. Peter van Inwagen echoes this objection when considering the following utilitarian theory of what it is to assign blame or to say ‘X is Y’s fault’ (with ‘X’ denoting a state of affairs and ‘Y’ referring to some agent):

- (6) X is an unfortunate state of affairs and Y brought about X and, for some group of people, it would maximize the general welfare if they were to do something unpleasant to Y and to describe their motive to Y and to the public as follows: We are doing this unpleasant thing to Y because Y brought about X.

Van Inwagen has us consider a mentally retarded person who, ignorant of his own strength

accepts moral responsibility would dispense rewards and punishments. More specifically, the behavioural hard determinist will offer more generous rewards (positive reinforcements) to the vicious individual (whenever that individual’s actions resemble the desired good behaviour) than to an individual who has been conditioned to act virtuously, since greater positive reinforcement is required to reform the reprobate’s behaviour than to sustain the virtuous individual’s behaviour. By contrast, someone who rejects hard determinism would offer praise whenever it is deserved, and the virtuous (on this view) always deserve greater praise than the vicious. «Behavioural hard determinism,» writes Waller, «rejects moral desert claims, and the schedule of rewards (positive reinforcement) it proposes is fundamentally *incompatible* with those based on moral desert/responsibility» (p.46, emphasis in the original). Furthermore, «the hard determinist *justification* of positive and aversive conditioning will be the opposite of the justification given for reward/punishment by believers in moral responsibility» (p.46, emphasis in the original). Strangely, in *Freedom Without Responsibility* (pp.135-40), Waller offers these observations as an *objection* to the pragmatist conception of moral responsibility. I would have thought that Waller’s behaviourist account fits quite nicely with the utilitarian view of moral responsibility.

It is important also not to overlook the fact that the hard determinist has the option of offering various non-utilitarian accounts of moral responsibility. A hard determinist may, for example, defer to virtue ethics in developing an account of moral responsibility, so that judgements about a person’s blameworthiness (or praiseworthiness) for performing a certain action are reduced to judgements about that person’s character. Such a view was, in fact, endorsed by Spinoza, as Michael Slote illustrates in «Ethics Without Free Will», pp.375-79. See also Pereboom, *Living Without Free Will*, pp.148-52, and Ben Vilhauer, «Hard Determinism, Remorse, and Virtue Ethics,» *Southern Journal of Philosophy* 42 (2004), pp.547-64, where Vilhauer attempts to accommodate the phenomenon of remorse within hard determinism by incorporating a circumscribed consequentialism about punishment into a broad-based virtue ethic. On Vilhauer’s approach, see also fns 38 and 39 below. For a more pessimistic view of the compatibility of virtue ethics and determinism, see Haji, *Deontic Morality and Control*, ch. 11.

²⁴. Richard J. Arneson, «The Smart Theory of Moral Responsibility and Desert,» in Serena Olsaretti (ed.), *Desert and Justice* (Oxford: Clarendon Press, 2003), p.233.

²⁵. A further objection to the instrumentalist account is that agents, on this view, would be treated as blameworthy (by, for example, expressing indignation toward them) even though they do not deserve to be treated in this way — and this, it might be claimed, is unfair and morally wrong (such an objection is voiced by Pereboom, *Living Without Free Will*, p.156). However, if the person who is treated as blameworthy stands to gain greatly (in moral terms) from being so treated, then perhaps the treatment is morally justified despite being undeserved. An equally plausible option, one advocated by Pereboom, is to do away with praise and blame, and instead engage in moral admonishment and encouragement. As Pereboom states,

One could explain to an offender that what he did was wrong and then encourage him to refrain from performing similar actions in the future.

The hard incompatibilist can maintain that by admonishing and encouraging a wrongdoer one might communicate a sense of what is right, and a respect for persons, and that these attitudes can lead to salutary change.

(*Living Without Free Will*, p.157)

For a comprehensive list of, and response to, the traditional objections to the instrumentalist account of moral responsibility, see Manuel Vargas, «Moral Influence, Moral Responsibility» (unpublished).

and goaded by cruel bystanders, causes a serious injury to someone — let's say he knocks down a parking inspector with a powerful left hook and as a result the inspector falls into a coma. The police and the courts, we are asked to suppose, are aware that the offender will be lynched if he is acquitted on grounds of incapacity, and so they conspire to deprive him of his legal rights and send him to prison for assault. The problem here, according to van Inwagen, is that if we were to adopt the utilitarian reading of 'blame' encapsulated in (6), we would be permitted, if not obliged, to treat the offender in this paternalistic and harsh way, given that such treatment produces the best overall consequences.²⁶

Hypothetical cases such as this are well-known in the literature on utilitarian accounts of punishment. H.J. McCloskey, in particular, made much use of these cases in an attempt to show that the utilitarian is committed to endorsing unjust forms of punishment.²⁷ But various problems have been identified with McCloskey's examples, and these same problems affect van Inwagen's hypothetical scenario. I will briefly mention only three of these problems here.²⁸

Firstly, as many others have pointed out, the examples employed by non-utilitarians are often fanciful or bizarre, with the result that the ensuing debate is usually conducted (as Smilansky puts it) «in a somewhat unreal atmosphere».²⁹ Van Inwagen's example also has this 'unreal' character. For one thing, it is strange that people would want to lynch the offender after having goaded him into committing the offence; moreover, we would expect the police to be able to protect the offender were he to be judged 'not guilty'. It seems, therefore, that scenarios such as those imagined by van Inwagen are unlikely to arise in the real world³⁰; and insofar as utilitarianism is put forward as an empirical (or contingently true) theory, it cannot be refuted merely by devising logically possible counterexamples.³¹

²⁶. See van Inwagen, «Response to Slote,» p.386.

²⁷. See H.J. McCloskey, «An Examination of Restricted Utilitarianism,» *Philosophical Review* 66 (1957), pp.468-69; «A Non-Utilitarian Approach to Punishment,» *Inquiry* 8 (1965), pp.255-59; and «Utilitarian and Retributive Punishment,» *Journal of Philosophy* 64 (1967), pp.91-102.

²⁸. It may be worth noting, however, that the account of moral responsibility defended here does not state that the attitudes and practices characteristic of moral responsibility are justified because of the utility such attitudes and practices may have for some *group* or *community*. Rather, the view is that our responsibility-characteristic practices, particularly the practices of praising and blaming, are justified when they are likely to influence or pressure *individuals* to behave in morally desirable ways. It may be possible, however, to raise the objection from injustice even against this latter account of moral responsibility, as is indicated in fn. 25 above.

²⁹. Smilansky, «Utilitarianism and the 'Punishment' of the Innocent: The General Problem,» *Analysis* 50 (1990), p.257. See also T.L.S. Sprigge, «A Utilitarian Reply to Dr. McCloskey,» *Inquiry* 8 (1965), pp.272-75.

³⁰. Van Inwagen has, of course, the option of filling out his example in various ways, thus rendering it more 'believable'. But this is likely to raise the problems mentioned below with respect to cases that are 'true to life'.

³¹. Admittedly, it is more common to view utilitarianism and other normative ethical theories as theories that attempt to specify necessary and sufficient conditions for the application of concepts such as moral obligation and moral rightness. In line with this view, it is open to the non-utilitarian to say that the principle of utility enshrined by utilitarianism (roughly, the principle that an act is right or good if and only if it has the greatest utility — e.g., it produces the most favourable balance of happiness over unhappiness) is to be thought of as describing what would be right and wrong under any conceivable or logically possible circumstances. But then the utilitarian account of punishment would hold that there is no logically possible world in which it is right or good (in the utilitarian sense of these terms) for people to be unjustly punished, and this view can be defeated simply by way of logically possible

A second and more telling criticism of McCloskey's examples, that is also applicable to van Inwagen's example, is that if such cases are thought to depict 'real life' situations, it is very likely that the overall outcomes, in terms of costs and benefits, would be quite different than what is envisaged by the originators of these examples. Consider, for example, Sprigge's comments in response to an example of McCloskey's that closely parallels van Inwagen's example:

One line of objection to this conclusion [that the punishment of an innocent person is mandated from a utilitarian point of view] appeals to the likelihood that the facts will become known. I may urge parenthetically that in the real world such a likelihood is likely (surely) to be pretty strong. The utilitarian may then insist on a variety of evils which would result from its becoming known, such as a loss of confidence in the impartiality and fairness of the legal system, of a belief that lawful behaviour pays, etc.³²

Precisely the same point can be made with respect to van Inwagen's scenario, where the suspect behaviour of the police and the courts is likely to be exposed, sooner or later. This view is further bolstered by the fact that, if on one occasion we are prepared in the name of some greater good to tell lies which will cause great harm to someone or even ruin their life, then we risk blunting our moral sensibilities in such a way that we will have few qualms about using similar methods when far less justification for doing so is available. This, in turn, will set in motion a train of behaviour amongst the legal authorities that will almost certainly become public knowledge, thus undermining further the community's system of justice.

Thirdly, one may adopt a more sophisticated form of utilitarianism that is not vulnerable to the kinds of counterexamples put forward by the likes of McCloskey and van Inwagen. A 'preference utilitarian', for example, would give great weight to a person's preferences — including their preference to have their autonomy respected — when judging the moral value of an action.³³ A utilitarian of this sort could therefore argue that respect for autonomy is one of the most important factors that goes into promoting 'the general welfare'. In other words,

counterexamples. Perhaps this is what lies behind McCloskey's view that,

Against the utilitarian who seeks to argue that utilitarianism does not involve unjust punishment, there is a very simple argument, namely, that whether or not unjust punishments are in fact useful, it is logically possible that they will at some time become useful, in which case utilitarians are committed to them. («A Non-Utilitarian Approach to Punishment,» pp.254-55)

However, even if the principle of utility is to be construed as a theory about what would be right or wrong under any conceivable circumstances, it is unlikely that commonsense morality should be thought of in the same terms. As Timothy Sprigge notes, «Plain men will probably admit that if the empirical nature of the world had been very different then different moral sentiments would often have been appropriate» («A Utilitarian Reply to Dr. McCloskey,» p.273). But if the principles of commonsense morality have force only in the actual and nearby worlds, then that would explain our abhorrence when faced with far-fetched cases of the punishment of innocents. Alleged logically possible 'counterexamples', in other words, only show up the limitations of commonsense morality, not of utilitarianism. See, however, McCloskey's different assessment of the matter in «Utilitarian and Retributive Punishment,» pp.93-94.

³². Sprigge, «A Utilitarian Reply to Dr. McCloskey,» p.275.

³³. See Jonathan Glover, *Causing Death and Saving Lives* (London: Penguin, 1977), pp.80-82, and Peter Singer, *Practical Ethics*, 2nd ed. (Cambridge: Cambridge University Press, 1993), pp.99-100. A further option is to adopt a 'rule' or 'restricted' form of utilitarianism, according to which it is *the practice* of putting into jail those found guilty after a fair legal trial, rather than any particular instances of this practice, that can be justified on utilitarian grounds. See, however, McCloskey's «An Examination of Restricted Utilitarianism» for some forceful arguments against this form of utilitarianism.

to maximize the general welfare we would normally need to maximize the degree of autonomy had by individuals. But then, if to do the right thing usually requires maximizing the general welfare by maximizing autonomy, we could permissibly override a person's autonomy only in very exceptional circumstances — specifically, when no other avenue is available for protecting the welfare of that individual or other members of the community. But clearly such circumstances do not hold in van Inwagen's hypothetical case, unless we are given good reason to think that the offender cannot be protected by the police if he were judged 'not guilty' by the courts.³⁴

Another common response to the instrumentalist rationale for moral responsibility is sparked by the feeling that such a rationale runs deeply counter to our ordinary moral practices. We do not ordinarily think that ascriptions of moral responsibility and praise or blame *merely* have a pragmatic role of influencing behaviour. Indeed, we often praise and blame people without intending at all thereby to modify their future behaviour. Jonathan Bennett voices this objection well in relation to an instrumentalist position developed by Moritz Schlick:

The Schlickian description of what accountability is — or of what the concept is *for* — strikes everyone as incomplete and strikes most people as wrong. The latter will say that although a distinction based on the utility of a certain sort of therapy or behaviour-control might *coincide* with accountability/non-accountability, it cannot give the latter's essence, and that the Schlickian account of what the line is *for* does nothing like justice to the real nature of our praise- and blame-related responses. When we express indignation for someone's cruelty, or admiration for his unselfishness, we usually are not engaged in any sort of therapy: blame-related responses all involve something like hostility towards the subject; whereas a moral-pressure therapist, though he may have to feign ill-feeling for therapeutic purposes, can in fact be in a perfectly sunlit frame of mind. And — to move briefly to the 'welcome' side of the fence — one may apply moral pressures to encourage a welcomed kind of behaviour while remaining in an ice-cold frame of mind, with no feelings of gratitude, admiration or the like.³⁵

It may be thought, therefore, that the instrumentalist rationale for moral responsibility is inadequate insofar as it rules out the following activities as unfitting or irrational:

- (a) expressing feelings such as indignation, disgust, and contempt toward others on account of their misbehaviour;
- (b) expressing feelings of admiration or respect for those who do what they ought to or go beyond the call of duty in performing supererogatory acts;
- (c) making people suffer for their past deeds;
- (d) rewarding people for special achievements (and not merely for positive reinforcement);
- (e) feeling remorseful about one's own past deeds and not excusing oneself; and
- (f) taking pride in having done what one thinks one ought to have done.³⁶

³⁴. As this indicates, the utilitarian should leave open the possibility that, although the punishment of innocents is unjust, in some circumstances it is permissible, or even obligatory, to override the dictates of justice. Interestingly, McCloskey himself endorses this position — see his «A Non-Utilitarian Approach to Punishment,» p.251.

³⁵. Jonathan Bennett, «Accountability», in Zak van Straaten (ed.), *Philosophical Subjects: Essays Presented to P.F. Strawson* (Oxford: Clarendon Press, 1980), p.20, emphasizes in the original.

³⁶. Cf. Richard Brandt's similar list in «Determinism and the Justifiability of Moral Blame,» in Sidney Hook (ed.), *Determinism and Freedom in the Age of Modern Science* (New York: New York University Press, 1958), p.140.

Despite the initial appeal of such objections, there are various ways of accommodating the above facets of human life within a hard determinist worldview.

Let's look first at (e), feelings of remorse and guilt.³⁷ It is plausible to think that if someone were deprived of such feelings, he would be incapable of mending any relationships with people that he has hurt. In addition, he would lack any motivation, after having done something wrong, to restore his own moral integrity and thus to develop morally. And so hard determinism, insofar as it prevents one from feeling blameworthy and hence guilty, undermines our ability to form good interpersonal relationships and to mature as moral beings.

This may seem plausible, but in the final analysis it is mistaken. As many hard determinists have insisted, if you perform a morally wrong act and yet you reject the claim that you are blameworthy (perhaps because you subscribe to hard determinism), you may nevertheless feel profound sorrow and regret that you were the agent of wrongdoing. And such feelings of intense sadness and regret are just what an attitude of remorse ordinarily amounts to. Pereboom in this context quotes an insightful passage from Hilary Bok, who employs an analogy with heartbreak to capture the idea that feelings of guilt may have little to do with feelings of blameworthiness:

The relation between the recognition that one has done something wrong and the guilt one suffers as a result...is like the relation between the recognition that one's relationship with someone one truly loves has collapsed and the pain of heartbreak. Heartbreak is not a pain one inflicts on oneself as a punishment for the loss of love; it is not something we undergo because we deserve it... Similarly,...the recognition that one has done something wrong causes pain. But this pain is not a form of suffering that we inflict on ourselves as a punishment but an entirely appropriate response to the recognition of what we have done.³⁸

But as Pereboom points out, even if guilt cannot be analysed solely in terms of feelings of sadness and regret, these feelings are likely to generate the same response as the experience of guilt: e.g., a resolution not to perform the immoral action again, an attempt to make amends by seeking to alleviate the suffering caused to other people, offering expressions of remorse towards those that have been wronged, etc.³⁹ Thus, the phenomena of guilt and remorse do not provide us with any reason to think that hard determinism threatens interpersonal

³⁷. Admittedly, (e) above only mentions feelings of remorse, not feelings of guilt, and the two should not be confused. There are, however, obvious connections between the two; hence my decision to discuss the two together.

³⁸. Hilary Bok, *Freedom and Responsibility* (Princeton, NJ: Princeton University Press, 1998), pp.168-69; quoted in Pereboom, *Living Without Free Will*, p.206. See, however, Ben Vilhauer's criticisms of Bok's analogy (on pp.557-58 of Vilhauer's «Hard Determinism, Remorse, and Virtue Ethics»), which lead Vilhauer to a different analogy, one which explains remorse on the model of suffering in sympathy with friends and loved ones (he writes, for example, that «sympathetic suffering prompts us to try to relieve the suffering of the friend or loved one, much as remorse prompts wrongdoers to try to make amends» [p.551]).

³⁹. See Pereboom, *Living Without Free Will*, p.206. Ben Vilhauer plausibly notes that remorse does not always have such beneficial behavioural consequences, and when it doesn't, the hard determinist will be unable to appeal to consequentialist considerations in justifying the appropriateness of feeling remorse. The hard determinist, furthermore, does not have the option of invoking the notion of desert, since the hard determinist cannot say that the remorseful person deserves to suffer emotionally on account of their wrongful actions. Vilhauer, however, contends that the option remains of appealing to virtue ethics. As he puts it,

[P]art of the solution to our problem about remorse is to stop thinking of it as punishment and, instead, to think of it as just one member of a set of emotional engagements, all of which depend upon suffering in sympathy with others, and none of which involve desert. Like love and friendship, remorse is a virtuous state of character that depends upon suffering in sympathy with people one cares about. («Hard Determinism, Remorse, and Virtue Ethics,» p.552.)

relationships and personal moral development.⁴⁰

Similar things can be said about the other items on the above list of (a)-(f). Consider, for example, the attitudes of indignation, disgust, and contempt as mentioned in (a). These are what P.F. Strawson has called ‘participant reactive attitudes’, that is, attitudes to which people are subject by virtue of engaging in ordinary interpersonal relationships. The attitudes are reactive in the sense that they are «natural human reactions to the good or ill will or indifference of others towards us, as displayed in *their* attitudes and actions.»⁴¹ Other reactive attitudes, beyond those given in (a), include gratitude, moral resentment, forgiveness, and love.

Some of these reactive attitudes, particularly love, are crucial to forming good and fulfilling relationships, but they are not clearly threatened by hard determinism. On the other hand, a hard determinist may have to concede that some reactive attitudes — such as indignation and gratitude — cannot be accommodated in a fully deterministic world. But as Pereboom points out, the attitudes that are incompatible with hard determinism typically fall into one of two categories: (i) they are not essential (or are even detrimental) to good relationships — resentment and contempt are cases in point; or else (ii) they have analogues or substitutes which lack any presupposition of moral responsibility while offering whatever is required for sustaining interpersonal relationships⁴² (for example, happiness and joy could play the same role as gratitude without involving any belief that some person is morally responsible or praiseworthy for some beneficial act that has brought about one’s contented state⁴³).

As for (c), making wrongdoers suffer for their past deeds, the hard determinist should not find it difficult to develop a reasonable position on managing and punishing criminal behaviour. A hard determinist, however, will not be able to justify punishment in retributive terms. The retributivist position sanctions the punishment of a wrongdoer on the grounds of desert: the wrongdoer deserves to have something bad happen to him — e.g., pain, deprivation, death — simply because he has done wrong. But hard determinism is incompatible with the notion of moral desert and must therefore renounce the retributivist justification for punishment. There are, nonetheless, a number of other options available to the hard determinist. For example, punishment might be endorsed by the hard determinist as a way of morally educating the wrongdoer or, alternatively, as a way of deterring the criminal and other prospective criminals from committing crimes. Even if these justifications for punishment are

⁴⁰. Similar views on guilt are expressed by Henry Sidgwick, *The Methods of Ethics*, p.71, and Bruce Waller, *Freedom Without Responsibility*, pp.164-69. Waller’s discussion of guilt is set within ch.12 of his book, where he provides some very good reasons for thinking that the denial of moral responsibility does not entail the rejection of all genuine morality.

⁴¹. P.F. Strawson, «Freedom and Resentment,» in his *Freedom and Resentment and Other Essays* (London: Methuen & Co., 1974), p.10, emphasis in the original. Strawson divides the reactive attitudes into those that are associated with demands on others for oneself (‘participant’ or ‘personal’ reactive attitudes), those that are associated with demands on others for others (‘generalised’ or ‘vicarious’ reactive attitudes), and those that are associated with demands on oneself for others (‘self-reactive attitudes’) — see pp.14-15 of Strawson’s «Freedom and Resentment». The argument of the main text is not restricted to any one of these three varieties of the reactive attitudes. (I should note that Strawson, like most compatibilists, does not think that determinism threatens merit-based moral responsibility, whereas I am assuming for the sake of argument that it does.)

⁴². See Pereboom, *Living Without Free Will*, pp.199-204.

⁴³. Cf. Pereboom, *Living Without Free Will*, pp.200-201, where a slightly different view on this matter is presented.

morally objectionable, they are not excluded by hard determinism. And even if all punitive measures against criminals were ruled out by hard determinism, there are a range of non-punitive or rehabilitative measures available, and these (as Pereboom indicates) may be more preferable and humane in any case.⁴⁴

It is not obvious, therefore, that hard determinism imperils the emotions and attitudes that underpin our social and moral lives, or that it jeopardizes our ability to respond to human evil in various punitive and non-punitive ways.⁴⁵ In fact, far from constituting a threat to the flourishing of human life, hard determinism might enable us to achieve a more fulfilling, meaningful, and moral life.

Compatibilists, going all the way back to the Stoics, have often promoted determinism on account of its perceived ethical advantages. Pereboom adopts a similar position, though from a hard incompatibilist perspective: «If we did in fact relinquish our presumption of free will and moral responsibility,» he writes, «then, perhaps surprisingly, our lives might well be better for it.»⁴⁶ Part of Pereboom's case in support of this view is predicated on the idea that hard determinism compels us to abandon expressions of moral anger, that is to say, the kind of anger that is directed toward someone who is represented (accurately or inaccurately) as having done wrong. Moral anger and closely related attitudes (such as indignation and contempt) are normally justified by way of the claim that the wrongdoer deserves blame in virtue of having committed some moral offense. This way of legitimizing moral anger, however, is clearly unavailable to the hard determinist. To be sure, the hard determinist may find a purely instrumentalist rationale for expressions of moral anger, while holding at the same time that the belief in moral desert that such anger presupposes is in fact false (as psychologists often tell us, false beliefs can sometimes be quite useful). Another option, however, is to jettison moral anger altogether, and this is the option Pereboom canvasses. This may provide substantial benefits, as we would be released from the harmful effects of moral anger — in particular, the physical and emotional pain it causes, and the damage it wreaks on our most cherished relationships — and this, in turn, would provide us with an increased sense of equanimity when we are wronged or treated unjustly.

But wouldn't the loss of moral anger engender an unwelcome degree of passivity? Moral anger motivates us to fight injustice and to address problems in our relationships with others. Without moral anger, therefore, there is a danger that we would make ourselves vulnerable to all manner of injustice, oppression, and abuse. Pereboom suggests a way of responding to this difficulty:

When the assumption that wrongdoers are blameworthy is withdrawn for hard incompatibilist reasons, the conviction that they have in fact done wrong could legitimately survive. Not implausibly, such a moral conviction could lead to a firm resolve to resist oppression, injustice, and abuse. As a result, hard incompatibilism might allow

⁴⁴. See Pereboom, *Living Without Free Will*, ch. 6.

⁴⁵. It may also be argued that even if hard determinism imperils certain aspects of commonsense morality, this may be offset by the resources hard determinism has in resolving some problems lying at the heart of our everyday moral thinking. Michael Slote takes such a view, arguing that the phenomenon of moral luck gives rise to a nest of inconsistencies within commonsense morality, but this can be overcome by going beyond our ordinary moral thinking and adopting a utilitarian account of moral responsibility. See Slote, «Ethics Without Free Will,» pp.372-75.

⁴⁶. Pereboom, *Living Without Free Will*, p.213.

for the benefits that moral anger may also produce, while avoiding its destructive consequences.⁴⁷

Perhaps, then, one's moral convictions (for example, one's conviction that something morally wrong has happened) in conjunction with the attendant feelings (such as feelings of hurt or alarm, particularly when one is personally affected by the morally wrong action in question) would be sufficient to motivate one to act to put things right. How plausible this is will depend in part on whether such convictions can be rationally upheld in a hard determinist world, an issue to be addressed in the ensuing Sections.

Saul Smilansky also highlights the ethical advantages of hard determinism. On his view, «the acceptance of hard determinism appears to create the possibility of a 'purer' morality than traditional free will based morality.»⁴⁸ Smilansky explains that traditional or commonsense morality has a dual nature: (i) a 'substantive' part, according to which certain things morally ought to be done (or not done), and (ii) an 'accountancy' part, according to which people ought to get positive reactions such as praise when they do things that morally ought to be done (and negative reactions such as blame when they do things that morally ought not to be done). If we accept hard determinism, we will of course have to give up the accountancy part of morality. But this, Smilansky writes, would be liberating and not destructive, as it would present us with a 'purer' morality. For given that the hard determinist agent chooses to do the right thing for its own sake, without any hope for praise or any fear of blame, he would be

...in one very curious sense a 'purer' moral agent than someone who is concerned with both parts of morality. For the hard determinist agent of the sort we are considering is not concerned with himself at all, he is solely focused on determining what he ought to do. By contrast, the 'non-hard-determinist' moral agent is inherently concerned with himself, with what action would make him praiseworthy or blameworthy in the eyes of others or of himself... The hard determinist agent we are considering enquires after the right thing to do and goes on to do it *despite* realizing that he is not praiseworthy for doing so and would not be blameworthy were he to act differently... If you like, he does not care about saving his moral soul, but about morality alone.⁴⁹

This I find a very attractive view. Pereboom, on the other hand, finds it unconvincing. He notes, first, that «one might doubt Smilansky's claim that for an agent who believes in moral responsibility, a concern with one's blameworthiness or praiseworthiness is always present.»⁵⁰ To be fair to Smilansky, however, following the above-quoted passage he goes on to note that someone who believes in moral responsibility need not be egoistic: «When one wants to be a decent human being and act like a creature worthy of the attribution of moral worth it would normally be misleading to call one egoistic.»⁵¹ Nevertheless, Smilansky continues to maintain that, with respect to the believer in moral responsibility, «a concern with

⁴⁷. Ibid., p.212.

⁴⁸. Smilansky, «The Ethical Advantages of Hard Determinism,» *Philosophy and Phenomenological Research* 54 (1994), p.356. See also Smilansky, *Free Will and Illusion*, ch. 10, where some further benefits of hard determinism are identified. It should be borne in mind that, according to Smilansky, these benefits of hard determinism are but a small consolation in the face of the disastrous personal, moral and social impact of accepting the absence of libertarian free will.

⁴⁹. Smilansky, «The Ethical Advantages of Hard Determinism,» pp.358-59, emphasis in the original.

⁵⁰. Pereboom, *Living Without Free Will*, p.152.

⁵¹. Smilansky, «The Ethical Advantages of Hard Determinism,» p.358.

the self's moral stature is...at least tacitly present.»⁵² I think Smilansky has overstated the mark, as it would have sufficed for his purposes to make the weaker but more plausible claim that a person who believes they are morally responsible is much more likely to exhibit the relevant kind of self-concern than their hard determinist counterpart. This view is entirely compatible with Pereboom's claim that «it seems possible that one believes in moral responsibility, but that this belief plays no role in one's moral decision-making.»⁵³

Pereboom, however, voices a second criticism of Smilansky's view:

For the hard incompatibilist agent, an analog to concern for one's praiseworthiness or blameworthiness seems as likely to be present as does this concern itself for the agent who believes in moral responsibility. This analog can also be characterized as a concern that one be a morally worthy person.⁵⁴

Pereboom goes on to add that, in a hard determinist world, an agent's moral worth would be determined by, for example, the degree to which they regulate their behaviour in accordance with moral reasons, the degree to which they are disposed to examine their past behaviour from the moral point of view, and their willingness to change their behavior when tendencies to immorality are recognized. This kind of moral worth, on Pereboom's view, would be just as much the concern of the hard determinist agent as would be the analogous kind of moral worth for the believer in moral responsibility.

Pereboom may be correct in holding that the hard determinist need only sever the connection between an agent's moral worth and their degree of praiseworthiness, rather than abandoning the notion of an agent's moral worth altogether. What Pereboom overlooks, however, is Smilansky's point that, without the accountancy side of morality, «there is one less major 'morally impure' source of motivation for acting as one should.»⁵⁵ In other words, the hard determinist agent, although concerned with their moral worth, will be motivated by, for example, the desire to become a better person or to promote the good. The believer in responsibility, on the other hand, is more likely to be motivated by more selfish reasons: viz., their wish to be rewarded or praised. Both the hard determinist and the non-hard determinist may display an equal degree of concern with their respective self's moral stature, but what is motivating them in each case is markedly different.

Smilansky, however, controversially assumes that the hard determinist can retain the 'substantive' aspect of morality, which includes such notions as moral obligation, moral rightness and wrongness, and moral goodness. Pereboom, as we saw earlier, makes a similar assumption when promoting the renunciation of moral anger as overall beneficial. But can the assumption that the hard determinist may reasonably retain a conviction in obligation, rightness, and goodness be justified? It is to this problem that I now turn.

⁵². Ibid.

⁵³. Pereboom, *Living Without Free Will*, p.152.

⁵⁴. Ibid., p.153.

⁵⁵. Smilansky, «The Ethical Advantages of Hard Determinism,» p.359.

III. Moral Obligation, and Moral Wrongness and Rightness

What happens to ‘ought’ judgements, and to our judgements that some actions are morally wrong while others are morally right, in a hard determinist world? It is commonly thought that the hard determinist must dispense with such judgements. Arguments in support of this view vary, but here is one inspired by Ishtiyaque Haji:⁵⁶

- (7) Assume that action A is something that person S does in fact perform.
- (8) **K**: Person S has a moral obligation to perform [not to perform] action A only if S can perform [refrain from performing] A.

Premise (8), or principle **K** (after its best-known exponent, Kant), is the ‘ought implies can’ principle, the principle that one ought morally to do A only if one can do A. Although the principle is intuitively plausible, it has become very contentious in virtue of the varying interpretations of ‘can’ that have been offered.⁵⁷ Haji has helpfully outlined three versions of this principle, each encapsulating a difference sense of ‘can’:⁵⁸

Weak Version

Agent S ought to do A only if S can do A, where this entails that:

- (i) S has the opportunity to do A — i.e., nothing prevents or would prevent S from successfully exercising the relevant abilities (whatever they are) to do A; and
- (ii) S is physically able to do A.

The first condition, relating to opportunity, expresses the idea that a person can do something only if they are ‘relevantly situated’ or have the cooperation of the environment. I can save a drowning child, for example, only if I can perceive that there is a child who is struggling to keep afloat, or only if the waters are not infested with sharks which would eat the child before I even jump into the water. The second condition, relating to physical ability, is regularly associated with uses of ‘can’ and ‘cannot’, as when we say that James cannot read the traffic sign because he is blind.

The problem with conditions (i) and (ii) is that they fail to capture some important uses of ‘can’ that are relevant to judgements of moral obligation. If, for example, you succeed in doing A but your accomplishment of A was strictly beyond your control — because your accomplishment was, say, a matter of sheer luck or predetermined by God — we would be reluctant to say that you could have done A in an obligation-conferring sense (that is, in a sense in which your ability to do A confers on you either the obligation to do A or the obligation not to do A). We might therefore opt for a stronger ‘ought implies can’ principle:

⁵⁶. The abbreviations of the moral principles outlined below, such as **K** and **OW**, are taken from Haji and Pereboom. Haji’s argument is developed in *Moral Appraisability: Puzzles, Proposals, and Perplexities* (New York: Oxford University Press, 1998), ch. 3, and in «Moral Anchors and Control,» pp.182-88.

⁵⁷. The controversial status of the principal that ‘ought implies can’ is also due to differing views as to what kind of implication relation holds between ‘ought’ and ‘can’. On this, see Walter Sinnott-Armstrong, «‘Ought’ Conversationally Implies ‘Can’,» *Philosophical Review* 93 (1984), pp.249-61. For a spirited defence of principle **K** and related principles, see Haji, *Deontic Morality and Control*, chs 2-4.

⁵⁸. See Haji, «Moral Anchors and Control,» pp.177-82, 202, and *Deontic Morality and Control*, pp.16-24.

Moderate Version

Agent S ought to do A only if S can do A, where this entail that:

- (i) S has the opportunity to do A — i.e., nothing prevents or would prevent S from successfully exercising the relevant abilities (whatever they are) to do A;
- (ii) S is physically able to do A; and
- (iii) A's accomplishment is not 'strictly out of S's control'.

Some have thought, however, that this version of the principle is also inadequate, since there is no provision made for various forms of knowledge — specifically, 'know-how' and 'knowledge that' — which undergird some morally relevant uses of 'can'. We say, for example, that James ought to save the drowning child, given that he can do so — and by this we mean, in part, that James knows how to swim and is not ignorant of the fact that the child is on the verge of drowning. We might therefore be led to an even stronger version of the 'ought implies can' principle:

Strong Version

Agent S ought to do A only if S can do A, where this entails that:

- (i) S has the opportunity to do A — i.e., nothing prevents or would prevent S from successfully exercising the relevant abilities (whatever they are) to do A;
- (ii) S is physically able to do A;
- (iii) A's accomplishment is not 'strictly out of S's control';
- (iv) S has the relevant 'know-how' to do A; and
- (v) S is not ignorant of any germane facts (the 'know-that' requirement).

Thankfully, for present purposes I do not need to decide which of these versions best accounts for the sense of 'can' in 'ought implies can' judgements. I will proceed, therefore, with the next step of the argument, which describes a standard principle of moral obligation:

- (9) **OW**: Person S has a moral obligation to perform [not to perform] action A if and only if it is morally wrong for S not to perform [to perform] A.⁵⁹

Premises (8) and (9) entail:

- (10) **WAP**: It is morally wrong for S to perform [not to perform] A only if S can refrain from performing [perform] A.

The plausibility of **WAP** may be highlighted by means of familiar Frankfurt-type cases. Suppose, for example, that I decide of my own accord to kill the President. Unbeknownst to me, however, Mad Scientist has the power to control my mind, so that if I were to hesitate to carry out my intentions, he would step in and stimulate parts of my brain in order to cause me to commit the crime. As it turns out, I carry out my plan without any hesitation, thus

⁵⁹. Although Haji considers both **K** and **OW** intuitively plausible, he also claims that these principles enjoy theory-based support, since both figure as theorems within (what Haji takes to be) some of our best current theories of moral obligation — for example, those developed by Michael Zimmerman in *The Concept of Moral Obligation* (Cambridge: Cambridge University Press, 1996) and Fred Feldman in *Doing the Best We Can* (Dordrecht: D. Reidel, 1986).

requiring no intervention from Mad Scientist. Nevertheless, I could not have acted otherwise; for instance, I could not have dropped the gun at the last minute, rather than pull the trigger. But given that I could not have refrained from killing the President even if I had wanted to, it seems that what I did was not wrong. Underlying this view, as Haji explains, is the idea that

Just as it is eminently reasonable to suppose that responsibility requires control — no one, for example, can be morally responsible for an action if one does not have «responsibility-grounding control» over the action — so it is reasonable to suppose that no one can perform an action that is morally right, or wrong, or obligatory unless one has appropriate «deontic-grounding control» over it.⁶⁰

If, therefore, I cannot refrain from killing the President, then I cannot exercise the requisite kind of control over my behaviour, and for this reason my behaviour cannot be deemed morally wrong.

The next premise draws out an obvious implication of the doctrine of hard determinism:

(11) **D:** If hard determinism were true, then no-one can do otherwise, that is, no-one can refrain from doing what they in fact do.⁶¹

The remaining steps of the argument then proceed as follows:

(12) If hard determinism were true, then it is not the case that it is morally wrong for S to perform A — that is to say, S's actions can never be judged to be morally wrong. (From 10 and 11)

(13) If hard determinism were true, then it is not the case that S has a moral obligation not to perform A — that is to say, S is never morally obligated to do otherwise than what S in fact does.⁶² (From 9 and 12)

What this argument indicates, assuming it to be sound, is that our commonsense system of morality — and in particular our commonly held beliefs that at least some of our actions are morally wrong and that in at least some cases we ought to have acted otherwise — are undermined by hard determinism. This does not mean that all our judgements regarding moral obligation and moral rightness or wrongness must be jettisoned. But it does mean that much of what we hold dear and common in the domain of morality is mere illusion.

I will now consider the attempt made by Pereboom to resist the conclusions of the above argument (or at least conclusion (12), as he is quite content to let moral obligation go⁶³). But before doing so, it may be helpful to briefly consider one possible line of response that those who are fond of compatibilism may be inclined to make. It may be held that a compatibilist

⁶⁰. Haji, *Deontic Morality and Control*, p.3.

⁶¹. I should note that Haji's argument is targeted against *soft* determinism, though I think no damage is done to his argument if it is steered towards hard determinism.

⁶². A further consequence is that, since you can never be morally obligated to act otherwise, there would be no point in conferring any moral obligations on you.

⁶³. Given that the above argument is applicable to both moral obligation and moral wrongness/rightness, it is somewhat strange that Pereboom concedes the argument's force in relation to the former category while disputing its cogency in relation to the latter category. I may note, however, that Pereboom has recently reported (via personal correspondence) that he now thinks that determinism only undermines some, but not all, uses of 'ought' (a view prefigured somewhat on pp.147-48 of Pereboom's *Living Without Free Will*).

reading of ‘S could have done otherwise’ would remove any reason to think that determinism (whether hard or soft) is incompatible with obligation and rightness or wrongness. According to compatibilists, to say that S could have done otherwise is just to assert a hypothetical of the form, ‘S would have done otherwise if S had desired, chosen, or willed to do otherwise’. But if ‘can’ is understood along these lines, then — *contra* (11) — people can do otherwise even in a hard determinist world.

Haji, however, demurs, and illustrates his objection with a Frankfurt-type case similar to that described earlier.⁶⁴ In a determinist world, he notes, we would all be in the same boat as the killer in the Frankfurt-type scenario because, like him, (i) we would not be able to do otherwise than what we in fact will to do, *and* (ii) we would not do otherwise even if we had so willed. Thus, even in the compatibilist’s sense of ‘could have done otherwise’, it remains the case that in a determinist world we could not act otherwise. But this does not seem right, for there is no reason to think that agents in a (hard or soft) determinist world would always find themselves in Frankfurt-style scenarios. And if that is so, then it follows that, even though in a determinist world we could not do anything other than what we in fact do, it remains possible that *had we willed to do something other than what we in fact did, we would have done something other than what we in fact did.*⁶⁵ In any case, I will let this criticism of premise (11) pass, and focus instead on the other steps of the argument.

Pereboom’s general line of response to Haji’s argument presented in (7)-(13) above is that the central premises of the argument — viz., principles **K** and **OW** — cannot plausibly be thought of as having «a justificatory status so strong that it immunizes them against disconfirming pressures from their unintuitive consequences.»⁶⁶ He goes on to explain that, «If the components of the [moral] theory derived from these principles [viz., **K** and **OW**] conform to our intuitions, that would provide theoretical support for them. But if such derived components do not conform to our intuitions, that would to some extent disconfirm these principles.»⁶⁷ To be sure, it is not (**K** & **OW**) alone that produces the counterintuitive results that none of our actions are morally wrong and that we never ought to have acted otherwise. Rather, it is (**D** & **K** & **OW**) that produces these results. So, how is Pereboom to be understood here? As will be seen shortly, Pereboom’s view is that **K** and **OW** are not as epistemically secure as might be thought at first glance, and that once we consider matters more deeply we will have no more reason to reject the claim that some of our actions are morally wrong as to accept this claim. Pereboom offers three arguments in support of this view.

His first argument begins with the following intuitively true principle, connecting moral goodness with moral rightness:

(14) **GR**: Sometimes, actions that bring about the greatest good overall in worlds accessible to S are right for S.

⁶⁴. See Haji, «Moral Anchors and Control,» p.189.

⁶⁵. I must confess that I was initially persuaded by Haji’s objection, but Pereboom (in personal correspondence) brought me to my senses.

⁶⁶. Pereboom, *Living Without Free Will*, p.144.

⁶⁷. *Ibid.*

Note, first, that Haji's argument has nothing explicitly to say about moral goodness (and moral badness). Thus, as far as the considerations advanced by Haji go, we have no reason to think determinism to be incompatible with moral goodness, though we are given some reason to think that determinism rules out judgements of moral wrongness and rightness (or at least a large class of our ordinary judgements regarding moral wrongness and rightness). This, however, commits Haji to an invidious demarcation between judgements about moral goodness and judgements about moral rightness. In other words, if assume the truth of **D** for the sake of argument, and if we follow Haji in accepting (**K** & **OW**), then we must also accept the following unintuitive result:

(15) **GR** is false.⁶⁸

But, Pereboom adds, it is far from clear that Haji's position, as laid out in (I) below,

(I) Accept: **D** & **K** & **OW** & goodness & blameworthiness

and therefore reject: obligation & rightness/wrongness & **GR**

is at all superior to a view (like Pereboom's) that bids us to

(II) Accept: **D** & **K** & rightness/wrongness & goodness & **GR**

and therefore reject: blameworthiness & obligation & **OW**.

Each position has unintuitive results, but it is not obvious that position (I) is in better shape than (II). And so, at the very least, we are faced with a kind of détente. This problem, of course, would not arise if you thought that, in a hard deterministic world, there can be no moral goodness, and not just no moral rightness or wrongness. But as I hope to show in the following Section, judgements of moral goodness are not undermined by hard determinism. This argument of Pereboom's, then, has some degree of force.⁶⁹

Pereboom's second argument is structurally similar to his first argument, but targets Haji's view that blameworthiness survives determinism while wrongness does not. According to Haji, the principle that an agent is blameworthy for performing an action only if that action is in fact wrong, despite its intuitive plausibility, is seriously mistaken. He offers a number of considerations in support of this view, including hypothetical cases such as the following (which he dubs 'Deadly's Defeat'):

In Deadly's Defeat, doing the best she can to murder her patient, Dr. Deadly does what credible evidence to which she has access indicates will kill the patient — she injects the patient with medicine C. But...happily for the patient,

⁶⁸. Pereboom notes that what holds for **GR** also holds for the following intuitively true principle:

GR': There is some correlation between an action's bringing about the greatest good overall in worlds accessible to S and that action's being right for S.

See Pereboom, *Living Without Free Will*, p.145, fn. 45.

⁶⁹. Haji has replied that Pereboom's view that **GR** is true is question-begging:

[I]f one finds **GR** intuitively plausible, presumably one does so independently of having reflected on the view that there is a requirement of alternative possibilities for rightness, wrongness, and obligatoriness, and hence that determinism is incompatible with deontic morality. It is not clear, consequently, that having engaged in the relevant sort of reflection, one would still find **GR** intuitively attractive. (*Deontic Morality and Control*, p.71)

In other words, **GR** appears plausible only when principles such as **K** and **OW** are overlooked or presupposed to be false. But I doubt that one needs to reject or overlook **K** and **OW** in order to appreciate the intuitive plausibility of **GR** (though this is not to say that **GR** ought to be accepted as true). I may add here that Haji does not reject **GR** outright, but thinks that **GR** inchoately expresses the truth of the matter, which is better expressed in terms of Zimmerman's 'world utilitarian' principle (stated on p.72 of Haji's *Deontic Morality and Control*).

there has been an innocent error in diagnosis. Contrary to what Deadly believes, her patient is suffering from a disease that can be cured only by taking C. Giving C results in the lucky patient's full recovery. Although it is arguably obligatory for Deadly to inject her patient with C, it appears that Deadly is to blame for the injection of this drug.⁷⁰

Cases such as this, according to Haji, strongly indicate that moral blameworthiness should be kept apart from objective wrongness and objective obligatoriness: Dr Deadly's behaviour may be open to blame, but he has done (albeit inadvertently) what he ought to have done and what was right for him to do. Haji therefore endorses a 'subjective view' of blameworthiness (and praiseworthiness), which roughly states that an agent is blameworthy for performing an action only under the condition that she has the belief that it is wrong for her to perform that action and this belief plays an appropriate role in the performance of that action — the belief in question, however, need not be epistemically justified, let alone true.⁷¹ This allows Haji to hold that, in a deterministic world, people can be blameworthy for what they do, even though their actions can never be morally wrong or obligatory. Pereboom, however, complains that Haji's view has the paradoxical result that when agents are blameworthy in a deterministic world, it is never because they have done something wrong. This, according to Pereboom, conflicts with another intuitively plausible moral principle, viz.:

(16) **BW**: Sometimes, when S is blameworthy for performing A, it was morally wrong for S to perform A.

And so, as before, if we assume the truth of **D** for the sake of argument, and if we follow Haji in accepting (**K** & **OW**), then we must also accept the following unintuitive result:

(17) **BW** is false.

And, once again, there is little reason to think that Haji's position,

(I*) Accept: **D** & **K** & **OW** & goodness & blameworthiness
and therefore reject: obligation & rightness/wrongness & **BW**

is preferable to the view that we should

(II*) Accept: **D** & **K** & rightness/wrongness & goodness & **BW**
and therefore reject: blameworthiness & obligation & **OW**.⁷²

⁷⁰. Haji, *Moral Appraisability*, p.146. Haji adds that we are to «construe this case as one in which Deadly is under no delusion that she takes herself to be doing wrong when she injects the patient with C» (p.152). There is a problem with the case of Deadly's Defeat that seems to have gone unnoticed. Haji assumes that it is obligatory for Deadly to inject her patient with medicine C. But given that the doctor is (non-culpably) ignorant of the fact that the patient requires C in order to be cured, how can the doctor have an obligation to administer C to the patient? This suggests that, in order to determine what is objectively right or obligatory for a person, we must take into account that person's epistemic predicament (e.g., what facts are accessible to that person).

⁷¹. See Haji, *Moral Appraisability*, chs 8 and 9, and *Deontic Morality and Control*, ch. 10.

⁷². Haji claims, however, that Pereboom is not entitled to accept **BW**. Haji first distinguishes two varieties of **BW**: **BW1**: Sometimes, when S is blameworthy for performing A in a deterministic world, it was morally wrong for S to perform A in that world. **BW2**: Sometimes, when S is blameworthy for performing A in an indeterministic world, it was morally wrong for S to perform A in that world. Clearly, as Haji observes, Pereboom is only concerned with **BW1**. Haji then goes on to state that a hard incompatibilist of the Pereboom variety «should reject this principle [i.e., **BW1**] (even though he might find it intuitive) because hard incompatibilism entails that both determinism and its denial are incompatible with blameworthiness» (*Deontic Morality and Control*, p.73). But this is mistaken, for Pereboom (unlike Haji) can accept **BW1**, understood

The first thing to note is that Pereboom's argument is only aimed at someone, like Haji, who thinks that in a deterministic world there can be blameworthiness and praiseworthiness (of the non-consequentialist, merit-based variety). Some support for Haji's view is provided by Frankfurt-type cases where the murderer appears to be morally responsible and deserving of blame, despite lacking the ability to have done otherwise. As noted in Section II above, Pereboom (correctly, I take it) disagrees with this assessment, for in his view no-one in a deterministic world is ever deserving of praise or blame. But what is important to note in the present context is that Haji's view that blameworthiness may survive in a deterministic world is not essential to the argument he develops, in (7)-(13) above, in support of the claim that determinism undermines some central concepts of deontic morality (in particular, moral obligation and moral wrongness). Pereboom, therefore, in attacking Haji's views on blameworthiness, fails to engage with the argument of (7)-(13).

In any event, there are problems with both Haji's view of blameworthiness and Pereboom's response to this view. To begin with Pereboom's reply, there is no reason why Haji must reject **BW**. On Haji's view, when a person S is blameworthy for performing some action A, it is never because S has done something that is objectively wrong; rather, it is (partly) because S has done something that is subjectively wrong (or wrong from S's perspective). But Haji can still agree that, in some cases, a person S is blameworthy for performing A *and* it was objectively wrong for S to perform A. Haji, therefore, can endorse **BW** (at least if the 'when' in **BW** is not read as 'because').

However, Haji's subjectivist account of blameworthiness is also problematic, as it seems to give insufficient weight to culpable ignorance. Consider, for example, a case where a Nazi officer kills a Jewish boy in front of its parents in the belief that it is right and obligatory to kill Jewish children. The Nazi's action would ordinarily be considered blameworthy, at least partly because he is culpable in forming the belief in question. In such cases, ignorance is no excuse: the offender ought to have known better and is therefore open to blame. But if Haji's account were correct, the Nazi officer could not be blameworthy since he does not believe that what he is doing is wrong. This, clearly, is a disappointing result of Haji's account.⁷³

Pereboom's third argument is intended to break the deadlock between positions (I) and (II), and their cognates, in favour of (II) and (II*). Pereboom's case here is built upon the premise that there is good reason to reject Haji's 'standard principle of moral obligation':

OW: Person S has a moral obligation to perform [not to perform] action A if and only if it is morally wrong for S not to perform [to perform] A.

as the claim that *If someone is blameworthy for doing A in a deterministic world, their performance of A in such a world is morally wrong*. For Pereboom, *qua* hard incompatibilist, this conditional would be true since its antecedent is false and its consequent is true.

⁷³. The disappointment is not removed by the defence offered by Haji against objections of this sort. In response to cases like that involving the Nazi officer, Haji writes that, «I strongly doubt that (in such cases) S ought (morally) to have realized that failing to acquire BF [the belief that torturing people for fun is wrong] is wrong, failed in this obligation, and was culpable for doing so» (*Moral Appraisability*, p.165; similar remarks are made in Haji, «On Psychopaths and Culpability,» *Law and Philosophy* 17 (1998), p.137). Unfortunately, Haji offers no explanation for the doubt he expresses here. It seems, however, that the doubt is quite misplaced, for most people would agree that (a) the Nazi officer ought to have realized that it was wrong of him to fail to acquire the germane belief (i.e., the belief that killing people solely on account of their race is wrong), (b) he failed in this obligation, and (c) was culpable for doing so.

OW, being a biconditional, consists of two distinct claims:

(**OW**₁) If S has a moral obligation to perform [not to perform] action A, then it is morally wrong for S not to perform [to perform] A, and

(**OW**₂) If it is morally wrong for S not to perform [to perform] A, then S has a moral obligation to perform [not to perform] A.

Pereboom focus on the bracketed version of **OW**₂, which reads as follows:

(**OW**₂*) If it is morally wrong for S to perform A, then S has a moral obligation not to perform A.

To produce a counterexample to **OW**₂*, one would need to devise a scenario in which it is morally wrong for you to do A and yet you do not have a moral obligation not to do A. Pereboom attempts to do just this by way of the following purported counterexample:

Suppose you say to an animal-abuser, «You ought not to abuse that animal,» but then you find out that he has a psychological condition (which he could have done nothing to prevent) that makes animal-abusing irresistible for him, so that he cannot help but abuse the animal. From my point of view, there is an appreciably strong pull to admitting that the «ought» judgment was false, but there is relatively little to denying that abusing the animal is morally wrong for him.⁷⁴

I am not convinced, as I doubt that moral wrongness can be entirely divorced from moral obligation. Indeed, it seems strange to say that, while having an obligation do something, A, presupposes having control over A-ing, its being wrong for one to do A does not. As Haji rightly asks, «Why the asymmetry in metaphysical presuppositions of control? After all, wrongness and obligatoriness are normative appraisals of the same family; they are deontic normative statuses and thus it would seem that the control-relevant presuppositions of the one should also be, barring cogent explanation to the contrary, those of the other.»⁷⁵ It seems unreasonable, then, to deem an action ‘morally wrong’ while denying that one *ought not* engage in such behaviour. And so if one has no moral obligations, there is nothing morally wrong that one can do. That is why describing the animal abuser’s actions as ‘morally wrong’ would be just as mistaken as describing a shark attack on a human being as ‘morally wrong’.

Summing up the results of this Section, it appears that only Pereboom’s first response to the Haji-inspired argument presented in (7)-(13) above has succeeded in providing us with some reason to doubt the claim that if hard determinism were true, none of our actions would be morally wrong. At best, therefore, we have just as much reason to believe that ascriptions of rightness and wrongness survive hard determinism as we have to reject this view. This is not an entirely positive result for the hard determinist, especially since no attempt has been made to respond to the above argument’s further claim that hard determinism is not compatible with moral obligation.

There are, however, various ways out for the hard determinist that seem to have gone unnoticed by both Pereboom and Haji. One such way out involves extending the consequentialist account of praise and blame, as defended in Section II, to our practice of making ‘ought’ judgements and judgements of rightness and wrongness. More specifically,

⁷⁴. Pereboom, *Living Without Free Will*, p.147.

⁷⁵. Haji, «Moral Anchors and Control,» p.185. See also Haji, *Deontic Morality and Control*, pp.52-53, including his fn. 12.

given a consequentialist reading of moral wrongness, **WAP** (i.e., premise (10)) would turn out to be false. For whether an action counts as morally wrong will depend on the (foreseeable) consequences of that action, perhaps in conjunction with some other conditions (such as that the action was performed intentionally). But the requirement, as expressed in **WAP**, that the agent be able to refrain from performing the action in question will play no part in the consequentialist story of moral wrongness. Alternatively, one may have grounds for rejecting **WAP** that do not rely on any particular theory in normative ethics — one could, for example, reject **WAP** on the grounds that the ‘ought implies can’ principle presupposed by **WAP** is false. But then a range of options become available to the hard determinist. Even a deontological account of moral rightness, wrongness and obligation can be endorsed by the hard determinist, for without **WAP** no threat will be posed to these moral categories by the absence of alternate possibilities. With the demise of **WAP**, however, the argument presented above in steps (7) to (13) collapses as well.

IV. Moral Goodness

The last moral category I will consider is moral goodness. Can moral goodness be accommodated in a hard determinist world? Pereboom thinks it can. He holds that even if, under hard determinism, ‘ought’ judgements are never true, it remains the case that judgements such as ‘It is morally good for person S to do A’ and ‘It is morally bad for S to do B’ can be true. He offers the following illustrations in support of this view:

Thus, for example, even if one is causally determined to refrain from giving to charity, and even if it is therefore false that one ought to give to charity, it still might be good to do so. Cheating on one’s taxes might be a bad thing to do, even if one’s act is causally determined, and thus, even if it is false that one ought not to do so.⁷⁶

At least one question this immediately raises is: What sense of ‘good’ is operative here? One obvious option, in the light of what has been said earlier, is to give ‘goodness’ a consequentialist reading. On this view, an action is considered morally good insofar as it has certain consequences (e.g., the greatest possible increase of pleasure over pain). But the hard determinist, as already mentioned, has a number of options available within normative ethical theory, even including a deontological understanding of moral goodness.

A slight problem, however, lurks here. Let’s suppose that we adopt a purely deontological moral framework. Let’s further suppose that, as Haji’s argument in the previous Section suggests, we are not entitled to think that moral obligation and moral rightness or wrongness (characterised in non-consequentialist terms) can be made to fit into a hard determinist world. Could we then say of at least some actions in such a world that they are *morally permissible* for some agent? In other words, if an action is not right, wrong, or obligatory, does it follow that it is permissible?⁷⁷ Haji considers this question and answers that it would be more plausible to say that actions that are not right, wrong, or obligatory are ‘amoral’ for a person, rather than permissible for a person. His reasoning in support of this view is that, if some of our actions are open to moral appraisal although they could never be wrong or obligatory, then each of our actions would either be amoral or morally permissible. But then if I were to kill someone in cold blood, my action would be, if not amoral, then morally permissible for me,

⁷⁶. Pereboom, *Living Without Free Will*, p.143.

⁷⁷. As indicated, the notions of ‘right’, ‘wrong’, ‘obligatory’, ‘permissible’, and ‘good’ are to be understood (in this and the following paragraph) in non-consequentialist terms.

«a result that is hard to swallow».⁷⁸

If Haji's reasoning were accepted, as I think it should, then we are faced with the following uncomfortable result: We can legitimately say of some action A that it is morally good, and yet we must deny that A is morally permissible. This sounds strange, however, only because we overlook the fact that judgements of permissibility are interchangeable with judgements of obligation (action A is not permissible for person S iff S ought not to do A; A is permissible for S iff it is not the case that S ought not to do A). That is why the collapse of our system of moral obligations also brings with it the end of our practice of judging actions to be permissible or impermissible. There should therefore be no hesitation in denying moral permissibility, given that we have already dispensed with moral obligation. But as pointed out by Pereboom, this need not lead us to reject moral goodness.

Concluding Remarks

Given the truth of hard determinism, the ordinary folk picture of moral responsibility (and hence of praiseworthiness and blameworthiness), moral obligation, moral rightness or wrongness, and moral goodness — since it presupposes a commitment to the notions of desert, the ability to do otherwise, and thus (compatibilist or libertarian) free will — must be uprooted, or at least revised in significant ways. The nature and function of judgements involving these moral categories must therefore be thoroughly reconceived. But even if there is much in our commonsense system of moral beliefs that must be uprooted or seen anew, the hard determinist is not without the resources needed to rebuild our moral edifice. We have seen, for example, that one among many options available to the hard determinist is to (re-)interpret moral judgements in broadly consequentialist fashion, so that a judgement that, say, a particular action is morally right is to be analyzed in terms of the overall consequences that would most likely result if that action were performed.

There is no reason to think, however, that this reorientation in our moral outlook brought about by hard determinism would have disastrous practical consequences for our lives. Despite initial appearances, the assumption that we lack the kind of free will required for moral responsibility does not threaten the emotions and reactive attitudes that are crucial to our social life and moral development. Indeed, as some hard determinists have plausibly argued, the rejection of responsibility-entailing free will holds out the promise of a life that is morally deeper (because less self-centred) and more fulfilling (because less prone to destructive

⁷⁸. Haji, «Moral Anchors and Control,» p.187. I might add that it will not help to bring the category of moral badness into the picture by holding that, even though the killing is not amoral, it remains impermissible because it is morally bad. For badness would then be intimately linked to obligation: to say that action A is bad would (in this case at least) entail that A is not permissible, which entails that one ought not to do A. The working assumption here, however, is that there are no moral obligations.

anger).⁷⁹

Nick Trakakis
Department of Philosophy
School of Philosophy & Bioethics
Monash University
Clayton, Victoria, 3800
Australia
Nick.Trakakis@arts.monash.edu.au

⁷⁹. For helpful comments on a previous draft, I must thank Daniel Cohen, Ish Haji, Saul Smilansky, and Derk Pereboom. I am also grateful for the feedback provided by audiences at Monash University and at the 2005 Conference of the Australasian Association for Philosophy (held at the University of Sydney), where an earlier version of this paper was presented.

ESSENTIAL DEPENDENCE AND REALISM

by Daniel Laurier

Let's start with the nearly trivial remark that a realist basically is someone who claims that there is «something» which exists/obtains independently of ourselves. It follows from this very minimal characterization that there are at least four main parameters to take into account in sorting out the various forms that the realism/irrealism disputes might take, namely: (i) the nature and extent of the «something» which is said to be «independent of ourselves», (ii) what it is for something to be «actual» (i.e., to exist, to obtain, or to be true) (iii) the nature of the relevant independence relation, and (iv) the nature of that of which the relevant reality is said to be independent, i.e. what is meant, in such contexts, by talking of «ourselves». In this paper, however, my main interest is in these forms of irrealist positions which grant that some of the states-of-affairs (or judgements) under consideration may obtain (be true) and deny only that they obtain (are true) «independently of ourselves». I will accordingly focus on the last two parameters that I have mentioned (and say next to nothing about the others).

Of what, exactly, is something said to be independent, when it is said to be independent of «ourselves»? This doesn't have to be construed as a question about the nature of selves or persons, or indeed, about the nature of any kind of entity (though it might be), for it will be readily admitted that when the realist claims that something exists/obtains independently of ourselves, he/she (most often) means to be saying not that it exists/obtains independently of the fact that we or «our minds» exist (which would mean that it exists/obtains independently of any mental property that we might have), but more modestly, that it exists/obtains independently of the fact that we have some specific mental feature or power M. As far as I can see (and historically, at least), there are three main candidates for playing the part of the designated mental feature M: (i) our capacity to know, (ii) our capacity to think or conceive and (iii) our capacity to perceive or experience.

Thus, in most cases, the realist who claims that something exists/obtains independently of ourselves is really claiming that it exists/obtains independently either (i) of our having the capacity to know it (i.e., of its being knowable by us), or (ii) of our having the capacity to conceive it (i.e., of its being conceivable by us), or (iii) of our having the capacity to experience it (i.e., of its being experientiable by us). This suggests that the selected mental feature or power M will more often than not be a relational one; that is, one which we are supposed to have *with respect to* a certain thing, in the sense that, e.g., while the state-of-affairs that P may be said to obtain independently of our possessing M *with respect to* P, some other state-of-affairs Q will be said to obtain independently of our possessing M *with respect to* Q. This will later prove to be of some relevance.

But for now, I want to consider how the notion of independence itself is to be construed. It will be convenient, for this purpose, to take *dependence* as the primitive notion, and to

understand independence as the denial of dependence (and hence as a non-symmetric relation). Thus, to say that something exists/obtains independently of our possessing mental feature M is just to say that (it exists/obtains and) its existence/obtaining doesn't depend on our possessing M. And to say that the existence/obtaining of something depends on our possessing M is to say that a certain state-of-affairs depends on another (or in the semantic mode, that the truth of some judgement depends on that of another). Thus, the notion of dependence must here be construed as a relation between states-of-affairs or judgements.

Two natural readings of the claim that the state-of-affairs (judgement) that P depends on the state-of-affairs (judgement) that Q are (i) the counterfactual one, according to which P depends on Q¹ iff (a) it would not be the case that P if it were not the case that Q and (b) if it were the case that Q it would be the case that P², and (ii) the modal one, according to which P depends on Q iff necessarily, if P then Q. In both cases, we are dealing with entailment relations: to say that P *counterfactually* depends on Q is to say that not-Q counterfactually entails not-P and Q counterfactually entails P, and to say that P *modally* depends on Q is to say that P necessitates Q (i.e., that not-Q necessitates not-P). As Lewis (1973: 167) remarks, when it is assumed that both P and Q are actually the case, P will counterfactually depend on Q just in case not-Q counterfactually entails not-P, since the other condition will trivially hold. On this assumption, modal dependence turns out to be strictly stronger than counterfactual dependence³.

At first sight, there would seem to be nothing to prevent either of these two construals of the dependence relation to be involved in some particular realist/irrealist dispute. But it has recently been argued (by Jenkins (2005)) that none of them is strong enough to capture what is «really» at stake in such disputes. According to Jenkins, Fact-realism about some subject matter *must* not be construed as the claim that the facts pertaining to this subject matter are *modally or counterfactually* independent from the mind, but as the claim that they are *essentially* independent from the mind, where *essential dependence* is taken to be stronger than either modal or counterfactual dependence⁴.

I'm now going to discuss this proposal in some detail. But it is worth mentioning at the outset that even if we were to grant Jenkins point that any «genuine» realist thesis must involve the claim that something is «essentially» independent of our possessing some mental feature, it would *not* follow that all forms of irrealism must similarly involve the claim that something is «essentially» dependent on our possessing some mental feature; for as was

¹. Saying that «P depends on Q» should be seen as a convenient way of saying either that the obtaining of the state-of-affairs that P depends on the obtaining of the state-of-affairs that Q, or that the truth of the judgement that P depends on the truth of the judgement that Q, as the case may be.

². This is a currently standard definition of counterfactual dependence, which derives from Lewis (1973: 167) and is to be found in several places such as Sanford (1989: 182) and Collins (2004: 109).

³. The claim that P necessitates Q is equivalent to the claim that in all possible worlds where Q is not the case, it is also not the case that P, while the claim that P counterfactually depends on Q, on the assumption that P and Q are both actually the case, comes down to the claim that in all the closest worlds where it is not the case that Q, it is also not the case that P.

⁴. Presumably, Jenkins would be prepared to extend this to Thing-realism; but since she doesn't discuss any form of Thing-realism, I'm going to ignore this in what follows. Though I think that the same points basically apply in this case as well.

pointed out above, there are forms of irrealism which oppose the realist's «actuality» claim without opposing his/her «independence» claim. Hence, it is to be understood that only what I call forms of «Independence-irrealism» are at issue in the following. That being so, it seems appropriate to assume that we'll be dealing only with states-of-affairs which do obtain (or are taken to obtain by both the realist and the irrealist), and thus with cases where modal dependence is strictly stronger than counterfactual dependence. The latter can accordingly be put aside in what follows, since it could not be «strong enough» unless modal dependence also is «strong enough»⁵.

Let me now make a simple observation about the dialectic of the debates opposing the realist and the irrealist. It would seem that anyone engaging in such a debate (or indeed, in any kind of theoretical debate) should strive to defend the strongest position one thinks is rationally defensible: the stronger the better. This means that the (Independence-)irrealist should aim at *establishing the strongest* form of mind-dependence he/she can (the stronger the form of mind-dependence he/she is able to establish, the better he/she should feel), while the realist should aim at *denying the weakest* form of mind-dependence he/she can (the weaker the form of mind-dependence he/she is able to deny, the better he/she should feel). If this is right then Jenkins' main contention, namely, that realism must be construed as involving not a denial of modal mind-dependence, but a denial of some stronger form of mind-dependence which she calls *essential* mind-dependence, can provide little more than a fall back position for any would-be realist. Accordingly, and correspondingly, it does challenge the (Independence-)irrealist to establish a stronger form of mind-dependence than any he/she may have wanted (or thought possible) to maintain.

I doubt that there is any clear and definite answer to the question which form of mind-dependence best captures the «true spirit» of all realist/irrealist theses. As I see it, various forms of mind-dependence are worth investigating, in various domains, and with respect to various mental features. It is quite obvious that one may well hold both that something modally (essentially) depends on our possessing some mental feature M, and that it doesn't modally (essentially) depend on some other mental feature M*. It should be no less obvious that just as there is nothing to prevent one from being both a counterfactual mind-dependence irrealist and a modal mind-dependence realist in a certain domain, there is nothing to prevent one from being both a modal mind-dependence irrealist and an essential mind-dependence realist in some other domain. The question whether someone in this position should be described as a «realist» or an «irrealist» *tout court* is largely a matter of taste and convenience.

That being said, it certainly makes sense to ask what is the strongest form of mind-dependence which could plausibly be maintained, relative to any given choice of (i) a kind of judgements or facts K and (ii) a specific mental feature M. And there is indeed some reason to think that the claim that something's being the case (the obtaining of some state-of-affairs) *modally* depends on our possessing mental feature M will not always succeed either in capturing the irrealist intuitions about the given subject-matter, or in defeating the corresponding realist intuitions. On the other hand, some forms of dependence might be so strong that virtually nothing could have this relation to anything other than itself (or to any

⁵. Where this notion could be useful is in the context of a *defence* of some realist thesis. For it would obviously be *sufficient*, in order to establish that some state-of-affairs doesn't depend (in some suitably strong sense) on our possessing M, to show it doesn't *counterfactually* depend on our possessing M.

mental feature thought to be relevant to the realism issue), thus threatening to trivialize any realist claim to the effect that something or other doesn't bear *this* relation of dependence to this or that mental feature. Clearly, to insist that the realist position be defined by its denial of any such strong dependence relation is to be unacceptably biased against irrealism. What I am going to argue is that, *if* the notion of essential dependence is going to be understood along the lines suggested by Jenkins, then appealing to it in this way runs a serious risk of expressing just such a bias.

Jenkins' characterization of the notion of essential dependence takes its inspiration from some work of Kit Fine (1994, 1995)⁶. It is, however, a matter of some importance that this work doesn't at all trade in the realism/irrealism business, but is concerned with finding an intuitively satisfying way of understanding a general notion of ontological dependence, conceived as a relation that some entity bears to some other entity, when the existence of the first can be said to depend on the existence of the second, or on the second's being a certain way.

From this perspective, it is easy to show that it would be inadequate to understand the claim that some entity A ontologically depends on some other entity B as meaning simply that necessarily, A exists only if B exists, or the claim that A ontologically depends on B's having a certain feature as meaning simply that necessarily, A exists only if B has this feature. For this would imply (i) that everything ontologically depends on every necessary existent, (ii) that any two things which necessarily co-exist ontologically depend on each other, and (iii) that nothing can ontologically depend on some other thing's being F without also depending on this other thing's having any feature necessitated by its being F (or for that matter, by its bare existence). But it is intuitively clear that one may want to deny that Socrates ontologically depends on the number two, even though the latter exists in all worlds where Socrates exists. It is just as clear that one may want to deny that Socrates ontologically depends on its singleton, despite the fact that the latter exists in all worlds where Socrates exists. Likewise, one may want to hold that X ontologically depends on Y's being F (and hence, on Y's existence) without being committed to the claim that X ontologically depends on Y's being F-or-G, despite the fact that Y is F-or-G in all worlds where it is F, or to the claim that X ontologically depends on Y's being G-or-not-G, despite the fact that Y is G-or-not-G in all worlds where it exists.

What this shows, is that the notion of modal dependence (as currently understood in the possible worlds framework) is too weak to account for some of our intuitions concerning ontological dependencies. As has been observed by several authors⁷, what must be captured, is the intuition that when some entity A ontologically depends on the fact that some entity B has a certain feature, it should be the case that it is *part of the nature or essence* of A that necessarily A exists only if B has this feature. This is the idea that ontological dependence should be understood in terms of some notion of essential dependence (i.e., of an entailment relation somehow *grounded* in essence).

This obviously has some bearing on issues concerning Thing-realism/irrealism, since the Thing-realist claims that the existence of entities of a certain kind depends on our possessing

⁶. Related work by Lowe (1994) and Correia (ms) is also relevant here, though Jenkins doesn't mention them. See also Mulligan, Simons and Smith (1984).

⁷. See e.g., Fine (1994, 1995), Lowe (1994) and Correia (ms).

some mental feature M, and this can be read as an ontological dependence claim; but it doesn't do anything to show that it is always inappropriate to frame Thing-realism in terms of modal dependence. However, Jenkins' view is that *all* realist/irrealist claims can and should be understood as involving essential dependence, which requires that the notion of essence be extended in such a way that it makes sense to talk of the «essence» of something's being the case (i.e., of a state-of-affairs' obtaining) and to say that something's being the case essentially depends on something else's being the case. The worry is not that there is no way in which this could be done, but that Jenkins' way of doing it results in a notion of essential dependence which is so strong as to make the corresponding realist claims almost trivially true.

It is certainly unclear what we should take to be the essence of *entities* of any kind. But it is even less clear (*pace* Jenkins 2005: 200) what the essence of something's being the case should be taken to be. For this is *not* to be viewed as the essence of any kind of *entity*: even if states-of-affairs are treated as entities, the question doesn't concern the essence of states-of-affairs, but the essence of the *obtaining* of a state-of-affairs⁸. Jenkins (2005: 200) introduces the generalized notion of essential dependence by saying⁹ that its being the case that P essentially depends on its being the case that Q (e.g., on our possessing mental feature M) when it is part of *what it is for P to be the case* that Q be the case (e.g., that we possess M). From this we learn that the essence of its being the case that P is *what it is for P to be the case*, or *what its being the case that P consists in*, and that it may have «parts». Given that she later adds (2005: 206-207) that what it is for P to be the case should be identified neither with *what makes it the case that P*, nor with *what it takes for P to be the case*, nor with *what it is in virtue of which it is the case that P*, this doesn't leave much to rely on in trying to understand what she means to be saying here.

Even so, it could at least be noticed that this way of generalizing the relation of essential dependence between entities to turn it into a relation between something's being the case and something else's being the case, departs somewhat from the original idea in the following respect. To say that it is part of the essence of some entity A, that necessarily A exists only if B has a certain feature, is *not* to say or imply that B is a part of A or that B's having a certain feature is part of what it is for A to exist (or to be what it is). It is just a way of saying that the essence of A is such that it requires that necessarily A exists only if B has a certain feature. Accordingly, one would have expected the claim that its being the case that P essentially depends on its being the case that Q to have been understood as meaning that it is part of what it is for P to be the case that necessarily it is the case that P only if it is the case that Q (i.e., as meaning that the «essence» of its being the case that P *requires* that necessarily it is the case that P only if it is the case that Q). But on Jenkins' explanation, for its being the case that P to depend essentially on its being the case that Q, its being the case that Q must be *part of what it is for P to be the case*, and hence, of its «essence».

Hence, we are confronted with two different ways of understanding the notion of essential dependence. On the one suggested by Fine's work, its being the case that P essentially depends on its being the case that Q iff the «essence» of its being the case that P requires that its being the case that P necessitates its being the case that Q. On the one proposed by

⁸. This suggests that, if the discussion was to be put in the semantic mode, the question would have to be seen as a question concerning the essence of truth.

⁹. Jenkins' own gloss is in terms of «independence», but nothing hangs on this.

Jenkins, its being the case that P essentially depends on its being the case that Q iff what it is to be the case that Q is *part* of what it is to be the case that P¹⁰.

On either interpretation, the realist who claims that its being the case that P doesn't *essentially* depend on our possessing mental feature M has the obvious advantage that he/she may eventually grant that its being the case that P may yet *modally* depend on our possessing M. But how reasonable is it to *require* that the notion of essential dependence involved in such a claim be construed in the way suggested by Jenkins?

Suppose the irrealist wants to advance the claim that it is «part» of what it is for P to be the case that we possess mental feature M. This is tantamount to the claim that its being the case that P is at least in part a *mental* fact (considering our possessing M as a mental fact), or that it reduces, in part, to a mental fact. Transposed in the semantic mode, this suggests that the judgement that P somehow *makes reference* to us and to feature M, i.e., that it is, *in part*, about our possessing M. But there seems to be no compelling reason to think that the irrealist, as such, must always be committed to such a strong claim (i.e., to think that it should be impossible to be irrealist about something completely *non-mental* in nature).

Here it may be useful to call attention to the fact that in most cases, the mental feature relevant to the realism issue will be a *capacity* to have some *intentional attitude* towards something's being the case. For example, it might be claimed that its being the case that P depends on our possessing the capacity to know that P, or on our possessing the capacity to judge (think, conceive) that P (where «our possessing this capacity» is short for «someone's possessing this capacity»). Reading this as involving Jenkins' notion of essential dependence yields the claim that its being the case that P consists, in part, in our possessing the capacity to know or judge that P. But it is hard to see how this could ever be the case. How could *its being the case that P* consist, even in part, in its being the case that we have the capacity to know or judge that *it is the case that P*? That would seem to require that the thought that P is in part a thought *about* our capacity to know/judge that P¹¹. But it is hard to see how anything can be such that being known or thought about, or being such that we can know or think about it, is literally *part* of what *it* is.

Suppose that *part* of what it is for it to be the case that P is that it be the case that we have the capacity to know/judge that P. It is at least arguable that this capacity involves a capacity to have a thought *about* its being the case that P. This is not (yet), however, to say that its being the case that P must in turn be part of what it is to have the capacity to know/judge that P (if only because one can have this capacity without ever exercising it), which suggests that our having the capacity to know/judge that P *could* be part of what it is for it to be the case that P, without being identical with it. Suppose that they are not identical. What, it can know be asked, is the other part of its being the case that P? And why should this other part be

¹⁰. It has been suggested (by Mark McCullagh) that these two «readings» may not be substantially different. If this is the case, then the objections I am going to raise against framing the realist/irrealist debates in terms of Jenkins'essential dependence would carry over to framing them in terms of Fine's essential dependence.

¹¹. As we have all learned at mother's knee, from the fact that its being the case that P consists (in part) in its being the case that Q, it doesn't follow that one cannot judge that P without (in part) judging that Q. But it does follow that any judgement that P is a judgement (in part) *about* Q. To take one objector's example, if Lois Lane thinks that Superman can fly, and its being the case that Superman can fly consists in its being the case that Clark Kent can fly, then her thought that Superman can fly *is* a thought *about* Clark Kent's ability to fly (without being a thought *that* Clark Kent can fly).

«mental» in nature? I don't see how such questions could be answered. Now suppose they are identical. Then, its being the case that P *is* our having the capacity to have a thought about it. But *that* doesn't specify any state-of-affairs (is there any way to specify what «it» refers to here?).

But whether they are identical or not, any thought *about* its being the case that P will at least in part be a thought about our capacity to know/judge that P, i.e., a thought about our capacity to have a thought about whatever this very thought is about. Perhaps there are such thoughts, and perhaps we can even have some of them. But it does seem unfair to saddle the irrealist with the view that whenever we think about a mind-dependent state of affairs we are in part thinking *about* our capacity to think about whatever we are thinking about, even if some irrealists may indeed have embraced some such views.

Moreover, it is not even clear that such «properties» as that of being thought *about* are «real» properties, or if they are, that they are «mental» properties. Does Vienna count as having a *mental* property, just in virtue of the fact that I am thinking about it? It would seem that the one who does the thinking is the one who has the mental property.

In any case, it seems doubtful that one who holds, e.g., that its being the case that $2 + 2 = 4$ depends on our having the capacity to prove it, or that its being the case that John believes that P depends on our having the capacity to judge that John believes that P, has ever meant to be claiming that part of *what it is* for $2 + 2$ to equal 4 is for us to be able to prove that $2 + 2 = 4$, or that part of *what it is* for John to believe that P is for us to be able to think that John believes that P (at least in any sense implying that the thoughts that $2 + 2 = 4$ or that John believes that P are thoughts *about* our capacity to know or judge *them*). It has been objected¹² that on some views about mathematics, for it to be the case that $2 + 2 = 4$ *is* for us to be able to prove it. This surely is a common way of expressing certain constructivist conceptions of mathematics. But as far as I can see, the views which tend to be expressed in this way are not really of the kind I am here discussing. The claim that its being the case that $2 + 2 = 4$ *is* our being able to prove it is not to be understood literally, as saying that its being the case that $2 + 2 = 4$ consists in our having the capacity to prove *that* $2 + 2 = 4$, but as a lousy way of saying that it consists in our having the capacity to prove the *formula* « $2 + 2 = 4$ » (or something like that).

At one point, Jenkins (2005: 204) mentions Berkeley's view that «truths about the physical world *consist in* truths about ideas in the mind of God» in support for the claim that irrealists generally consider themselves as being committed to essential dependence. I agree that the view just alluded to, according to which physical facts actually are identical to (some) mental facts, does involve essential dependence¹³. But on the face of it, it is just not the kind of position which would currently most often be associated with irrealism. For one thing, the irrealist as such is not necessarily committed to any form of reductionism (there must be room for non-reductive forms of irrealism); for another, the relevant mental facts will nowadays

¹². By Gregory Lavers.

¹³. Jenkins (2005: 204) also mentions the kantian view that «phenomenal reality *is* (in part) a product of our sensibility» in support for his contention. But here I shall deny that the view in question is best understood in terms of essential dependence; there is no clear sense in which «our sensibility» (or for that matter, «our mind») has to be seen as a part of its products. In my opinion, this kind of view is best construed as involving some sort of «constitutive» dependence (see below).

(most often) be taken to be facts *about us*. But more importantly, most standard formulations of realism/irrealism involve dependence with respect to *intentional* attitudes, and it strikes me as just implausible to suggest that part of *what it is* for something to be the case (for a judgement to be true, or a state of affairs to obtain) is for *it* to be (or possibly be) the content of any such attitude. Hence, it could not be objected that I am saddling Jenkins with an implausibly strong construal of the notion of essential dependence. The trouble doesn't have anything to do with the strongness of the relation in itself. It is reasonably clear, at an intuitive level, what is meant by saying that part of what it is for something to be the case is for something else to be the case, and I don't want to suggest that such claims are so strong as to never be true. The trouble comes when it is suggested that this kind of relation might hold between something and (a capacity for) an intentional attitude that one may have about or towards it.

None of this, however, is to deny that in many contexts, the irrealist may and will have in mind a relation stronger than modal dependence. One candidate is the notion of essential dependence, *as understood in the way suggested by Fine*, another is what might be called a relation of «constitutive» dependence. To take an analogy, there seems to be a sense in which, e.g., its being the case that Sherlock Holmes smokes the pipe strongly depends on someone's having said so. But no one would want to suggest, I take it, that (even part of) what it is for it to be the case that S.H. smokes the pipe (for S.H. to smoke the pipe) *is* for it to be the case that someone said so. In such a case, it is tempting to say that its being the case that S.H. smokes the pipe is somehow *constituted* by the fact that someone said that he smoked the pipe (but note that saying *that S.H. smokes the pipe* doesn't involve saying *that one is saying so*). The trouble with such cases, of course, is that they deal with the realm of fiction, where statements are often claimed not to be «strictly» true, while we want to consider cases where it could «strictly» be the case that P, even though its being the case that P «constitutively» depends on our capacity to know/judge that P¹⁴. They may nonetheless give some idea of what is the intuition behind the notion of constitutive dependence.

A question might be raised, as to whether constitutive dependence really must be understood as being strictly stronger than modal dependence. The same question is raised by Jenkins (2005: 203), with respect to essential dependence¹⁵. According to her, there is no question that essential dependence entails modal dependence¹⁶, when *modal* claims are understood in the usual sense. But she contends that they are sometimes understood in such a way that modal independence fails to entail essential independence. She illustrates this by means of the following example. An irrealist about mathematics might be happy to admit that « $2 + 2 = 4$ » is true even in worlds where nobody exists, while maintaining both (what Jenkins (2005: 203) takes as an *essential* dependence claim, namely) that «what it is for $2 + 2 = 4$ to be the case is for us to construct mathematical structures of a certain kind» *and* that it is

¹⁴. Unfortunately, I have no space here to discuss the question whether those who think that there is some metaphysically neutral notion of «minimal truth» would or should count true fictional statements as true *simpliciter*, in this minimal sense. The obvious analogy, here, is with some versions of constructivism in mathematics, where the question may arise, as to whether true arithmetical statements should be seen as on a par with true fictional statements.

¹⁵. In this and the next paragraph, I'm indebted to Mark McCullagh.

¹⁶. This holds on either of the two ways of construing this notion that I have mentioned.

necessary that $2 + 2 = 4$. The funny thing about this example, is that it doesn't involve anything like a non-standard or unusual understanding of *modality*: both the statement that necessarily, $2 + 2 = 4$, and the statement that necessarily, $2 + 2 = 4$ only if we have constructed a certain structure could be understood in the usual way, as meaning, respectively, that $2 + 2 = 4$ in all possible worlds (including those where we don't exist), and that we have constructed this structure in all worlds where $2 + 2 = 4$. It is just that since the first statement is admitted as true, the second has to be rejected as false. But if there has been no shift in the notion of modal truth, then there must have been a shift in the notion of essential dependence itself.

What the example actually relies on, then, must be a new reading of the notion of essential dependence itself, one which departs from Jenkins' own «official» interpretation and conflates it with a form of what I call «constitutive dependence». On the face of it, it will easily be granted that one who claims that what it is for $2 + 2 = 4$ to be the case is for us to construct mathematical structures of a certain kind is not likely to be implying that the fact that we have constructed such and such structures is *part of* the fact that $2 + 2 = 4$ (or to put it otherwise, it is unlikely that a constructivist about mathematics will express his/her view by using these words, except perhaps as a matter of stylistic convenience). For it is clear (since Jenkins' essential dependence *does* entail *standard* modal dependence) that if he/she were, then he/she would be committed to the corresponding modal dependence claim.

This is not to deny that there may be some non-standard reading of the claim that necessarily, $2 + 2 = 4$ only if we have constructed a certain structure (i.e., a non-standard interpretation of modal dependence), on which it will follow from (or at least be compatible with) the claim that its being the case that $2 + 2 = 4$ «constitutively» depends on our having constructed the relevant structure, and/or *fail* to follow from the corresponding essential dependence claim. Thus, I'm willing to acknowledge that there may be some construal of the modalities on which the resulting (non-standard) modal dependence claims will not entail Jenkins' essential dependence claims. But this is irrelevant to the point at issue, for it does nothing to show that the irrealist who understands modal claims in such a way is nonetheless concerned with Jenkins's essential dependence (quite the contrary).

If this is right, then what Jenkins' example shows is that *constitutive* dependence can be construed in such a way that it doesn't entail modal dependence. I see no reason, however, to conclude that it must be so construed, or more importantly, to conclude that no specific constitutive dependence claim can entail the corresponding modal dependence claim. Either way, I think it is now clearer why to insist that the realism/irrealism debates must be put in terms of *Jenkins'* notion of essential dependence threatens to trivialize any realist position. Admittedly, insisting that they be put in terms of *Fine's* notion of essential dependence (or for that matter, in terms of constitutive dependence) would not have this consequence, but I fail to see any point in doing so.¹⁷

References

- Collins, John (2004) «Preemptive Prevention», Collins, Hall and Paul eds (2004) 107-117
Collins, John, Ned Hall and L. A. Paul eds (2004) *Causation and Counterfactuals*, Cambridge

¹⁷. This paper has been read at the annual meeting of the Canadian Philosophical Association (York, may 2006). Special thanks to Mark McCullagh for his thoughtful comments.

(Mass.), MIT Press

Correia, Fabrice (unpublished ms) «Existential Dependence»

Fine, Kit (1994) «Essence and Modality», Tomberlin *ed.* (1994) 1-16

Fine, Kit (1995) «Ontological Dependence», *Proc. of the Aristot. Soc.* 95, 269-290

Jenkins, C. S. (2005) «Realism and Independence», *American Phil. Quart.* 42, 199-209

Lewis, David (1973) «Causation», Lewis (1986) 159-172

Lewis, David (1986) *Philosophical Papers 2*, Oxford, Oxford U. Press

Lowe, E. J. (1994) «Ontological Dependency», *Philosophical Papers* 23, 31-48

Sanford, David (1989) *If P then Q*, London, Routledge

Tomberlin, James E. *ed.* (1994) *Philosophical Perspectives 8: Logic and Language*, Atascadero, Ridgeview.

Daniel Laurier
Département de philosophie
Université de Montréal
C.P. 6128, succ. Centre-Ville
Montréal, Qc H3C 3J7
Canada
daniel.laurier@umontreal.ca

THE LOGIC OF 'IF'

OR HOW TO PHILOSOPHICALLY ELIMINATE CONDITIONAL RELATIONS

by Rani Lill Anjum

Introduction

In this paper I present some of Robert N. McLaughlin's critique of a truth functional approach to conditionals as it appears in his book *On the Logic of Ordinary Conditionals*. Based on his criticism I argue that the basic principles of logic together amount to epistemological and metaphysical implications that can only be accepted from a logical atomist perspective. Attempts to account for conditional relations within this philosophical framework will necessarily fail. I thus argue that it is not truth functionality as such that is the problem, but the philosophical foundation of modern logic.

1. Six principles of logic

In classical logic before Frege, there were three basic principles of logic that Copi calls «The three laws of thought»:¹

1. **The principle of identity:** If a statement is true, then it is true.
2. **The principle of contradiction:** No statement can be both true and false.
3. **The principle of excluded middle:** A statement is either true or false.

With the introduction of Fregean logic, three new principles were added. This made modern formal logic purely extensional:

4. **The principle of composition:** The meaning of a complex expression is determined by its structure and the meanings of its constituents.
5. **The principle of truth functionality:** The truth-value of a complex expression is determined by its structure and the truth of its constituents.
6. **The principle of form:** The logic of a complex expression is determined by its (syntactic) form, not its (semantic) content.

¹. Copi (1968), p. 244.

2. The material conditional

Although Fregean logic has some obvious advantages over Aristotelian logic, being formally more advanced and flexible, some problems follow from the introduction of the three extensional principles. For instance, where one within classical logic can distinguish between the two statements ‘If Socrates is a man, then he is mortal’ and ‘If Socrates is mortal, then he is a man’ by judging the first as true and the second as false, Fregean propositional logic will make both these statements come out as true, since Socrates was both a man and mortal.

This is commonly thought to be an unfortunate consequence of applying the principle of truth functionality to conditionals. Interpreted truth functionally, the only combination that makes a conditional ‘if p then q’ false is when the antecedent ‘p’ is true and the consequent ‘q’ is false at the same time, something that renders the conditional true for all other combinations:

- i) A conditional is true *whenever its antecedent is false*, which means that for instance the inference ‘not-p, therefore, (if p then q) and (if p then not-q)’ is valid.
- ii) A conditional is true *whenever its consequent is true*, which means that for instance ‘q, therefore, (if p then q) and (if not-p then q)’ is a valid inference.
- iii) A conditional is true *whenever both its antecedent and consequent are true*, so that ‘p and q, therefore, (if p then q) and (if q then p)’ is a valid inference.
- iv) A conditional is true *whenever both its antecedent and consequent are false*, which means that the inference ‘not-p and not-q, therefore, (if p then q) and (if q then p)’ is valid.

McLaughlin argues that truth functionality represents a dead-end for understanding conditionals. According to him, a truth functional logic cannot be a satisfactory model of a language that includes conditionals:

The truth functional calculus of propositions is one of the foundation stones in modern logic. It is so eminently reasonable, and its principles perform so important a function in all branches of standard logic, that doing logic without it is unthinkable. And yet, it is a fact that the calculus is only satisfactory as a logical model of a language in which conditionals do not exist. Even such humble assertions as ‘If Smith works hard, then he will get a raise’ fall outside the range of statements that can serve as interpretations of the forms of truth functional logic. This would not be an embarrassment if conditionals were marginal characters on the linguistic stage. But our conviction that this is not so, that they are important actors with major roles to play, is evidenced by the many attempts that have been made to find models that properly express their logic.²

3. Three alternatives to the material conditional

Now many logicians have tried, as McLaughlin notes, to search for alternative and more satisfactory models for conditionals in order to avoid the problems faced by the truth functional account of conditionals.

One tradition, represented by Adams and Edgington,³ argues that the abovementioned problems with a truth functional interpretation of conditionals indicate that natural language conditionals do not have truth conditions at all, but rather probability conditions. Adams’ thesis is that the conditional probability a subject assigns to ‘q’ given ‘p’ is defined as equal

². McLaughlin (1990), p. 1.

³. Adams (1975), Edgington (1991) and (1995).

to the probability she assigns to 'p & q' divided on the probability she assigns to 'p', thereby preventing the consequent from being less probable than the antecedent.

Another tradition, influenced by Adams and Grice, and represented by Jackson,⁴ claims that a conditional has truth conditions according to the material conditional, but assertion conditions according to whether the conditional can be asserted or believed. One then distinguishes between some truth functional content of a conditional on the one hand and its communicated pragmatic content, which is therefore claimed not to be relevant for the truth of the conditional, on the other. In defining assertability as the conditional being robust with respect to the truth of its antecedent, Jackson uses Adams' thesis for conditional probability of conditionals. This means that the assertability of a conditional is still regarded as calculable from the antecedent and the consequent.

A third tradition that is widely accepted is the possible-worlds theory of conditionals, represented by Lewis⁵ and Stalnaker⁶. They give an account of conditionals according to which a conditional is true (in the actual world) if in the closest possible world where 'p' is true, 'q' is true as well. In this approach, one avoids that a conditional is true whenever its antecedent is false. However, in every possible world, the material conditional is claimed to hold as a model for conditionals, so that in a possible or an actual world, all true statements entail each other:

According to Stalnaker if 'p' is true in our world, then our world is the closest accessible world containing p and, because the conditional is proved by its confirmation, the conditional relation holds between p and all its world's happenings. If this stone is on the ground and the sun shines, then we are bound to accept both 'If the sun shines, then the stone is on the ground' and 'If the stone is on the ground, then the sun shines'.⁷

4. Calculability and reduction

In all the approaches sketched above, the conditional is treated as a function of the value of its antecedent and consequent. Hence one generally expects to be able to *calculate* the truth, probability or assertability of 'If p then q' from the truth, probability or assertability of 'p' and 'q' as such. McLaughlin argues against such a reductionist approach to conditionals:

Consider, for instance, what is usually taken as the strongest evidence for the truth of 'If p then q', the joint fulfillment of 'p' and 'q'. This by no means furnishes proof of 'If p then q'... The truth of 'If p then q' cannot be extracted, as a matter of logical necessity, from the truths of 'p' and 'q'... Even some who reject the truth functional model of the singular conditional propose systems in which the truth of the conditional can be equated with the joint occurrence, in certain possible worlds, of the events mentioned in the 'if' clause and that mentioned in the 'then' clause.⁸

McLaughlin here points to the fact that a conditional 'If p then q' is in general defined to follow from 'p and q'. However, even though it is true that grass is green and snow is white, this is not sufficient for me to infer that 'If grass is green, then snow is white', nor that 'If

⁴. Jackson (1987) and (1991).

⁵. Lewis (1973).

⁶. Stalnaker (1968).

⁷. McLaughlin (1990), p. 31.

⁸. McLaughlin (1990), pp. 8-9.

snow is not white, then grass is not green'. Moreover, even if I start out with a conditional, for instance 'If Norway joins EU, then the interest rate will fall', I cannot from the fulfilment of both the antecedent and the consequent infer the truth of the conditional. It might for instance be the case that the interest rate falls for other reasons than Norway joining EU.

So even though many logicians acknowledge the problems with the material conditional and try to solve them by finding alternative ways of representing conditionals in a logical system, all the alternatives seem to have the character of being pure *ad hoc* solutions to problems related to the material conditional. Also, all the approaches are attempts to include conditionals in extensional logical systems. Adams, Edgington, Grice, Jackson, Stalnaker and Lewis offer alternative solutions that are more intricate than the material conditional approach; however, all the different approaches basically *depend on* the material conditional.

There seems to be a common craving for a pure formalism including calculability and extensionality. This is what Sören Stenlund refers to as a «calculus conception of language». Characteristic for this view is that one considers language to be in principle a calculus, or a formal system that can be isolated from the situations in which it is used.⁹ As long as one is not willing to give up on this conception of language, is it then at all possible to find a satisfactory approach to natural language? I think not.

5. A (truth) functional world

Let us now assume that conditionals can in fact be adequately accounted for within an extensional, truth functional, syntactic, compositional approach. What would be the consequences? According to McLaughlin, a truth functional logic can only be a model of a language addressed to a certain kind of world. So given that the truth functional logic *is* a satisfactory model of natural language, then natural language must have properties according to the six basic principles of modern propositional logic, as referred above. Moreover, given that natural language *is* truth functional, so that a truth functional logic is in fact an adequate tool for deciding truth and validity of our statements and arguments; then natural language must be a language of a world that has its *structure* according to these principles.

The propositional calculus is well designed to serve as the logic of a language addressed to a world of which all one had to say was that certain events have or have not occurred or will or will not occur. And this arrangement would be acceptable if we really thought it the case that the world is a collection of discrete or unrelated events, the occurrence or nonoccurrence of which can be expressed by simple assertions and their denials. Because if reality consists wholly of elementary facts that can, in appropriate circumstances, be determined by observation to obtain or not, and if the sentence letters of the calculus are interpreted by statements or propositions that express these facts, then if the facts are known, all the truths about the world that are capable of rational representation can be embraced in a long conjunction of simple or atomic propositions and their negations of the form 'p.q.-r.-s, ...' each element of which states the presence or absence of a fact.¹⁰

This picture of the world addressed by a truth functional language seems to fit neatly into the philosophical picture of logical atomism. Within this tradition, the world is taken to consist of atomic facts that are described by simple statements. These atomic facts form part of a structure that corresponds to the truth functional structure of language. What McLaughlin points out is that the basic principles of logic, taken together, involve a reduction, but also

⁹. See Stenlund (1990) for a detailed account of the calculus conception of language.

¹⁰. McLaughlin (1990), p. 2.

dissolution, of conditional relations as such. To reduce the conditional form ‘If p then q’ to a negated truth functional conjunction ‘ $\neg(p \ \& \ \neg q)$ ’ or to a truth functional disjunction ‘ $\neg p \vee q$ ’ eliminates the very notion of a conditional relation.

But this is to trivialize an important region of speech. The relations between events are just as significant to us as the question of their occurrence. We want to know the circumstances under which things happen — and this is so whether the subject matter is particles of light, cells in the body, commercial responses to movements of prices on the open market, or the behavior or feelings of animals and humans. To ground a claim that ‘if p then q’ is true on the fact that ‘p’ is false or ‘q’ true misses the point of what the conditional is meant to say. A conditional, when applied to the experienced world, attempts to report a dependency between events, a dependency that obtains as a result of physical connection, social convention, or human purposes; this objective is not fulfilled by the fact alone that our observations show that the assertion has not been disconfirmed.¹¹

Most of our interactions with the world and each other depend on us handling conditionals in a totally different way from the observing of occurrence and non-occurrence of singular events. According to Grice, however, «no one would be interested in knowing that a particular relation (truth functional or otherwise) holds between two propositions without being interested in the truth-value of at least one of the propositions concerned, unless his interest were of an academic or theoretical kind».¹² But is this really so? If someone tells me that ‘If you eat that mushroom, you’ll die’, I am not first of all interested in the truth or probability of the antecedent or consequent. I am interested in the conditional relation.

6. Treating conditionals extensionally

Assuming that our interest in the world is primarily an interest in the occurrence or non-occurrence of single events, so that our interest in relations between events is reduced to a question about whether or not we observe ‘p’ and ‘not-q’, most of our conditionals would be useless and senseless:

We know that there are countless assertions that we regard as being true or false but that employ concepts that do not take observable instances. There are no observable properties through which such concepts as pressure, energy, force, gravity, and time are immediately instantiated.¹³

Take for instance the following statement: ‘If a body is not subject to any net external force, it will continue in a uniform movement or stay at rest.’ We know that no bodies can be unmoved by forces. According to our truth functional approach, any conditional that cannot be directly tested must be true, since it then cannot be proved false. This means that the following conditional will also come out as true: ‘If a body is not subject to any net external force, it will search for the nearest cat to kill it.’ As we saw above, the Stalnaker/Lewis tradition wants to reduce this problem by referring to the closest accessible possible worlds in which the antecedent ‘p’ is true. So if the consequent ‘q’ is true in all the possible worlds where ‘p’ is true, the conditional ‘if p then q’ is (necessarily) true. However, such a theory does not seem to get the job done for conditionals.

In helping us to impose order on the world, they (conditionals) help us to determine our own actions. Consequently, we are not concerned to know, when ‘p’ is false, whether ‘q’ is true in the closest (but by definition unactual) world in which ‘p’ is true. (How could one find out such a thing?) We want to know what

¹¹. McLaughlin (1990), p. 6.

¹². Grice (1989), p. 61.

¹³. McLaughlin (1990), pp. 6-7.

would happen, what conclusion one might draw, if the event expressed by 'p' is actualized by its occurrence.¹⁴

Extensional logic does not only represent a reduction and downgrading of conditionals. It seems to dissolve the whole idea of conditional relations. If anything, we should have learned from Hume that we cannot reduce a conditional relation to the direct observation of joint occurrence or constant conjunction of two distinct events. But this seems to be an unavoidable consequence of treating conditionals extensionally.

7. Concluding remarks

We have seen how difficult it is to cut loose from the basic principles of logic, including functionality and calculability. McLaughlin makes clear that our models of language must be consistent with our models of the world and the relation between language and the world. Put differently, he makes clear that it would be inconsistent to maintain a set of logical principles and at the same time deny their epistemological and metaphysical implications. What we need, then, is a metaphysics that allows conditional relations as basic and primitive. Only then will we have a basis for developing an adequate logic, modelling a language, addressing a world that we can recognise as ours. This is not a world of unrelated events and accidental regularities, but a world of potentials, connections and dependencies.

Acknowledgments

Thanks to Johan Arnt Myrstad for constructive comments and countless discussions about conditionals.

References

- Adams, Ernest W.: *The Logic of Conditionals*, Dordrecht, Reidel 1975.
- Copi, Irving: *Introduction to Logic*, New York, Macmillan 1968.
- Edgington, Dorothy: 'Do Conditionals Have Truth-Conditions?', in Jackson (ed.) 1991, pp. 176-201.
- Edgington, Dorothy: 'On Conditionals', *Mind* 104.414., April 1995, pp. 235-329.
- Grice, Paul: *Studies in the Way of Words*, Cambridge, Harvard University Press 1989.
- Jackson, Frank: *Conditionals*, Oxford, Blackwell 1987.
- Jackson, Frank (ed.): *Conditionals*, Oxford, Oxford University Press 1991.
- Jackson, Frank: 'On Assertion and Indicative Conditionals' in Jackson (ed.) 1991.
- Lewis, David: *Counterfactuals*, Oxford, Blackwell 1973.
- McLaughlin, Robert N.: *On the Logic of Ordinary Conditionals*, Albany N.Y, State University of New York Press 1990.
- Stalnaker, Robert: 'A Theory of Conditionals', *Studies in Logical Theory: American philosophical quarterly monograph*, Oxford, Blackwell 1968, pp. 98-112.

¹⁴. McLaughlin (1990), p. 35.

Stenlund, Sören: *Language and Philosophical Problems*, London, Routledge 1990.

Rani Lill Anjum
Department of Philosophy
University of Tromsø
Norway
rani.anjum@sv.uit.no

WITTGENSTEIN AND THE SORITES PARADOX

by David Michael Wolach

Two considerations might let allow us to understand Wittgenstein's analysis of the Sorites Paradox. The first has to do with Wittgenstein's notes on vagueness in natural languages. The second arises from his remarks on purity.

Contra Frege, Wittgenstein considers vagueness to be pervasive in natural language, unproblematic (at points, useful), and unshakable. Its pervasiveness is handled directly in §§104-108 of the *Investigations*. Its usefulness, and that vague sentences do not lack sense, is defended *par exemple* in §75 and §76. And Wittgenstein speaks about the unavoidability of vagueness most centrally at §§72-73, §§77-81, §§104-108.

Wittgenstein asks us to consider (§72) the narrow definition of an item by ostension (I point to a color patch and say «this color patch is called 'yellow ochre'.») This is akin to giving someone a color patch with a word written under it. As of yet the person has no fixed rule at his disposal which tells him whether or not some new, similar color, may be called 'yellow ochre' — he has only the rule: *this color is called 'yellow ochre'*. The respondent complains: «here is this new color, and I have no way of saying whether *it* is called 'yellow ochre'.» And so I give him a range of color patches lined up on a chart, each slightly different from the one next to it, and say: «only color patches falling within *this* range may be called 'yellow ochre'.» Now I have given a precise definition of the term, by giving examples, by marking off the territory in which the term is applicable (the instances under which the predication is appropriate).

Here we may be tempted to say that we have 'discovered yellow ochre', Wittgenstein tells us, or that we have resolved the fuzziness of the term's application *tout court*. The first temptation arises when we assume that we have seen our language's horizon, and therefore caught a glimpse of what lies beyond it — that we have glimpsed some fact that is pure and exhaustive. The second, connected, temptation stems from our mistaken assumption that a localized use of a term (hence, the local language game itself), exists in isolation from all others and is thereby not threatened by the possibility they will infect the game under consideration.

In the first case we have not made a discovery but a *decision* — «To repeat, we can draw a boundary» (§69). We have *decided* that the boundary goes like this, or the constraints relative to a «special purpose» (painting the house, say) have made the decision ahead of time. In seeking precision, or in attaining it in the case of a narrow domain such as logic, we want to say that we have seen past the bounds of our language game and stared in the face of the *ideal* or *true* predication. But when we sense that we have made a *choice* to bound the extension of the term in this way rather than that, we become frustrated — «This means: it has

impurities, and what I am interested in at present is the pure article» (§100).

In the second case, we are all the more apt, especially in making a choice as to the use of a word or sentence, to think that this choice has resolved all (or orthogonal) ambiguities, that this resolve gets us closer to the meaning of the term. We may say of «Moses» (§79) that he was so and so, and decide that such a set of criteria for the ascription of «Moses» are necessary and sufficient. But this is no stopgap to the possibility of appropriating the term for different uses, to fill the spaces for different language games. Hence, the business of identifying necessary and sufficient conditions for a term's application is as slippery as tracking its use, its various functions. We should not mistake the resolution of an ambiguity within a local neighborhood of our language as a definitive victory over vagueness, so much as a relocation of it. But again, that we cannot isolate necessary and sufficient conditions for the application of a term or concept, save by making a choice (or by having a choice foisted upon us by customs and practices surrounding our particular endeavor), does not render the term or concept useless, lacking in sense.

If we approach the Sorites Paradox with these considerations at hand, we come to see that the «paradox» is rather a «grammatical joke» (cf. *Culture and Value*). We start by using our language in the ordinary way, saying: «well, I know that by 'child' I am talking about a human that is closer to birth than to death, given the average life span these days.» And then I say: «surely a one day-old human is a child, by my lights.» And then I ask: «If a human were a child and n days-old, then surely he would be a child at $n+1$ days-old?» But by inductive inference I have apparently committed myself to saying that a 100,001-day-old human being is a child, which I clearly would not like to do. My first reaction might be to reject the second premise. But by doing this I apparently commit myself to an *incredibly* stringent definition of «is a child,» which is again an unhappy consequence.

Wittgenstein's response, however indirectly, is to suggest that the problem is fiat. «Instead of philosophizing,» we might say, «look and see how the language functions.» Let us trace the use of the terms being employed, trace it back to its familiar application. We might find that the predicate «is a child» is usefully vague. It might turn out that «child» is sometimes used in this context, and now that, and all the while there is no necessary and sufficient set of criteria for application. It might turn out that all we have, as it were, are instances of use and accepted application. The problem looks like a philosophical one because we mistake the broader problem of predication as one relating propositions to states of affairs, wherein we actually are in a continual process of relating propositions to propositions. It might turn out, that is, that «is a child» simply does not admit of the kind of highest generality and greatest precision that we want. And so our only recourse is to make the term do what we want, give it a strict definition (or such a general one as to come out empty)—but this is to simply change the rules arbitrarily, and of course we can do that. The use of the predicate may have definite applications on either end of a spectrum, but in the middle it is unclear: «is *this* a child? I don't know, the question hasn't come up — I've never *used* the term in this way.»

Bringing ordinary language into philosophical waters in this regard is like training a microscope on the boundary between black and white squares of a chess board—we are now lost because the terrain is somehow extraordinary. It is precisely the open texture of certain predicates, and the demand that they function with sharper boundaries, that allows the sorites paradox to get off the ground. Hence, Wittgenstein suggests of the Sorites problem that «there is no problem,» and would say of vagueness generally that «the problem, insofar as there is one, should only be of practical interest—otherwise we are all on holiday.» Insofar as this is

a «solution,» it is the kind that only claims to trace the steps of how we arrived at the supposed problem in the first place—to show the fly the way out of the bottle.

David Michael Wolach
The Evergreen State College
<dwuaw@yahoo.com>

SAYING THE UNSAYABLE: WITTGENSTEIN'S EARLY ETHICAL THOUGHT

by Paul Formosa

1. Introduction

In a letter to Ludwig von Ficker, Wittgenstein explains that the key to understanding the *Tractatus* is to grasp that the overall «sense of the book is an *ethical one*».¹ Despite this, the *Tractatus* is often read as a book that seeks to define the limits of thought through a logical analysis of language. Indeed, such an approach is lent much credence from the simple fact that the vast majority of what is *in* the book deals with precisely this. Further, Wittgenstein also states in his preface that this is indeed the aim of the book.² However, on reading the final few pages of the *Tractatus*, one can be forgiven for wondering if things are really quite so clear. Bertrand Russell, in his introduction to the *Tractatus*, recognises, somewhat reluctantly, that it is not the logical analysis of language, but rather the mystical (or ethical) finale, which Wittgenstein himself «would wish to lay most stress» on.³ In this paper I will attempt to flesh out what the ethical sense of the *Tractatus* might be.

In the recent literature on the *Tractatus* two broad interpretative camps have emerged.⁴ On

¹. Quoted from P Engelmann, *Letters from Ludwig Wittgenstein, with a Memoir* (Oxford: Basil Blackwell, 1967) 143-44.

². «Thus the *aim of the book* is to draw a limit to thought, or rather ... to the expression of thought» — Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, trans. D F Pears and B F McGuinness (London: Routledge & Kegan Paul, 1974) 3. Italics are mine.

³. Russell, on page xxiv of his introduction, puts forward the following objection: that while there are things that cannot be said within a language, there is some meta-language in which they can be said. This hypothesis, Russell claims, if it proved tenable, would «leave untouched a very large part of Mr Wittgenstein's theory, though possibly not the part upon which he himself would wish to lay most stress» (xxiv-xxv). Russell's hypothesis would indeed leave the 'picture theory' and so on, i.e. the vast majority of the work, untouched, but it would make the mystical, the unsayable that can only be shown, actually sayable in some meta-language.

⁴. «My propositions serve as elucidations in the following way: anyone who understands me eventually recognizes them as nonsensical, when he has used them as steps — to climb up beyond them. (He must, so to speak, throw away the ladder after he has climbed up it.)» (6.54). There has been much debate on exactly how to interpret this finale and what implications it has for interpreting the *Tractatus* as a whole. This debate focuses mainly on whether Wittgenstein considered that there were two types of nonsense: illuminating nonsense and mere gibberish nonsense. On one reading of the text, forwarded by Diamond and Conant, among others, all nonsense is gibberish nonsense; illuminating nonsense is only different in that it gives the false impression that there is an illuminating ineffable truth lurking behind it. Hacker, among others, rejects this reading. On this alternative reading, there are two types of nonsense in the *Tractatus*: illuminating nonsense, which tries to manifest some deep, ineffable truth,

the one hand, there are those who see the *Tractatus* as a book that tries to teach us that metaphysics is nonsense and thus that it is not a practice that we should engage in. On the other hand, there are the more traditional or ‘metaphysical’ readings that see Wittgenstein as attempting to communicate some sort of illuminating or ineffable ‘nonsense’, as opposed to mere gibberish or nonsense proper. My account of Wittgenstein’s ethics will only make sense within this latter ‘metaphysical’ reading of the *Tractatus*. It is my intention in this paper to show that this ‘metaphysical’ reading of Wittgenstein is a *plausible* way of reading what is said in the *Tractatus*, and to show what account of Wittgenstein’s ethics follows *if* one accepts this reading. This should at least provide a burden of proof challenge to those who read Wittgenstein differently.⁵

In order to make sense of Wittgenstein’s ethical account I argue that we need to divide it into two distinct parts. The first part I shall refer to as Wittgenstein’s ethical *negative thesis*. This negative thesis is roughly the position that the propositions of ethics are nonsense. The second part I shall refer to as Wittgenstein’s ethical *positive thesis*. The positive thesis, which tries to state positively what ethics is, relies heavily on Wittgenstein’s understanding of the ‘metaphysical subject’ and the claim that there is ‘illuminating nonsense’.

2. Sense and Nonsense

I shall first take a very cursory detour via Wittgenstein’s account of language in order to set the foundation for Wittgenstein’s two ethical theses. In order to do this I shall show why some sentences are nonsense and how we might find room for illuminating nonsense in Wittgenstein’s system.

Wittgenstein divides the sentences of ordinary or natural language⁶ into those that ‘make sense’ [*sinn*], which are called propositions, those that ‘lack sense’ [*sinnlos*], which are called pseudo-propositions and those that are ‘nonsense’ [*unsinn*]. Every ordinary language sentence that makes sense can necessarily be analysed into propositional form. This is because every natural language sentence that makes sense is composed of «either a truth-functional composite of other simpler sentences or an atomic sentence consisting of a concatenation of

and plain nonsense, or gibberish, which manifests nothing. For bibliographical details of this debate see P M S Hacker, «Was He Trying to Whistle It?,» in *The New Wittgenstein*, ed. Alice Crary and Rupert Read (London: Routledge, 2000), 356-57, 82. Also see Cora Diamond, «Ethics, Imagination and the Method of Wittgenstein’s *Tractatus*,» in *The New Wittgenstein*, ed. Alice Crary and Rupert Read (London: Routledge, 2000).

⁵. It is not my intention, however, to defend this view against its various critiques, but merely to show that it is a plausible and intuitively appealing way to read what Wittgenstein says. Nor is it my intention here to mount a serious case for a rejection of the alternative reading offered by Diamond and others. I think Wittgenstein’s work is rich and broad enough, not to mention sufficiently elusive, to support more than one understanding of his work. It may be objected that my reading assigns Wittgenstein a ‘metaphysical’ position of the kind that he struggled against throughout his career. This objection, however, begs the question. My reading of the *Tractatus* implies that there is a bigger and more fundamental break between the early and late Wittgenstein than is sometimes thought (e.g. by Diamond).

⁶. Ordinary languages, such as English and German, are composed of *signs*. A sign can be meaningless and is just a convention (i.e. it is a convention that ‘hello’ means hello). A *symbol* is the part of a proposition that characterises its sense (3.31). We *recognise* a symbol from its sign (3.326). Symbols are not conventions and they cannot be meaningless. A symbol is necessarily a symbol of *something*, or it wouldn’t be a symbol at all.

simple names».⁷ Thus, in order to analyse natural language sentences into propositional form we must reduce them to a truth-functional combination of simpler sentences, and so on, until the analysis is complete. The analysis of a sentence into a proposition is complete when there is nothing left but a concatenation of simples that pictorially portray the relationship between themselves.

A state of affairs is simply a combination of objects (2.01) composed of simpler objects and ultimately simples themselves. A state of affairs is analysable into a concatenation of simples that pictorially represent their relationship to one another. Hence, both a proposition and a state of affairs pictorially portray the relationships between simples, and it is for this reason that Wittgenstein holds that propositions picture states of affairs. A proposition represents a state of affairs linguistically via a shared picture of simples that it has in common with the state of affairs it depicts.⁸ What makes only propositions have sense is that it is only propositions that picture a *possible* (or actual) state of affairs and thus only propositions lie within logical space.⁹ A proposition makes a 'cut' in logical space by picturing a particular possible state of affairs.

The «logic of the world» (6.22) is shown by pseudo-propositions that lack sense (6.2). These include the tautological (and contradictory) pseudo-propositions of logic, as well as the equations of mathematics (6.22) and certain *a priori* laws of science, such as the law of conservation.¹⁰ What these pseudo-propositions have in common is that they all express the «possibility of logical form» (6.33) by defining the very forms in which propositions can be cast (6.34). These pseudo-propositions all lack sense because they do not refer to a single possible state of affairs, as a proposition does, but rather they show something about *all* possible states of affairs; they pervade *all* of logical space (5.61). They are part of the symbolism (4.4611) and represent the very scaffolding which defines the forms out of which propositions that make sense can be constructed. Pseudo-propositions thus *lack* sense, but they can nevertheless *show* their sense because they say what is manifested by *all* possible propositions.

Outside of logical space lie nonsense [*unsinn*] sentences. One way a sentence can be nonsense is if it contains meaningless [*bedeutungslos*] signs. If a sign is useless then it is meaningless (3.328). For example, sentences which contain gibberish, such as 'over there lies

⁷. Hans Sluga and David G Stern, *Cambridge Companion to Wittgenstein* (Cambridge: Cambridge University Press, 1996) 10.

⁸. A picture is a representation of something by virtue of the fact that it shares something in common with what it represents and this is what Wittgenstein called 'pictorial form'. Inclusive in this term is 'logical form', which is the minimum amount of commonality that must be present between a picture and a state of affairs for one to be a portrayal of the other. Pictorial form is the possible relationships between constituent elements and it is these relationships between simples that a picture shares with the state of affairs it depicts. Likewise, a proposition will have a logical form in common with a picture that represents a state of affairs and so it is via the medium of a picture that a proposition represents a state of affairs, in which it maintains the original's logical form. Thus an identical logical form is what ties a proposition to the state of affairs it pictures. See 2.1-2.202.

⁹. Logical space is the space of all *logically* possible combinations of simples. The sum total of all possible propositions represents exactly the domain of logical space. 'The world' is that part of logical space that is the case.

¹⁰. Wittgenstein includes here the «law of conservation» (6.33), «principle of sufficient reason, the laws of continuity in nature and of least effort in nature» (6.34).

glurg slurgs', are nonsense because they contain signs, namely 'glurg' and 'slurgs', that are (for argument's sake) useless.¹¹ A sentence can also be nonsense, not because it contains signs that are useless, but because it does not picture any possible state of affairs. For example, the sentence 'God is a transcendent being' does not contain any useless signs, but it is still nonsense as the sentence does not picture a possible state of affairs. Nonsense sentences need not be gibberish even if they do try and say what cannot possibly be the case, or what is the same, cannot be said. Hence it is not absolutely ruled out that at least some non-gibberish nonsense sentences might prove not to be completely useless, as I shall try to show below.

Wittgenstein writes: «there are, indeed, things which cannot be put into words» (6.522).

But as everything that is possible can in-principle be put into words, it follows that if there are 'things' which cannot be put into words then such 'things' cannot exist within logical space. This reinforces the claim that some nonsense might have an important use, namely, to try and *illuminate* those 'truths' which cannot be said. By attempting to say what cannot be said, by speaking nonsense (but not meaningless gibberish), one can try to make the things which cannot be put into words mystically «manifest» themselves through words (6.522). While we will inevitably pass over such things in silence, this does not necessarily imply that we *should* be silent as we can try to *manifest* them in order to help others «see the world aright» (6.54). Indeed, this is what I take Wittgenstein to be trying to do throughout the *Tractatus*. For Wittgenstein there are ethical 'truths' that cannot, strictly speaking, be said at all, but can only be 'manifested' through illuminating nonsense, or so I shall argue. This, I wish to say, is the ethical 'sense' of the *Tractatus*.

3. Why ethics is nonsense: The negative thesis

Wittgenstein's negative ethical thesis claims that no ethical proposition can have sense. Ethical 'propositions' are, for Wittgenstein, absolute judgements of value of the form 'A makes the judgement that p is good'.¹² However, this immediately raises the following difficulty: in the *Tractatus*, Wittgenstein seemingly claims that it is *impossible* for a judgement of the form, 'A makes the judgement p', to be nonsensical (5.5422). This seems to directly contradict the claim that ethical judgments, which for Wittgenstein are of this form, can be nonsensical. However, what Wittgenstein's own analysis reveals is that it is impossible for judgements to be nonsensical *only if* the result of the judgement (i.e. p) is itself not nonsensical. When dealing with a proposition of the form, 'A makes the judgement p', Wittgenstein points out that although it 'superficially' appears to be the case that the proposition p stood in some relation to an object A, this is in fact not so. This is because Wittgenstein claims that when such propositions are analysed, the object A drops out of the proposition altogether, as its correct logical form is: «p» says p'.¹³ This analysis reveals that we do not have a correlation of a fact (the judgement p) with an object (judger A) at all, but

¹¹. «If a sign is *useless* it is meaningless...(If everything behaves as if a sign had a meaning, then it does have a meaning.)» (3.328). Thus if no one behaves as if a sign (i.e. such as 'glurg' or 'slurgs') has a meaning, as is the case with genuine gibberish, then such signs are meaningless because they have no use.

¹². What an 'absolute' judgement of value is and how it differs from a 'relative' one shall be discussed shortly.

¹³. How Wittgenstein gets this analysis is not made explicit in the *Tractatus*. Very briefly, his argument would (probably) go something like this: For A to make the judgement p is to assert or say «p». So 'A makes the judgement' is therefore analysable to the utterance «p». Thus we are left with something like '«p» says p'.

rather we have «the correlation of facts by means of the correlation of their objects» (5.542). The facts that must correlate here, by means of their objects, are «p» and p; i.e. the utterance «p» and the fact, or state of affairs, p — such a statement is true just in case the fact «p» correlates with the fact p. This implies that in order for a judgement of the form «A makes the judgement p» to have sense, p must be a possible fact, for otherwise the correlation of facts cannot occur.

For example, the proposition ‘Socrates makes the judgement that the moon is smaller than the sun’, has sense and is true (if Socrates did indeed make this utterance), as the utterance ‘the moon is smaller than the sun’ correlates to the possible fact that the moon is smaller than the sun. Consider, now, a second example: the ethical ‘proposition’, ‘Socrates makes the judgement that pity is good’. This ethical judgement will likewise only have sense if ‘pity is good’ is a possible fact or state of affairs, otherwise it will necessarily be nonsensical. Therefore, Wittgenstein’s analysis of judgements reveals that in order for an ethical judgement (e.g. ‘Socrates makes the judgment pity is good’) to have sense, and thus be a proposition, the result of the judgement (i.e. ‘pity is good’) *must* refer to a possible state of affairs. That is, there needs to be possible ethical facts in order for there to be ethical propositions.

In his *Lecture on Ethics*, Wittgenstein points out that all ethical terms, such as ‘good’, ‘right’, and so on, when used in judgements of value, have two senses: a «trivial or relative sense...and the ethical or absolute sense». ¹⁴ For example, judgements of value, such as ‘this is a good chair’ and ‘this is the right road’, as generally used, are relative judgements of value. They are relative judgements because the ‘ethical’ term, ‘good’, ‘right’ and so on, is used in the sense of good or right *for* something, and is thus assessed relative to some criteria. Further, *all* relative judgements of value are actually statements of facts and can be analysed into such a form so as to make this explicit. Thus, for example, ‘This is the right road’, can be explicitly analysed into, ‘This is the right road if you want to get to the park in the shortest time, without stopping off for milk and by following only conventional paths and so on’. Such a proposition clearly asserts a possible fact. As all relative judgements of value can be likewise analysed, it follows that all relative judgements of value can be propositions.

Absolute judgements of value, unlike relative ones, use ethical terms, such as ‘good’ and ‘evil’, in an absolute sense. An absolute value is not a value *for* something, but is a value-in-itself. Hence, it is like saying ‘this is the absolutely right way’ without regard for goal, destination or criteria. Wittgenstein defines the ethical good as follows: «the *absolute good*, if it is a describable state of affairs, would be one which everybody, independent of his tastes and inclinations, would *necessarily* bring about or feel guilty for not bringing about». ¹⁵ Wittgenstein then makes the crucial claim that *all* ethical ‘propositions’ are *absolute* judgements of value, because all ethical assertions employ ethical terms only in an absolute (and never relative) sense. ¹⁶ For Wittgenstein we say that others ought to do such and such a thing, ethically speaking, *because* it is good in-itself (in an absolute sense), not because it satisfies some criteria. Therefore, if all ethical propositions are absolute judgements of value, then this implies that in order for an ‘ethical proposition’ to actually be a proposition (i.e. to

¹⁴. Ludwig Wittgenstein, «A Lecture on Ethics,» *The Philosophical Review* 74, no. 1 (1965): 5.

¹⁵. *Ibid.*: 7.

¹⁶. *Ibid.*: 5.

have sense), the result of the ethical judgement must refer to a possible state of affairs (i.e. an ethical fact). If such an ethical fact can possibly exist, then it would alone be sufficient to make the corresponding ethical judgement have sense. But, Wittgenstein argues, this is impossible, because «no state of affairs has, in itself...the coercive power of an absolute judge».¹⁷ That is, no possible state of affairs in logical space *can* justify the absolute claims of an ethical judgment, from which it follows that it is «impossible for there to be propositions of ethics» (6.42).

This claim can be strengthened by reiterating that for Wittgenstein all propositions simply state possible facts and as all facts are of equal value, it follows that «all propositions are of equal value» (6.4). But we have seen that ethical assertions try to point to absolute facts, which implies the need for there to be some facts which have a greater value than other facts. But such ethical facts cannot exist, as Wittgenstein explains:

A stone, the body of a beast, the body of a man, my body, all stand on the same level. That is why what happens, whether it comes from a stone or from my body is neither good nor bad.¹⁸

Hence, just as we do not say that some fact about a stone is, in-itself, absolutely good or bad, it equally follows that some other fact about a human being can likewise not imply an absolute ethical value. For example, imagine all the facts related to a murder, or some other horrendous crime. All of these facts, physical and psychological, will be just plain facts as equally valueless as all other facts. No fact about the murder, such as a person having the psychological intention to kill someone, *can* have the absolute value required to justify an ethical claim. That is, no fact, in-itself, can possibly be absolutely good or evil — thus no ethical facts and therefore no ethical propositions.

Despite his conclusion that there can be no ethical propositions, Wittgenstein still maintains the common «tendency» or «temptation» to assert them. But this tendency to assert ethical propositions is not only one that Wittgenstein recognises, but one that he «personally cannot help respecting deeply».¹⁹ Wittgenstein explains this temptation to use ethical assertions by claiming that certain personal experiences, such as wonder at the existence of the world,²⁰ *seem* to have an absolute or intrinsic value. Clearly, however, these personal experiences that *seem* to have an absolute character are still just facts; they occur at such and such a time and so on. Indeed, Wittgenstein goes so far as to call it a «paradox» that «an experience, a fact, should *seem* to have supernatural [absolute/intrinsic] value».²¹ Of course, it will be

¹⁷. Ibid.: 7.

¹⁸. Ludwig Wittgenstein, *Notebooks 1914-1916*, trans. G E M Anscombe (Oxford: Basil Blackwell, 1979) 85.

¹⁹. «Ethics ... is a document of a *tendency* in the human mind which I personally cannot help respecting deeply and I would not for my life ridicule it.» — Wittgenstein, «A Lecture on Ethics,»: 12. Wittgenstein later discussed this lecture with some members of the Vienna circle. He states, on the 17th December 1930, at Schlick's house: «At the end of the lecture on ethics I spoke in the first person. I think that is something very essential. Here there is nothing to be stated any more; all I can do is to step forth as an individual and speak in the first person...I do not scoff at this tendency in man; I hold it in reverence...[but] I am speaking *about myself*». - Quoted from Cyril Barrett, *Wittgenstein on Ethics and Religious Belief* (Oxford: Basil Blackwell, 1991) 48.

²⁰. Wittgenstein mentions this both in the *Lecture on Ethics* and also in the *Tractatus* — «It is not *how* things are in the world that is mystical, but *that* it exists» (6.44).

²¹. Wittgenstein, «A Lecture on Ethics,»: 10.

nonsensical to assert that an experience has a 'supernatural' value, so the paradox is only in our tendency, our temptation to assert such nonsense.

4. Trying to say the unsayable: The positive thesis

Wittgenstein's ethical *positive thesis* tries to say what ethics is, as opposed to the negative thesis, which only states what ethics is not. It might seem that Wittgenstein's rejection of the possibility of ethical propositions has left him with no room for forwarding a positive ethical thesis. This is only partly true — indeed, *in* the world, so to speak, there is no room for ethics. But for Wittgenstein, as is made clear by the mystical finale of the *Tractatus*, there is still a *beyond* the world which, although one cannot speak of it, can nevertheless *manifest* itself through the world. For example, statements such as «there are, indeed, things which cannot be put into words» (6.522), only make sense if there are indeed 'things' *beyond* the world, because everything *in* the world can necessarily be put into words. The reason why some nonsense is illuminating is because it tries to point to that which is beyond the world, but of which in the world we can say nothing. In contrast, nonsense proper tries to manifest nothing. The thing beyond the world is, for Wittgenstein, the 'metaphysical subject', or synonymously, the 'will'.

In the *Tractatus*, Wittgenstein makes the following, rather cryptic, remark:

For what the solipsist *means* is quite correct; only it cannot be *said*, but makes itself manifest. The world is *my* world: this is manifest in the fact that the limits of *language* (of that language which I alone understand) means the limits of *my* world (6.52).²²

Firstly, Wittgenstein seems to be implying that what the solipsist means but cannot say is that «the world is *my* world». This, it should be noted, is in fact *not* the solipsist position at all, at least as it is generally understood.²³ Secondly, the phrase in brackets seems to imply some conception of a private language, but this, however, would be a serious misinterpretation.²⁴ What Wittgenstein means is that while there are many different languages, such as English and German, technically different systems of *signs*, there is still fundamentally only one *symbolism*, one language; this is because language is «a mirror-image of the world» (6.13). This then implies that there is only *one* world, namely the world of which our language is a mirror-image, and thus there is only one language and so only one language that I *can* understand. Further, it is a fact that this language defines the limits of the world (5.6), because «the world is all that is the case» (1) and language says exactly all that can possibly be the case (4-4.002). This account then implies that there are two concepts that need investigation here: '*the* world', of which our language is a mirror-image, and '*my* world' (referred to in the

²². Wittgenstein here implies that a thing can be 'correct' even if it cannot be said, which again seems to imply some sort of notion of an 'ineffable truth' beyond the world.

²³. Solipsism is «a belief that not only is there no *evidence* that anything else besides oneself exists, but that *this is the case*» — Barrett, *Wittgenstein on Ethics and Religious Belief* 65. Wittgenstein, as will be shown soon, is a realist about the existence of the world. He also thinks that there are other metaphysical subjects. So Wittgenstein is in effect putting forward a very different position to the traditional solipsist; he considers it to be solipsism that is 'properly thought out' (5.64).

²⁴. Mounce claims that this translation misses the sense of the original, and should be translated as «the only language that I understand». By this, Mounce claims, Wittgenstein is referring to what is common to all human language, its logical structure. See H O Mounce, *Wittgenstein's Tractatus: An Introduction* (Oxford: Basil Blackwell, 1981) 91.

above quote), which shall be shown to belong to *a* metaphysical subject.

Wittgenstein goes to some pains to distinguish between the ‘psychological subject’ and the ‘metaphysical subject’.²⁵ On the metaphysical subject Wittgenstein writes: «The [metaphysical] subject does not belong to the world; rather, it is *a* limit of the world» (5.632). Wittgenstein attempts to explain this obscure claim through an analogy to the eye in the visual field.²⁶ Just as the eye looks upon but is not *in* the visual field, similarly the metaphysical subject looks upon but is not *in* the world. Further, just as the eye is the outer limit just beyond the visual field, similarly the metaphysical subject is the outer limits of its world (i.e. ‘*my* world’) and thus beyond, or not in, ‘*the* world’. Thus, as the metaphysical subject, or synonymously the will, is not in ‘*the* world’, it follows that we can say nothing about it (5.631).

Further, just as each eye *has* a unique visual world that belongs only to that particular eye (i.e. what that eye alone sees), so too each metaphysical subject *has* a world, ‘*my* world’, which belongs only to that will. Each metaphysical subject is thus the limit of their world and their world only. Each ‘*my* world’ is unique because each has a unique limit, namely the will of a particular metaphysical subject, just as the visual field of each eye is unique to it. Further, as the metaphysical subject is only *a* limit of ‘*the* world’, this implies that there are *many* limits to ‘*the* world’. Indeed, each metaphysical subject is the limit of their own world, so there are as many ‘*my* worlds’ as there are metaphysical subjects. However, Wittgenstein is not here postulating multiple ontologically distinct worlds, for there is only one world of which our one language is a mirror-image, but rather multiple ontologically distinct wills (i.e. contra strict solipsism). Again, to employ the eye analogy, even though each eye sees the world from its own unique perspective, all eyes still look out onto the same world and so likewise for metaphysical subjects. This definition of the metaphysical subject leads Wittgenstein to the conclusion that what he calls solipsism coincides with «pure realism» because «the [metaphysical] self...shrinks to a point without extension, and there remains the reality co-ordinated with it» (5.64).²⁷ Here we find the agreement of the realist, that there is a world (‘*the* world’) and it is all that is the case, with the solipsist’s idea of a metaphysical self (‘*my* world’), which is a point without extension, that is, a unique limit of the world that it is attached to.

Wittgenstein uses his account of the metaphysical self to sharply distinguish it from, and

²⁵. Wittgenstein does not use a consistent terminology in the *Tractatus*. In 5.631 Wittgenstein talks of a «subject» and «will» and it is not clear if he means the psychological or metaphysical self; or rather he seems to mean both. In 5.632 we still find talk of a «subject», but by 5.633 we see the introduction of a «metaphysical subject»; 5.64 has «the self of solipsism», and then 5.641 has «the self in a non-psychological way»; finally, in 5.641 we have: «The philosophical self is not the human being, nor the human body, or the human soul, with which psychology deals, but rather the metaphysical subject, the limit of the world — not a part of it.» Wittgenstein’s inconsistent terminology makes him somewhat difficult to interpret. However, he is in fact defining two terms; this essay will exclusively and consistently employ the terms, the ‘psychological self’ and the ‘metaphysical self’, to represent this.

²⁶. However, it should be noted that Wittgenstein here forwards only an analogy and not an actual argument for the existence of this metaphysical self. Further, this analogy is surprisingly weak, as it compares the relation of two things *in* the world (the eye and its visual field) with the world as a whole and a thing ‘beyond’ it.

²⁷. Again, Wittgenstein is using solipsism in an unusual way. Probably the solipsist from here may claim that there is only *the* metaphysical subject (‘I’) and the world, but the solipsist would not grant, as does Wittgenstein, that there are *other* metaphysical subjects likewise co-ordinated with the world.

critique, what he calls the 'psychological self'. The psychological self is the self *in* the world; thus we can talk about it and state facts. Wittgenstein claims «that there is no such thing as the soul — the [psychological] subject etc — as it is conceived in the superficial psychology of the present day» (5.5421).²⁸ The argument Wittgenstein employs to justify this claim begins by pointing out that, given the Tractarian framework, all facts are ultimately composed of a relationship between simples and thus all facts are composite. But the psychological self, as it is conceived, is *necessarily* a whole, an indivisible unity, because «a composite soul would no longer be a soul» (5.5421). However, as the psychological self is clearly not a simple, and as the psychological self cannot be composite, it follows that the psychological self cannot be a fact in the world. Therefore, there is indeed no such *unity* as is referred to as the psychological self; at best we have only an unconnected collection of psychological facts about a particular brain and so on. It is only philosophy, Wittgenstein claims, speaking in a «non-psychological way», which can talk about the metaphysical self, even though nothing can actually be *said*, as it is *beyond* the world (5.641).

The metaphysical self, or will, being *a* limit of '*the world*', is thus necessarily related to the world in some way. The realism of Wittgenstein's position necessarily implies that there is only one world, '*the world*'. Obviously '*the world*' will continue to exist regardless of which, if any, metaphysical subjects are co-ordinated with it. However, '*my world*' cannot exist if '*the world*' did not, because in such a case there would be no reality for the metaphysical subject to be co-ordinated with. Therefore, the *ontological* relationship between the will and the world is one where the will is ontologically dependent on the world and the world is independent of the will.

Wittgenstein also examines the *causal* relationship between the will and the world. He argues that there are no logical grounds for believing in the law of causality, or for believing that there is a compulsion making one thing happen because another has happened (6.361). That is, things *can*, logically speaking, happen otherwise than they do, and for Wittgenstein there is *only* logical necessity (6.37). But as there «is no *logical* connexion between the will and the world» (6.374), it follows that my will cannot interfere with the facts of the world because everything in the world that happens is, for Wittgenstein, merely accidental (except what is logically necessary). But any action that is merely accidental cannot be a willed action, because an action is, by definition, not willed if it is merely accidental. Thus it follows that the facts of the world are causally independent of my will. Hence, freedom of will, in terms of action *in* the world, is nothing more than the impossibility of knowing what actions lie in the future (5.1362), outside of those which are logically necessary or impossible. Therefore, there is neither a logical nor contingent physical relationship between my will and the world (6.374) and thus the two are causally independent.

However, while I have already shown that the metaphysical subject cannot will any actions *in* the world, this does not imply that the will is completely impotent. To return to the eye analogy, though the eye is impotent in causing *what* is seen in the visual field, it can still choose, in some sense, *how* the visual field is seen. Wittgenstein states the analogous case, in the *Notebooks*, as follows: «The will is an *attitude* of the [metaphysical] subject to the world».²⁹ Thus we find that the only non-ontological relationship of will to world is one of

²⁸. This is repeated slightly differently in the *Notebooks* as: «The thinking subject is surely mere illusion» (80).

²⁹. Wittgenstein, *Notebooks 1914-1916* 87.

the *attitude* of the former to the latter.

Further, it is the metaphysical subject, beyond the world, which for Wittgenstein is «first and foremost the *bearer* of good and evil».³⁰ It is not the will itself which is good or evil, but rather it is predicated as such as a consequence of its *exercise*. This «good or bad *exercise* of the will», though, cannot alter the world itself (as the will is causally impotent), but only «the limits of the world» (6.43). However, I have already shown that the only exercise the metaphysical subject can make of its will is to change its *attitude* towards the world. This then implies that some attitudes of the metaphysical subject can be predicated as ethically good and others as ethically bad and further, that this attitude can directly affect the *limits* of the world.

However, this exercising of the will cannot change what is *in* the world, so its effect can only be to change the *entire* world itself, through altering its limits — of course, it is only ‘*my* world’ that can be changed by this exercise of the will and not ‘*the* world’. The metaphysical self is a limit of ‘*the* world’ and *the* limit of ‘*my* world’; hence ‘*my* world’ will entirely change depending on any change in the limit relationship between *my* will and ‘*the* world’. So through the will’s exercise, through its ‘expansion and contraction’, so to speak, the limits of ‘*my* world’ can change, thus causing ‘*my* world’ itself to «wax and wane, as a whole» (6.43).³¹ This implies that a change in the will’s attitude towards the world can change that will’s entire world.

Further, Wittgenstein also argues that there «must indeed be some kind of ethical reward and ethical punishment, but they must reside in the action itself» and not in the «consequences» or «events» that result from the action (6.422). Firstly, Wittgenstein is here repeating a claim made earlier that the consequences of an action cannot be ethically important because all facts are accidental and of equal value, which implies that the consequences of any action *in* the world are ethically neutral. However ‘consequences’ not *in* the world, but occurring to the metaphysical subject beyond the world, are ethically important. The action itself, that is, the will’s exercise of changing its attitude towards the world, must in itself be ethically rewarding or punishing. But this reward and punishment can only be reflected in the limits of ‘*my* world’ changing, as this is the only possible consequence of the will’s exercise. From this it follows that the actual changing of the will’s attitude towards the world is in-itself rewarding or punishing.

In the *Notebooks* Wittgenstein makes explicit what is only implied in the *Tractatus*: that «the happy life is good, the unhappy bad»,³² or more precisely, the good exercise of the will results in a happy world, the bad exercise of the will in an unhappy world. Clearly, when Wittgenstein is talking about the ‘happy life’ he is not referring to psychological or physiological states, because: «Physiological life is of course not ‘Life’. And neither is psychological life. Life is the world».³³ So the happy life is not a life where the

³⁰. Ibid. 76. Italics are mine.

³¹. «The good or bad *exercise* of the will does not alter the world, it can alter only the limits of the world, not the facts...In short the effect must be that it becomes an altogether different world. It must, so to speak, wax and wane as a whole» (6.43).

³². Wittgenstein, *Notebooks 1914-1916* 78.

³³. Ibid.

psychological subject, if one can even speak meaningfully of such an entity, experiences much happiness, but rather it consists in the metaphysical subject's world being a 'happy world'. The *Tractatus* puts this as follows: «The *world* of the happy man is a different one from that of the unhappy man» (6.43). However, this happiness can only be a result of the will's attitude towards the world, so that it cannot be anything *in* the world that makes the difference between a happy world and an unhappy one, but rather the *whole* world itself must be different. Thus, we find that what is rewarding about the ethically good exercise of the will is that it results in a happy world; likewise, an unhappy world is the punishment of an unethical exercising of the will.

These considerations lead to the conclusion that the happy world is ethically good and that it is brought about by the good exercise of the will. Thus the happy world, or happy life, is good in-itself; for Wittgenstein the happy life is tautologically the *only* right life to live.³⁴ Wittgenstein then claims that the happy life, which we ought to live, «is in some sense more *harmonious* than the unhappy».³⁵ This then leads us to the problem of defining what a 'harmonious' life is. But Wittgenstein, understandably, states that no such 'objective' mark of the happy, harmonious life *can* be given, because even if there is any such mark it will be beyond language; this is because it would apply only to the metaphysical subject who is beyond the world.

However, even given such a disclaimer, Wittgenstein does still try to 'point', in some way, to this ineffable happy life. For example, Wittgenstein writes:

that the man who is happy is fulfilling the purpose of existence...We could say the man is fulfilling the purpose of existence who no longer needs to have any purpose except to live. That is to say, who is *content*.³⁶

This statement then implies that the *attitude* of the happy subject is one of complete *contented* acceptance of the world, exactly as it is. The unhappy subject is one whose attitude towards the world is one of a desire for it to be different to how it is. Such a person remains forever unhappy, because one cannot causally bend the happenings of the world to one's will. The *contentment* of the ethically good happy life is its own reward, just as the discontentment of the ethically evil unhappy life is its own punishment.

However, given that there is much evil in the world, why ought we be happy with the world, even if we can't change it? Wittgenstein's answer is found in his understanding of God. In the *Tractatus* Wittgenstein writes: «*How* things are in the world is a matter of complete indifference for what is higher. God does not reveal himself *in* the world» (6.4321). This repeats an earlier claim, that all actions *in* the world, that is, all facts, are ethically neutral. But it also makes the important claim that *how* the world is does not concern God and this is just as well, because we have already seen that the will cannot influence *how* the world is, but can only take up an attitude towards it. Wittgenstein expands these considerations in the *Notebooks*, where he writes:

What we are dependent on we can call God. In this sense God would simply be fate, or, what is the same thing:

³⁴. Ibid. 78.

³⁵. Ibid.

³⁶. Ibid. 73.

The world-which is independent of our will.³⁷

It has been shown that it is *how* the world is that is independent of our will. So, in Wittgenstein's sense, it is God that we are dependent on and this implies that, as Wittgenstein makes explicit: «How things stand, is God. God is, how things stand».³⁸ So although God does not reveal himself *in* the world, he reveals himself *as* the way the world is, that is, what we are necessarily dependent on.

These considerations at last allow us to fully comprehend Wittgenstein's ethical thought, by explaining *why* he takes living in harmony with the world to be ethically good in-itself. If God is the way the world is, it follows that by living happily we are living:

in agreement with that alien will on which I appear dependent [God]. That is to say: 'I am doing the will of God'.³⁹

By being happy with the way the world is, however it is, we are in a sense doing God's will, because God *is* the way the world is; we thus are living 'harmoniously' with God. This is *why* it is ethically good to take an attitude of happy contentment towards the world, because the world is, in some sense, God. Thus Wittgenstein here confronts the so-called 'problem of evil' by claiming that *in* the world there is no evil (or good), but that the world as a whole is necessarily good, as it is a manifestation of God's will. By being in agreement with the world as a whole we are in agreement with God and this is *why*, for Wittgenstein, we ought to live happily, or equivalently, live ethically.

5. Conclusion

Wittgenstein's negative ethical thesis claims that there can be no ethical propositions. His positive ethical thesis, which relies upon, but goes beyond, the negative thesis, holds that the ethical good, or good in-itself, is the rewarding happy life. The happy life involves living in perfect contented harmony with the world, however it is, because how the world is, is a manifestation of God's will. Given the negative thesis, the positive thesis cannot strictly speaking even be said. We can only make sense of this by assuming that Wittgenstein takes this positive thesis to be 'illuminating nonsense' and not mere gibberish.⁴⁰

This approach is not new to philosophy. For example, Hannah Arendt writes:

both [Plato and Aristotle] ...considered this dialogical thought process to be the way to prepare the soul and lead the mind to a beholding of truth beyond thought and beyond speech — a truth that is *arrhēton*, incapable of being communicated through words, as Plato put it, or beyond speech, as in Aristotle.⁴¹

Similarly, Lynette Reid quotes the following Elizabeth Anscombe discussion of Wittgenstein:

³⁷. Ibid. 74.

³⁸. Ibid. 79.

³⁹. Ibid. 75.

⁴⁰. As Lugg notes, if Wittgenstein's propositions are nonsensical (as the conclusion states), then what are we to make of the claim that his thoughts express truth (as the preface states) —see Andrew Lugg, «Wittgenstein's *Tractatus*: True Thoughts and Nonsensical Propositions,» *Philosophical Investigations* 26, no. 4 (2003) 57.

⁴¹. Hannah Arendt, *The Human Condition* (Chicago: The University of Chicago Press, 1958) 291.

«There is indeed much which is inexpressible — which we must not try to state, but must *contemplate* without words». ⁴² Reid criticises Anscombe, but her criticism rests on an equation of *contemplation* with *thinking*. As Arendt makes clear, the two are very different things. Thinking is a kind of linguistic activity, whereas contemplation is non-active and non-linguistic — it involves *seeing* the world aright, not *saying* how it is. We might think of Wittgenstein's ladder as a way out of Plato's cave, but as Plato once pointed out, when we bring outside 'Ideas' back into the cave, they will become laughable, or rather, nonsensical. ⁴³

The positive thesis is of course highly problematic, not merely because it relies on the strange notion of 'ineffable truth', but also because Wittgenstein offers no good argument for why we ought to accept that the metaphysical subject exists 'mystically' beyond the world. One might defend this view along Kantian lines by arguing that we must affirm that a metaphysical subject exists beyond the world because otherwise we cannot think of ourselves as free moral beings. Such an argument seems to fit in well with the general Kantian (or Schopenhauerian) overtones of Wittgenstein's account of the will. ⁴⁴ However, the main difference between Kant and Wittgenstein here is that Kant sees the will as able to somehow effect the way the world unfolds through forming maxims of action, whereas for Wittgenstein the will can do no more than take up a very extreme form of stoicism and just be happy with the way things inevitably are.

Paul Formosa
School of History, Philosophy, Religion and Classics
University of Queensland
Brisbane, Queensland, 4072
Australia
p.formosa@uq.edu.au

⁴². Lynette Reid, «Wittgenstein's Ladder: The *Tractatus* and Nonsense,» *Philosophical Investigations* 21, no. 2 (1998) 99. Italics mine.

⁴³. In a discussion of Plato, Arendt writes: «It is not the *logos* that convinces them, but what they see with the eyes of the mind, and the Parable of the Cave is also in part a tale of the impossibility of translating convincingly such seen evidence into words and argument» — Hannah Arendt, «Some Questions of Moral Philosophy,» in *Responsibility and Judgment*, ed. Jerome Kohn (New York: Schocken Books, 2003) 88. By this I do not mean to imply that Wittgenstein was a Platonist, but rather only that Plato's metaphor of the cave gives us an interesting way to think about Wittgenstein's metaphor of the ladder.

⁴⁴. For a discussion of Kant and Wittgenstein see John M Weyls, «Wittgenstein: Transcendental Idealist?,» *Sorites* 14 (2002).

IS THE YELLOW BALL GREEN?

by Jack Lee

It appears that the proposition «the yellow ball is green» is contradictory. It is believed that yellow is not green. If yellow is not green, then the yellow ball is not green, either. Given this, to claim that «the yellow ball is green» would entail that «the yellow ball is *not* green, and, at the same time, *is* green.» In this paper, however, I want to show that this proposition is intelligible. In other words, I propose to argue that the yellow ball is, in a very important sense, green.

I

In physical science, there are basically two kinds of signals: (1) the analog signals, and (2) the digital signals. An analog signal is a physical quantity (or quality) which conveys information in a smooth and continuous form. For example, $f(t) = \sin(t)$ can be described as analog signal. In contrast to analog signal, a digital signal is an ordered sequence of information selected from a finite number of discrete quantities (or values). An easy and popular example is the dichotomy signal. We can use high/low, on/off, or 1/0...etc. to express the quantity (or value) of the signal. This makes the digital signals easier to handle. In short, we can manipulate digital signals with greater accuracy and efficiency than analog models would allow. Fortunately, an analog signal can be transformed into a digital one.

Likewise, in expressing (or describing) the objects or the states of affairs, there are basically two kinds of concepts: (1) digital concepts, and (2) analog concepts. The digital concepts are the concepts that generally cover (or include) merely a finite number of discrete categories (or paradigms). «Husband», «wife», «parent», and «sibling» are all among the digital concepts. Indeed, a person P is either a husband or *not* a husband. There is no such thing as being a little of a husband.¹ Of course, «ball», «stone», and «cat» are also among the digital concepts. On the other hand, the analog concepts are continuous concepts. Therefore, in theory, they can cover (or include) an infinite number of categories (or paradigms). «Color», «speed», «height», for example, are all among them. Given the characteristic mentioned above, we seldom (or never) apply our analog concepts directly. Conversely, we are inclined to apply them digitally. As Ronald de Sousa correctly says:

Whatever we know, we must categorize. From the continuum of our experience, we need to extract a finite and relatively fixed number of categories...²

¹. Ronald de Sousa, «Love Undigitized», in *Love Analyzed*, ed. Roger E. Lamb (Boulder, Colo.: Westview Press, 1997), p. 196.

². Ibid., p. 197.

Yet, to efficiently apply our analog concepts, we should digitalize them first. Indeed, digitalization can provide the needed convenience of categories that we consider as discrete.

We can sometimes see a beautiful rainbow in the sky after it rains. How many colors does the rainbow have? To this question, some might answer, «Of course, five!» The others may insist, «Seven!» In fact, «color» is a continuous (or analog) concept, so, in theory, there can be an infinite number of colors in rainbow.³ In practice, we assign color names to *roughly* mark out areas within a continuous spectrum of hues. But this practice may be different among different peoples or cultures. For example, it is possible for a group of people (the S tribe) living near the polar circle to respectively assign several color names to the corresponding areas within a spectrum of hues which people living in the USA call «white». Put loosely, different whites may have different meanings for the S tribe—danger, safety, predatory moments...and so forth. How precisely should we digitalize our analog concepts? Or how many parts should we divide an analog object or state of affairs (e.g., «color») into? In a sense, the digitalization is quite arbitrary. Indeed, we can divide «color» into any number of parts in terms of our common needs. In practical exercise, we merely assign a finite number of color names to mark out areas within a continuous spectrum of hues. For normally people are not able to sharply discriminate quite a lot of colors (say, 50 colors).

The digital concepts conform to the principle of non-contradiction; whereas the analog concepts do not. The claim that «this is a ball and this is not a ball in the room» is contradictory. So is the claim that «P is Q's father and P is not Q's father.» On the other hand, given that the speed of my bike ranges from 0 km/hr to 100 km/hr, it is true that the speed of my bike is 50 km/hr. It is also true that the speed of my bike is 90km/hr. Accordingly, it is not contradictory to claim that the speed of my bike is both 50 km/hr and 90 km/hr.⁴

In any case, my bike has an infinity of speeds within the range of 0 km/hr and 100 km/hr. Now the question here is, «Would the digitalized (analog) concepts (e.g., quick/slow, yellow/green, big/small) conform to the principle of non-contradiction?» It appears that they conform to this principle. Indeed, as shown above, the digital concepts *do* conform to the principle of non-contradiction. And the digitalized (analog) concepts are a certain kind of digital concepts. One might argue for this point as follows.

Suppose the speed of my bike ranges from 0 km/hr to 150 km/hr. We can, therefore, digitalize its speed-range into three zones: the low-speed zone (0 km/hr-50 km/hr); the middle-speed zone (50 km/hr-100 km/hr); and the high-speed zone (100 km/hr-150 km/hr). The high-speed zone is also called «the 125 km/hr-speed zone» because all speeds within this zone are assigned as 125 km/hr. Likewise, the middle-speed zone is also called «the 75 km/hr-speed zone»; and the low-speed zone called «the 25 km/hr-speed zone». Now speed 100 km/hr is both in the middle-speed zone and in the high-speed zone. However, «high-speed» is not

³. Austen Clark also dresses this point. He says:

If one examines the sky at sunset on a clear night, one seems to see a continuum of colors from reds, oranges, and yellows to a deep blue-black. Between any two colored points in the sky there seem to be other colored points. Furthermore, the changes in color across the sky appear to be continuous. Although the colors at the zenith and the horizon are obviously distinct, nowhere in the sky can one see any color borders, and every sufficiently small region of the sky is made up of regions that all seem to be of the same color.

See Austen Clark, «The Particulate instantiation of Homogeneous Pink», *Synthese* 80 (1989), p. 277.

⁴. Note that «the speed of my bike at $t=85$ km/hr» is a digital concept. Thus, it is contradictory to claim that «the speed of my bike is both 85 km/hr and 90 km/hr at t ».

«middle-speed». Thus, to say that «speed 100 km/hr is in the middle-speed zone and in the high-speed zone» would seem to entail that «speed 100 km/hr is in the middle-speed zone and *not* in the middle-speed zone.» This appears contradictory. However, I suggest that, in fact, it is *not*. Remember that these digitalized concepts are in reality analog. Let me explicate this point.

In the 75 km/hr-speed zone (or the middle-speed zone), only speed 75 km/hr can precisely be assigned as 75 km/hr. As for speed 80 km/hr, it is said that it is very *close* to 75 km/hr, and, therefore, is also treated as 75 km/hr. Similarly, the speed 100 km/hr is assigned as 75 km/hr. On the other hand, in the 125 km/hr-speed zone (or the high-speed zone), only speed 125 km/hr can be precisely assigned as 125 km/hr. As to speed 110 km/hr, precisely speaking, it is very *close* to 125 km/hr and, thus, also can be treated as 125 km/hr. Similarly, the speed 100 km/hr is assigned as 125 km/hr. Given this, speed 100 km/hr is in a sense in the 75 km/hr-speed zone (or the middle-speed zone), and in another sense in the 125 km/hr-speed zone (or the high-speed zone). In other words, 100 km/hr is in a sense assigned as 75 km/hr, and in another sense assigned as 125 km/hr. Therefore, there is no contradiction here.

At this point, I would like to further argue that «high-speed» can also in a sense be «middle-speed». As we have shown, since speed 100 km/hr is *close* to 75 km/hr, we assign it as 75 km/hr. What about speed 101 km/hr? Isn't it *close* to 75 km/hr, as well? If so, we can also, in this sense, assign it as 75 km/hr (or call it «middle-speed».⁵ Since 101 km/hr is in the high-speed zone, and since it can also be assigned as «middle-speed» as shown above, «high-speed» is «middle-speed» in this very sense. We are now in a good position to justify that the yellow ball is green.

II

The lights we can see are the ones with frequencies ranging from about 4.3×10^{14} Hz to about 7.5×10^{14} Hz. Within this frequency range, a light, in an appropriate condition, with a certain frequency would cause us to perceive a certain color. There are an infinite number of frequency values (or quantities) within this range. Put in another way, this frequency-range can be infinitely divided. Thus, *in theory*, there are an infinity of colors within this frequency-range.⁶ In order to handle it well, we should digitalize it. We customarily use the term «yellow» to refer to those colors caused by lights with frequencies ranging from about 5.0×10^{14} Hz to about 5.5×10^{14} Hz. The term «yellow» in fact indicates many (or an infinity of) colors within a certain area of spectrum of hues. Next to yellow is green. The term «green» is customarily used to call those colors whose frequencies range from about 5.5×10^{14} Hz to about 6.0×10^{14} Hz.⁷ Here, the exact yellow color can be stipulated as the color which is caused by the light with frequency at 5.25×10^{14} Hz. Call it color-525.⁸ As to color-530,

⁵. In fact, how precisely should we digitalize our analogous objects (or states or affairs) can be very arbitrary in terms of our common needs. We can also, for some reason, stipulate «middle-speed» as the speed-range between 50 km/hr and 120 km/hr, and «high-speed» as the speed-range between 120 km/hr and 150 km/hr.

⁶. In claiming that there are, in theory, an infinity of colors within this frequency-range, I do not rule out the possibility that, in practice, we perceive colors digitally.

⁷. See Vincent P. Coletta, *College Physics* (New York: McGraw-Hill, 1995), p. 633.

⁸. For the rest of this paper, I call a color caused by the light with its frequency at $P \times 10^{12}$ Hz, «color-P».

precisely speaking, it is very *close* to the exact yellow color. In fact, in a strictly philosophical sense, we cannot say that it is yellow. However, in practical exercise, both color-525 and color-530 are called «yellow». Or both are assigned to the same category—«yellow». Likewise, the exact green color can be stipulated as color-575. If so, color-570 can only be said to be very *close* to the exact green color. Since it is very close to the exact green color, it can be assigned as «green». Here, we can actually locate a color just at the border of yellow and green. Suppose this is color-550. Since color-550 is *close* to color-525, it is called (or assigned as) «yellow». On the other hand, it is *close* to color-575 too, it is also assigned as «green». Therefore, color-550 is yellow, and, at the same time, green. There is no contradiction here.

Suppose there is a ball whose color is color-545. As mentioned above, color-550 is *close* to the exact yellow color (i.e., color-525), it is called «yellow». Clearly, color-545 is closer to the exact yellow than color-550 (or put loosely, color-545 is *yellower* than color-550). Since color-550 is assigned as «yellow», color-545 should be assigned as «yellow» too. Therefore, this is a yellow ball.

On the other hand, as stated above, color-550 can also be assigned as «green» because it is *close* to color-575. Color-545 is very close to color-550, and, in a sense, also *close* to color-575.⁹ Thus, it can also be assigned as green, in this sense. Given this, *this (yellow) ball can also be green in a sense*. There is no contradiction here. Indeed, in terms of «vagueness», V. J. McGill and W. T. Parry claim:

...[i]n any concrete continuum there is a stretch where something is both *A* and $\sim A$... There is a sense in which the ranges of application of red and non-red [in so far as «red» is vague] overlap, and the law of non-contradiction does not hold.¹⁰

Similarly, Dominic Hyde writes, «...a man growing a beard is at some stage both bearded and not bearded.»¹¹ He goes further, saying that:

When a vague predicate is applied to a borderline case we are confronted by a sentence which is neither determinately true...nor determinately false...but indeterminate, which now amounts to the claim that the sentence is both true and false. It is true since true on some admissible precisifications and false since false on some.¹²

To sum up, the analog concepts do not conform to the principle of non-contradiction. To be handled well, they must be digitalized first. But note that these digitalized concepts are in reality analog. Because of this, and, on the basis of our above discussion, the following propositions are all plausible:

- (1) The high New York World Trade Center tower is low.
- (2) The heavy stone is light.

⁹. There is no reason for us to insist that «green» should cover merely those colors ranging from color-550 to color-600. We can also stipulate that «green» covers those colors ranging from color-540 to color-610 in terms of our common needs.

¹⁰. V. J. McGill and W. T. Parry, «The Unity of Opposites: A Dialectical Principle», *Science and Society* 12 (1948), p. 428.

¹¹. Dominic Hyde, «From Heaps and Gaps to Heaps of Gluts», *Mind* 106 (1997), p. 645.

¹². *Ibid.*, p. 649.

(3) The long Nile River is short.¹³

Jack Lee
Center for General Education and Core Curriculum
Tamkang University, *Taiwan*
jlee@mail.tku.edu.tw

¹³. I am very grateful to Dr. Graham Priest (University of Melbourne) and Dr. Dominic Hyde (University of Queensland) for their comments on earlier drafts of this paper. Thanks also to Dr. Cheng-Hung Lin (Soochow University) for his help on the topic of paradox.

INCOMMENSURABILITY AND INTERPRETATION

by **Anthony D. Baldino**

«‘Incommensurable’ is, of course, Kuhn and Feyerabend’s word for ‘not intertranslatable’», says Donald Davidson (1984, 190) in ‘On the Very Idea of a Conceptual Scheme’, and so, in presenting a general argument against the possibility of making sense of languages that are not intertranslatable, Davidson thinks he has an argument against incommensurability as advocated by Thomas Kuhn and Paul Feyerabend. I think that, after trying to pinpoint just what that argument would look like, we will find no satisfactory formulation and will be forced to conclude that the thesis of incommensurability is more subtle and resilient than Davidson imagines. I will try to indicate why, after a period of being held in low philosophical esteem in some quarters, the thesis is due for a reexamination and reassessment and will argue that it should return to being regarded the deeply illuminating thesis about scientific progress that many considered it when it was first introduced.

I. Davidson’s Arguments Against Incommensurability

[A] An argument based on shared agreement

According to the defenders of incommensurability, many important competing scientific theories, theories that replace each other through scientific revolution (e.g. Ptolemaic and Copernican astronomy, Newtonian mechanics and Einsteinian relativity), are said to be incommensurable. The theories in question are committed to radically different central or fundamental principles; the principles of one theory undermine or exclude those of its rival; they thereby undermine or exclude the whole of the rival theory built upon those fundamentals. That is, when the fundamental principles of one theory are in place, those of the other theory cannot be, and all of the concepts and so-called facts that have their foundation in these displaced principles are rendered absurd or meaningless — the theories are thus said to be incommensurable with one another.

Davidson’s argument against non-intertranslatable languages and incommensurable conceptual schemes is based on his theory of meaning, which in turn rests on his conception of radical interpretation. Meaning is given in the production of a truth theory for the agent being interpreted. The truth-theory is produced via radical interpretation — the interpreter must proceed only upon the empirical evidence of which sentences under which circumstances are held true by the interpretee. Since Davidson does not think that under such constraints the interpreter can assign beliefs to the interpretee independent of assigning meanings to her utterances, the production of a truth-theory must proceed by simultaneously producing the interrelated constructs of belief and meaning. Such a truth theory, which will take the form

of T-sentences¹, will be accepted as an interpretation of the sentences of the interpretee if the totality of T-sentences optimally accounts for all the empirical evidence we have assumed access to. Such a theory, then, does not produce meaning sentence by sentence but holistically; a sentence is given an interpretation by the right hand side of the relevant T-sentence if the interpreter knows the empirically adequate truth-theory to which that sentence belongs.

A fuller description of Davidson's theory of meaning will, of course, have to say more about how we go about producing such a truth-theory, that is, how beliefs and meanings are to be produced simultaneously and the related question of how a truth-theory is judged empirically adequate. But with even just this much said about radical interpretation, we might glimpse a speedy dismissal of incommensurability. We might be tempted to use an argument Davidson uses explicitly against the skeptic for a quick refutation of the possibility of concepts of one point of view being incommensurable with those of some other viewpoint.

Davidson writes:

We can, however, take it as given that *most* beliefs are correct. The reason for this is that a belief is identified by its location in a pattern of beliefs; it is this pattern that determines the subject matter of the belief, what the belief is about. Before some object in, or aspect of, the world can become part of the subject matter of the belief (true or false) there must be endless true beliefs about the subject matter. (Davidson 1984, 168)

The point here is that meaningful error presupposes a foundation in truth, but essentially the same point is sometimes put by Davidson more explicitly in terms of the project of interpretation, that is, that meaningful disagreement between interpreter and interpretee must have a foundation in agreement.

A skeptic may or may not be committed to meaningful error, and so perhaps this is a good anti-skeptical argument. But the incommensurabilist is absolutely not committed to meaningful disagreement — it is exactly that that the incommensurabilist wants to deny — holders of incommensurable theories are doomed to misapprehend and talk past one another. And so Davidson's argument, stated as baldly as that, cannot refute the incommensurabilist. The incommensurabilist denies both meaningful agreement and meaningful disagreement between holders of incommensurable theories, and so it does not matter to him that the latter entails the former.

[B] An argument based on the principle of charity

It would seem that if the radical interpretation approach to meaning is to be the source of an anti-incommensurabilist argument, we must look more deeply into what constitutes that approach to reveal that argument. In particular, we must pay more attention to what conditions on an acceptable truth-theory must be imposed and what must be done in the process of interpretation to meet those conditions.

The principal difficulty confronting the radical interpreter concerns the interdependence of belief and meaning that was mentioned earlier. As a result of this interdependence, the interpreter has to find a way, given the evidence of which sentences an agent holds true under which conditions, to arrive at both a theory of meaning and a theory of belief. Davidson's familiar proposal for accomplishing this is to hold one of these unknowns constant: we hold belief constant and solve for meaning. Davidson writes, 'We cannot take even a first step

¹. A sentence of the form, 'S is true if and only if R', where 'S' is the sentence to be interpreted and 'R' is the interpretation.

towards interpretation without knowing or assuming a great deal about the speaker's beliefs.' (Davidson 1984, 196)

The equally familiar principle by which we manage the task of holding belief constant is what Davidson calls the 'principle of charity' whereby we attribute beliefs to the agent so that we minimize unexplained error on the part of the agent. Given our empirical evidence which consists in assent and dissent to sentences in particular circumstances, we begin to generate correlations between sentences and truth conditions by 'assigning to sentences of a speaker conditions of truth that actually obtain (in our own opinion) just when the speaker holds those sentences true' (Davidson 1984, 196) and, presumably, do not obtain when the speaker holds the sentences false. This policy is 'to be modified in a host of obvious ways' (Davidson 1984, 152) — the goal is not the absurd one of making error or disagreement impossible, nor is it to simply maximize agreement and minimize error, but it is to 'countenance() error where it can be best explained' (Davidson 1984, 318). A truth-theory is an acceptable interpretation for a speaker, Davidson thinks, if it satisfies certain formal constraints and if it accounts for the empirical evidence of assent and dissent under the charitable constraint outlined.

The significant point about this principle of charity is that, since it is 'not an option, but a condition of having a workable theory, it is meaningless to suggest that we might fall into massive error by endorsing it' (Davidson 1984, 318). Charity is not an option because it is that principle which supplies the truth-theory with empirical content. It is only by some such principle that we can link what a speaker says with what we observe in the world (the conditions of utterance). This very basic methodology of interpretation instructs the interpreter that most of the time he may take it to be the case that what the speaker holds true is true, and it is by virtue of this methodological principle that the interpreter can 'take into account the causal interaction between world and speaker in order to find out what the speaker means, and hence what he believes' (Davidson 1986, 332).

The argument against conceptual schemes based on the principle of charity is then this: since charity is necessary to interpretation, not optional, shared agreement and conceptual contact between any two language-users is thereby necessary as well, since charity guarantees agreement and contact. The incommensurabilist appears to be confronted by the following uncomfortable dilemma: if he denies the principle of charity, he severs any ties between the interpretee's language and the observable features of the world, and it becomes rather mysterious how the person who uses such a language could have ever learnt the language in the first place; if he accepts the principle of charity as necessary, he accepts shared agreement as necessary and can no longer make his radical semantic claims.

This argument seems far more promising than the first simple one that disagreement rests on a foundation of shared agreement. This one seems to ensure agreement and, *a fortiori*, conceptual contact between interpreter and interpretee and thus seems to block the very possibility of large-scale failures of conceptual overlap and to render incoherent the very notion of incommensurable conceptual schemes. And it does so without appeal to meaningful disagreement, something the incommensurabilist explicitly denies.

However, I think if we look more closely we will see that this is not the case, and that, in fact, the two arguments resemble each other more closely than it may first appear and may ultimately be rejected for very similar reasons.

Whenever we approach a language to be interpreted, the argument we have been considering tells us that, in order to succeed, we have no choice but to employ the principle

of charity and thus that we can never be led into failure because we have employed that principle. It appears to follow from this that there can be no learnable language that does not share substantial agreement and conceptual contact with the language we come to the interpretation with, and thus, it seems Davidson has the argument that he claims he is after in 'On the Very Idea of a Conceptual Scheme', one that has as a conclusion that 'translatability into a familiar tongue (is) a criterion of languagehood' (Davidson 1984, 186), and, therefore, that there can be no languages that are not intertranslatable.

If one is convinced, as I am, that Davidson's account of meaning is essentially correct, is one also forced to abandon the thesis of incommensurability as untenable? Must one deny Davidson's principle of charity to be an incommensurabilist? I do not think so.

It might seem that the incommensurabilist must deny charity since charity implies agreement, which in turn implies conceptual contact. But note that interpretation in violation of charity is interpretation without shared agreement and therefore interpretation that posits wholesale disagreement between interpreter and interpretee. Davidson's compelling argument prohibits this, but that is not a problem for the incommensurabilist because, just as we had reason to point out before, the incommensurabilist is absolutely not committed to interpretation that posits wholesale disagreement. According to the descriptions given by Kuhn and Feyerabend, it is not that holders of incommensurable points of view disagree with one another — *rather, they do not understand each other*. But if they do not understand one another, they have *failed* to interpret one another. And if they have failed to interpret one another they cannot have possibly interpreted each other in violation of the principle of charity. They have simply come up short in their effort to produce an interpretation — they cannot clearly make out what the other is trying to say, what the concepts they are using can possibly mean. The incommensurabilist describes wholesale incomprehension not wholesale disagreement, not interpretation that violates the principle of charity.

The dilemma we set for the incommensurabilist is a false one — the incommensurabilist does not need to deny the very plausible principle of charity described by Davidson and thus does not cut off language from the world and rob words of all possible empirical content. Holders of incommensurable points of view do not come up with uncharitable interpretations of one another. They fail to interpret; they do not grasp what the other means at all. Why do such failures of interpretation take place? Ironically, maybe, they take place for a very Davidsonian reason. Interpretations, in cases of incommensurable theories, are impossible because in such cases the principle of charity cannot be satisfied. And, as Davidson tells us, nothing could count as a successful interpretation if it violates the principle of charity.

So the incommensurabilist can wholeheartedly agree that interpretation cannot be successful unless the interpreter satisfies the principle of charity, unless the interpreter can use what he observes to be true in the world to assign meanings to the interpretee's utterances. Indeed he can also concede that whenever the principle of charity is met in a satisfactory way a successful interpretation is achieved. But, importantly, he need not concede because it has not been shown that, whenever an observer uses what he observes to be true in the world to assign meanings to utterances, the interpreter has satisfied the principle of charity or succeeded in interpretation. The incommensurabilist may hold that a successful interpretation depends on the use of what is observed to be true in the world, but he need not also hold that all interpreters are capable of making the same observations — in fact that is just what he denies in stating the incommensurabilist position. (Surely it is wise to deny this. Although these are clearly not cases of incommensurability, it is still the case, for example, that I cannot make

the same observations as someone who understands cricket while watching a match with him, and that someone who does not read music or comprehend principles of tonality cannot make the same observations when looking at a musical score as someone who can.)

A successful interpretation surely rests on shared agreement. It simply does not follow from this that the interpreter, without going well beyond the language he came to the interpretation with, has the conceptual resources to agree with the interpretee, and it does not therefore follow that the interpreter is able to give an interpretation from that original language. That is precisely what the incommensurabilist has been denying from the start, and the present argument does nothing whatever to block such a denial.

It is important to see that the response I offer on behalf of the incommensurabilist does not deny that the principle of charity is a necessary constraint on interpretation nor that any interpretation that adequately satisfies this principle (and certain other formal constraints) is a successful interpretation. The incommensurabilist merely denies that, in all instances, conceptual contact between interpreter and interpretee is sufficient for the would-be interpreter to describe the truth conditions of the interpretee's utterances — that is, the interpreter lacks the conceptual resources to successfully meet the constraint of charity. In cases of incommensurability, one may use what one observes to be true of the world to assign meanings, one may look as best one might for shared agreement, but one will never succeed in satisfying the principle of charity, will never succeed in minimizing unexplained error. One may, to be sure, do the best one can to minimize unexplained error with the conceptual resources one has, but if the points of view are incommensurable in the way Kuhn and Feyerabend imagine, such an 'interpretation' will be just a great mass of unexplained error — the alternative point of view appears absurd, almost as if it were not a point of view at all.

In short, just because it has been shown by Davidson that you cannot possibly succeed without charity, it simply does not follow that you always have the concepts required to be truly charitable. In a case like this, the only way you could truly minimize error, truly satisfy charity, would be to experience a conversion of the sort the incommensurabilist describes, that is, to go native with respect to the concepts of the interpretee. In such a case only absurdity is detected when you obstinately insist on matching your interpretee's comments with observations about the world at the time of utterance that you are able to make only from within the confines of your original point of view. Understanding — true interpretation — is only achieved when the grip of your old concepts are loosened and you allow yourself to learn your interpretee's point of view from scratch, by going native, without detour through your own native tongue.

After such a conversion, successful interpretation becomes available.

After you have learned to observe the world as your former rival does, you have no trouble using what you are now able to observe to be true of the world to assign meanings to her sentences, no trouble locating a mass of shared agreement, no trouble minimizing unexplained error. The incommensurabilist has not severed meaning and interpretation from what can be observed in the world as he has been accused of doing; he merely claims that not all languages are capable of expressing the same observations, and, therefore, not all language users are prepared to make the same observations.

[C] An argument based on nonsense

We are almost ready to conclude that the argument from the principle of charity neither succeeds in showing that the criterion of languagehood is translatability into a familiar tongue, nor that there could be no languages that fail of intertranslatability, nor that the semantic claims of the incommensurabilist are incoherent.

One final attempt at a critique of incommensurability from the point of view of radical interpretation, however, suggests itself based on what we have said so far. We have noted that when an interpreter confronts an incommensurable conceptual scheme or untranslatable language, the so-called language or scheme appears absurd, almost as if it were not a scheme or a language at all. Davidson has said that is ‘tempting’ to say of such a situation that ‘nothing ... could count as evidence that some form of activity could not be interpreted in our language that was not at the same time evidence that that form of activity was not speech behavior.’ (Davidson 1984, 185) Why should the interpreter not follow this hunch and conclude that the sounds issuing from his ‘rival’s’ mouth are not really the statement of a point of view at all — they are merely sounds? Why, after being as charitable as he can be from his own point of view and still coming up with nonsense, should the would-be interpreter not just conclude that there was never anything there to interpret in the first place?

As to this argument, however, Kuhn, Feyerabend and I are all almost perfectly in accord with what Davidson goes on to say: ‘As fiat, the thesis (that a form of activity that cannot be interpreted as language in our language is not speech behavior) lacks the appeal of self-evidence.’ He adds, ‘If it is a truth, as I think it is, it should emerge as the conclusion of an argument.’ (Davidson 1984, 185,6) We all can agree that there should be an argument for such a thesis. Davidson thinks that he has produced one, and so that the thesis is true. If I am right in what I have said so far, no such argument has been produced.

To say that the thesis in question ‘lacks the appeal of self-evidence’ is not at all to say that the thesis has no intuitive appeal: why should we persist in seeking an interpretation when all we can see, from our native point of view, is nonsense? But it seems to me that the temptation to say that we are not dealing with speech behavior in such a case only derives from the harmless fact that, if we find that the person we try to interpret is speaking of none of the things we know how to speak about, we are left unable to even clearly imagine what he or she is talking about. The fact that we cannot imagine what our would-be interpretee is talking about is, as I say, harmless for Kuhn and Feyerabend since that is exactly what they think the meeting of incommensurable points of view is like. So the fact that we cannot imagine what they are speaking about could not be considered proof that they are not speaking about anything without obviously begging the question against the incommensurabilist. Moreover, not only do Kuhn and Feyerabend acknowledge the fact that one is tempted in these cases to conclude that one is dealing with mere nonsense, they exploit the fact to explain the stubborn resistance scientists of the old-guard typically display in the face of scientific revolution. The temptation may be natural (often only the greatest of scientists succeed in overcoming it), but there has been no argument that has successfully shown that succumbing to such a temptation would always be justified.

All that has been said in this essay up till now still leaves open the question of whether there actually are alternative conceptual schemes or languages that are not intertranslatable. It seems to me, though, that if there is any hope of finding such schemes, the history of science would be one of the more likely places to uncover them. How far do we really think

the Aristotelian would get toward interpreting what is said in the modern physics laboratory if he limited himself to matching what is said there to what he knew how to say before he entered the time-machine that transported him there? How confident ought we be that there is enough in common between the schemes in question for the Aristotelian to build modern concepts out of the ones he came to the interpretation with?

The conclusions reached thus far in this essay indicate that if we wanted to answer these last two questions we would have to consider the actual contents of the Aristotelian and modern points of view. That is, we cannot rule out, via some *a priori* argument, the very possibility of alternative conceptual schemes, and so, the incommensurabilist's commitment to such a possibility does not render his view of scientific theories untenable.

II. The Semantic Soundness of Incommensurability

[A] The consistency of incommensurability and radical interpretation

Of course, if we are going to say that there may be alternative conceptual schemes, that there are schemes that the radical interpreter has no access to while he remains within the conceptual confines of his original point of view, we should be prepared to give an account of how the stymied interpreter could go about grasping those foreign points of view — if we could not, we would begin to wonder how those who are already comfortable with that point of view ever got that way. The incommensurabilist does have such an account, one that I have alluded to earlier: incommensurable points of view are not grasped via translation but by a process which can be called 'going native' — they are learned from scratch without attempting to translate them into one's native scheme.²

But this recourse to the notion of going native might appear problematic within the larger context of this paper since the expressed goal is to show that nothing in the radical interpretation model militates against the thesis of incommensurability. It might be thought that, once we admit that to grasp the meaning of some foreign sentences it may be necessary to go native, we have simply abandoned the radical interpretation model. For upon that model, all there is to meaning is what is given in the radical interpreter's truth-theory, and that theory is constructed by correlating interpretee sentences with truth conditions given in the language of the interpreter. If the view is that all there is to meaning is given in the truth conditions that the radical interpreter must give *in his own language*, what place is there for going native once the view is adopted?

This apparent tension, however, rests on a fairly trivial confusion. The radical interpretation account is meant to be a theoretical account of meaning or linguistic competence — what a speaker of a language knows in knowing how to speak her language is what is given in the truth-theory of the radical interpreter. But the radical interpretation account, obviously, is not and was never meant to be an account of how we *come* to know a language. The radical interpreter must have a language to interpret into, and we all *come* to know at least our first language without a language to interpret into. In fact, not only do we *not* learn our first

². Feyerabend describes going native: 'Analogies, metaphors, negative characterizations, bits and pieces of cultural history are used to present a new semantic landscape with new concepts and new connections between them. Historians of science proceed in a similar way, but more systematically. Explaining, say, the notion of 'impetus' in sixteenth- and seventeenth-century science, they first teach their readers the physics, metaphysics, technology, and even the theology of the time: in other words, they too introduce a new and unfamiliar semantic landscape, and then show where impetus is located in it.' (Feyerabend 1987, 266.)

language from the point of view of the radical interpreter, we rarely learn any language or interpret any speech behavior from that point of view. Radical interpretation, if actually carried out, would be an immensely arduous affair, and in most ordinary circumstances we interpret others, whether they speak our native language or some foreign language we are trying to learn, by making use of all sorts of linguistic conventions, shortcuts which the radical interpreter cannot allow himself. But, to repeat, admitting this fact does not strip from the radical interpretation account its ability to fulfill its purpose, that of giving an account of what meaning is, not of how we come to know how to produce and understand sounds and signs that have meaning.

But if radical interpretation is silent regarding how we come to know a language, going native is not incompatible with that account. That is, to be in a position to be a radical interpreter, one must know a language, and so some language must be learned outside the process of radical interpretation, learned ‘from scratch, as a child learns’ (Feyerabend 1987, 266). If the radical interpretation model is compatible with a child’s learning from scratch, it cannot be incompatible with an adult’s (one who already has a language) learning that way — and indeed it was never meant to be incompatible with this.

Nothing the incommensurabilist says prevents him from agreeing that all there is to linguistic competence is what is given in the truth-theory of a successful radical interpretation (one that meets the empirical and formal constraints). Nothing the advocate of radical interpretation holds should prevent him from recognizing that not just anyone can succeed in the enterprise of radical interpretation. Surely, for instance, someone without a language would first have to learn a language from scratch in order to succeed, and there is nothing essential to the radical interpretation model which prevents one from holding that someone without the appropriate native tongue would also inevitably fail if she took up the stance of the radical interpreter without first going native with respect to the language interpreted. Once a language is learned by the would-be interpreter from scratch or by going native, the meaning of the sentences of that language are readily given in the homophonic translations which the interpreter may now give as the truth-theory for the language learned.

[B] Incommensurability and the scheme content distinction

Davidson’s main target in his compelling essay ‘On the Very Idea of a Conceptual Scheme’ is the third dogma — the scheme/content distinction. A central premise of that argument involves Davidson’s also compelling principle of charity as a constraint on interpretation. Davidson goes on to assert, as a sort of corollary of the main argument, that the principle of charity also contravenes the thesis of incommensurability as advocated by Kuhn and Feyerabend.

This last move, I have tried to establish, is a misstep. But this misstep does not impinge on the main thrust of Davidson’s essay — it does not vitiate the argument against the third dogma. That is because incommensurability does not imply any objectionable scheme/content distinction. Holders of incommensurable theories have different concepts. These concepts are supposed to somehow exclude one another, but in all of the ways Kuhn and Feyerabend have tried to explain how this happens, they never try to say that they exclude each other by carving up some content differently by using different schemes.

In fact, Davidson never really accuses the incommensurabilist of making this mistake. His argument against the incommensurabilist, as I say, stands apart from the main argument against scheme/content. It is the principle of charity that is supposed to undermine

incommensurability and, as I have tried to show, that component of Davidson's discussion is flawed. If this is so, then the denial of the scheme/content distinction, the principle of charity, and incommensurability can all coexist.

This is an important result in that it might help reconnect some segment of philosophy of science with a segment of philosophy of language. Many greeted incommensurability as a radical, intriguing, and powerfully informative conjecture about the history of science. But misgivings, like Davidson's, about the thesis's semantic integrity did a lot to cast a dark shadow of doubt over it and restrained much of the enthusiasm that first greeted the thesis. Over time, interestingly, some in the philosophy of science, leaving aside the semantic doubts, have come back to refer to and discuss incommensurability positively. Often enough, the thesis is referred to as a way of shedding light on some specific scientific impasse or controversy³. Those who use the thesis as this kind of elucidative tool do so by simply ignoring the semantic disrepute that the thesis had fallen into in some circles. In this essay, I have tried to establish that instead of ignoring the thesis's suspect reputation, we should challenge it.

References

- Albert, David. 1992. *Quantum Mechanics and Experience*. Cambridge: Harvard University Press.
- Davidson, Donald. 1986. A Coherence Theory of Truth and Knowledge. In *Truth and Interpretation*, ed. by Ernest LePore. Oxford: Basil Blackwell Ltd.
- Davidson, Donald. 1986. Empirical Content. In *Truth and Interpretation*, ed. by Ernest LePore. Oxford: Basil Blackwell Ltd.
- Davidson, Donald. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- Feyerabend, Paul. 1987. *Farewell to Reason*. London: Verso.

Anthony D. Baldino
Rutgers University
Anthony.Baldino@gfigroup.com

³. A very compelling instance of this as it relates to quantum mechanics can be found in Albert 1992.

ON THE SEMANTIC INDECISION OF VAGUE SINGULAR TERMS

by **Dan López de Sa**

According to a popular, plausible, but also controversial view about the nature of vagueness, vagueness is a matter of semantic indecision. In the words of one of its main defenders, David Lewis:

The only intelligible account of vagueness locates it in our thought and language. The reason it's vague where the outback begins is not that there's this thing, the outback, with imprecise borders; rather there are many things, with different borders, and nobody has been fool enough to try to enforce a choice of one of them as the official referent of the word 'outback.' Vagueness is semantic indecision. (Lewis 1986, 213)

Likewise for other vague singular terms, like 'Tibbles,' 'Everest,' 'Toronto,' and 'I.' Donald Smith (2006) argues that if 'I' is indeed vague, and the view of vagueness as semantic indecision is correct after all, then 'I' cannot refer to a composite material object. But his considerations would, if sound, also establish that 'Tibbles,' 'Everest,' or 'Toronto,' do not refer to composite material objects either—nor hence, presumably, to cats, mountains, or cities. And the considerations can be resisted, anyway. Or so I argue.

Smith observes that, according to the view of vagueness as semantic indecision, although it is definitely the case that there is something to which a vague singular term refers, nothing is such that it is definitely the case that it is what a vague singular term refers to—at least, on standard ways of characterizing what it is for something to satisfy a definiteness-involving matrix, see (McGee 1998). Therefore $(Tibbles_{dd})$ and (I_{dd}) are true, according to the view, but $(Tibbles_{dr})$ and (I_{dr}) are not:

$$(Tibbles_{dd}) \quad D\exists x(Tibbles = x);$$

$$(Tibbles_{dr}) \quad \exists xD(Tibbles = x);$$

$$(I_{dd}) \quad D\exists x(I = x);$$

$$(I_{dr}) \quad \exists xD(I = x);$$

(where 'D' stands for the notion of definiteness, and assuming variables are precise).

Next Smith observes that (if one believes that one exists) rejecting (I_{dr}) commits one to

ODD I exist and for all x either the use of 'I' in this sentence fails to refer to x or it is indefinite whether the use of 'I' in this sentence refers to x .

Although he does not say, in a similar manner (if one believes that Tibbles exists) rejecting $(Tibbles_{dr})$ commits one to

ODD? Tibbles exists and for all x either the use of ‘Tibbles’ in this sentence fails to refer to x or it is indefinite whether the use of ‘Tibbles’ in this sentence refers to x .

According to Smith, ODD is odd. The reason he gives, however, would also make ODD? odd, which it is not.

And the consideration is flawed, anyway. Here it is, in a variant involving Tibbles:

If I assert it, then I take myself to have successfully referred to Tibbles with the use of ‘Tibbles’ in ODD? . But Tibbles is one of the things the universal quantifier in ODD? quantifies over. So, if I assert ODD? , then I assert of myself that either I definitely fail to refer to Tibbles with the use of ‘Tibbles’ in ODD? or it is indefinite whether I refer to Tibbles with that use of ‘Tibbles’. So, if I assert ODD? , then I take myself to have successfully referred to Tibbles with the use of ‘Tibbles’ in ODD? and either to be such that the use of ‘Tibbles’ in ODD? definitely fails to refer to Tibbles or to be such that it is indefinite whether that use refers to Tibbles. But if I sincerely take myself to have successfully referred to Tibbles with a use of ‘Tibbles’, then I cannot sincerely take myself to be such that the use of ‘Tibbles’ in question definitely fails to refer to Tibbles. And it also seems that I cannot sincerely take myself to be such that it is indefinite whether that use refers to Tibbles. It is puzzling to be committed, and to see that one is committed, to the truth of something that one cannot sincerely assert. But this is just the position I would be in if I believed that Tibbles exist and that the negation of (Tibbles_{dr}) is true.

Obviously, the problem lies with the very first conditional. If ‘Tibbles’ is vague, and the view of vagueness as semantic indecision is correct, then, when I assert sentence containing it, I do not need to take myself to having successfully referred to any particular thing—if that is understood as definitely referring to something. Rather, I aim my statement to turn out true on any admissible way of making the semantic decisions that are not (and should not, and maybe cannot, be) made. But surely, on any admissible sharpening of ‘Tibbles,’ ‘Tibbles exists’ turns out to be true. Hence the first conjunct of ODD? is true no matter what, and perfectly compatible with the vagueness of ‘Tibbles’ in a way that also vindicates the second conjunct, according to the view of vagueness as semantic indecision. *Mutatis mutatis*, of course, for ‘T’ and ODD.

The other one consideration Smith offers seems more important, and concerns the problem of the many.¹ According to the view of vagueness as semantic indecision, nothing in the thoughts, experiences and practices of language users, nor in the way things are, determines any particular thing as the cat referred to by ‘Tibbles.’ Rather there is a number of different candidates that are equally eligible candidate referents for ‘Tibbles.’ Now, given that it is definitely the case that Tibbles is a cat, the different candidates seem to share those features that are relevant for something being a cat, and hence seem to have equal claim to be a cat. Hence the problem of the many: where it seemed to be one and just one cat, there turn out to be many candidates with equal claim to be the cat, and there is nothing in the vicinity with a better claim. So, instead of one cat, we seem to have many. *Mutatis mutandis*, again,

¹. Strictly speaking, Smith also submits the following «Cartesian» consideration: «In order for me to believe that something is the case, I must *definitely* be around to believe it. And this requires there to be something that is definitely identical with me...» But this certainly does *not* require that: it suffices that it is definitely the case that I am around to believe it. In any event, Smith himself does not seem to put much weight in his «Cartesian» consideration.

regarding the different candidates to which ‘I’ indeterminately refers, if it is vague and the view of vagueness as semantic indecision is correct: each seem to have equal claim to be a person.

One solution to the problem of the many accepts that the many are indeed cats (or people), and explain why this notwithstanding, the adequate response to the question ‘How many cats are there on the mat?’ is (typically) ‘Just one.,’ appealing to independently motivated facts about the pragmatics of counting, see (Lewis 1993). To the possibility that the many are people, Smith simply responds: «There simply is not, so it seems to me, this massive collection, pain-experiencing, action-performing entities.» One would certainly appreciate some reason for thinking that the Lewisian account of why it seemed to us that there simply is not a massive collection—even if strictly speaking there *is* one such—is faulty. For what it is worth, my own view is that, actually, one such «many» solution is the most promising one to the problem of the many, and certainly the solution that I think defenders of the view of vagueness as semantic indecision should ultimately adopt, see (López de Sa 2004).

But be this as it may, it is certainly not the *only* solution that defenders of the view of vagueness as semantic indecision can adopt—and have indeed adopted. One rival solution by disqualification is the so-called «supervaluationist» solution, mentioned also in (Lewis 1993), and more recently defended by (McGee & McLaughlin 2000), (Varzi 2001) and (Weatherson 2003). According to this alternative solution, each sharpening of ‘is a cat’ or ‘is a person’ selects just one of the many candidates—different ones in the different sharpenings, thus respecting the arbitrariness felt in denying that they all had an equal claim. ‘Tibbles is a cat’ serves as a penumbral connection, guarantying that it is rendered inadmissible any sharpening that selects a different candidate as the referent of ‘Tibbles’ from the one that is selected as belonging to the extension of ‘is a cat’—inasmuch as ‘If it is not red, then it is orange’ serves to exclude sharpenings in which borderline rose Fifi is assigned both to the extension of ‘is red’ and to that of ‘is orange.’ Thus the many candidates are indeed equally eligible as referents of ‘Tibbles,’ but it definitely the case that one and just one of them is a cat after all. *Mutatis mutandis*, once again, for me.

Thus the view that the many candidates are all indeed people (or cats), «very substantive and wildly implausible» as it might be, it is by no means «a metaphysical commitment incurred by anyone who accepts the negation of (I_{dr})» (or of ($Tibbles_{dr}$) for that matter), against what Smith contends. On the contrary, any reason one could have for thinking that it is indeed implausible would be a reason for the defender of the view of vagueness as semantic indecision to adopt a so-called «supervaluationist» solution to the problem of the many.

I conclude that, even if ‘I’ is vague and the view of vagueness as semantic indecision is correct, I could be a material composite object all the same.²

References

- Lewis, David 1986: *On the Plurality of Worlds*, Blackwell, Oxford.
- Lewis, David 1993: ‘Many, but Almost One,’ in John Bacon, Keith Campbell, and Lloyd Reinhardt (eds.) *Ontology, Causality and Mind*, Cambridge University Press.

². For discussions on these topics, thanks to Robbie Williams. Research funded by project HUM2004-05609-C02-01 and a GenCat-Fulbright Postdoctoral Fellowship.

-
- López de Sa, Dan 2004: 'Lewis vs Lewis on the Problem of the Many,' unpublished.
- McGee, Vann 1998: 'Kilimanjaro,' *Canadian Journal of Philosophy* 23, 141-63.
- McGee, Vann and Brian McLaughlin 2000: 'The Lessons of the Many,' *Philosophical Topics* 28, 129-51.
- Smith, Donald 2006: 'Vague Singulars, Semantic Indecision, and the Metaphysics of Persons,' *Philosophy and Phenomenological Research*, forthcoming.
- Varzi, Achille 2001: 'Vagueness in Geography,' *Philosophy & Geography* 4, 49-65.
- Weatherson, Brian 2003: 'Many Many Problems,' *Philosophical Quarterly* 53, 481-501.

Dan López de Sa
LOGOS—Grup de Recerca en Lògica, Llenguatge i Cognició
Universitat de Barcelona
Arché—The AHRC
Research Centre for the Philosophy of Logic, Language, Mathematics and Mind
University of St Andrews
dlopezdesa@ub.edu

TRUTHMAKERS FOR NEGATIVE TRUTHS¹

by Yuki Miyoshi

1. Introduction

What are truthmakers for negative truths? This is the central topic of this essay. It seems that propositions are made true by the following things: particulars, properties, relations, particulars having properties, or the relations holding among particulars.² (Let us call the last two things, particulars having properties and particulars being related to one another facts or states of affairs.) This is easy to see in the case of positive propositions, such as <Smith has blond hair>.³ What makes this proposition true, or the truthmaker for this proposition, is Smith's (the particular's) having the property of having blond hair.⁴ However, the search for truthmakers for true negative propositions has proved to be difficult. For instance, take the truths <Unicorns do not exist> and <Smith does not have dark hair>. What could be the truthmakers for these truths?

I shall present and discuss the answers to this problem offered by Bertrand Russell, Raphael Demos, and D. M. Armstrong. Then, I shall present my solution which involves the consideration of the world in which nothing exists. After that, I shall briefly talk about the implications of my solution on truthmaker theory and possibilities. Before going any further, however, I have first to present so-called entailment principle which will often be used in what follows.

2. Entailment Principle

When a proposition <p> entails another proposition <q>, we have the feeling that either of the following two things is going on. The content of <q> is identical to, or a part of, the content of <p>; the proposition <q> expresses the reality which is identical to, or a proper part of, the reality that the proposition <p> expresses. For example, <Smith is blond and Smith is 6 feet tall> entails that <Smith is blond> because the content of the latter proposition is merely a part of the content of the former. We think that the proposition <The fork is to the

¹. This title was suggested by Dr. Bernard Linsky.

². Particulars, properties, relations by themselves can be truthmakers. For example, a particular, Smith, can be the truthmaker for the proposition that Smith exists.

³. I will use the symbols '<' and '>' to mean whatever is enclosed by them is a proposition.

⁴. Although it is true that other things, such as the whole world, can also be truthmakers for <Smith has blond hair>, I say, 'The truthmaker for <Smith has blond hair> is Smith's having blond hair'; for in this paper, I am mostly concerned with the minimal truthmakers.

right of the spoon> and the proposition <The spoon is to the left of the fork> have different contents, but we think that one entails the other because these two propositions express the same reality.

It seems, then, that when <p> entails <q>, the truthmaker for <p> can be a truthmaker for <q>. Firstly, if the contents of <p> and <q> are exactly the same, what makes one true should also make the other proposition true. Secondly, if the content of <q> is only a part of the content of <p>, then whatever makes <p> true should also make <q> true. This is clear from the example used in the above. If some truthmaker makes <Smith is blond and Smith is 6 foot tall> true, then the same truthmaker should also make <Smith is blond> true. Thirdly, it seems trivial to say that <p> and <q> have the same truthmaker if they both express the same reality in different ways. For instance, <The fork is to the right of the spoon> and <The spoon is to the left of the fork> have the same truthmaker. Finally, if <q> expresses only a part of the reality that <p> expresses, it seems that the truthmaker for <p> can also be a truthmaker for <q>. Suppose there is a reality such that three objects, a knife, a fork and a spoon, are lined up from right to left in that order. Then, this reality is the truthmaker for the proposition <The knife is to the right of the fork and the fork is to the right of the spoon> as well as the proposition <The spoon is to the left of the knife>.

In the above we seem to have justified the entailment principle, which states: the truthmaker, T, which makes some proposition <p> true, also makes another proposition <q> true if <p> entails <q>. Following Armstrong's symbolization more or less, the principle can be represented as:

1. $T \rightarrow \langle p \rangle$
2. $\langle p \rangle$ entails $\langle q \rangle$
3. $T \rightarrow \langle q \rangle$ ^{5,6}

3. Negative Facts

Russell thought that negative truths are made true by the special type of facts, negative facts, but he did not investigate the structure of these special facts (1985, sec. III). So, we shall take up where Russell left off and try to come up with the logical form of negative facts here. In doing this, we will assume two things: that the only elements that constitute the world are particulars, properties, relations, and facts; that negative facts are distinct from positive facts which make positive truths true.

We might think that what makes negative facts distinct from positive facts is that they have negative particulars in their constituents, where negative particulars are the counterparts of regular particulars such that the presence of them in the facts make negative propositions true. So, the logical form of negative facts is negative particulars having positive properties or being related to one another by positive relations. To illustrate, suppose that <Smith does not have dark hair> is true. The truthmaker for this negative truth would be the fact that the negative Smith has dark hair.

Alternatively, we might try to introduce into our ontology negative properties and relations

⁵. See Armstrong 2004, 10.

⁶. I use ' $T \rightarrow \langle p \rangle$ ' to mean T is a truthmaker for <p>.

(e.g., being not-red and being not-left-of). By doing so, the logical form of negative facts would be particulars having negative properties or being related by negative relations. So, the truthmaker for the negative truth that Smith does not have dark hair would be the fact that Smith's hair has the property of being not-dark.

Objections to these two suggestions for the logical form of negative facts are as follows. First of all, postulating negative particulars or negative properties and relations would be ontologically uneconomical. Assuming that we ought to abide by Occam's razor, we should try to find another way to explain negative truths. Secondly, the employment of negative particulars or properties (and relations) seems like an ad-hoc solution; the only motivation that we have to postulate them is to explain negative truths.⁷ Thirdly, both of the above suggestions for the logical form of negative facts cannot handle negative existential truths. For example, what would be the truthmaker for <Unicorns do not exist>? Employing the first approach to negative facts, we seem to need yet another type of particulars, non-existing particulars. This seems like a bad idea with regard to ontological economy, not to mention the strangeness of the notion of existing non-existing particulars. The second approach would be even worse off, under which view, the truthmaker for this negative truth would be each unicorn's having the property of being not-existing. Obviously, this is problematic. Unicorns do not exist; they cannot have any property, let alone the property of being not-existing.⁸

After seeing the ontological cost incurred by introducing negative particulars or negative properties (and relations), we might try to economize by letting negative facts have a special relation, the *not*-relation, where the *not*-relation is a relation such that facts including this relation serve to be truthmakers for negative truths. Using the *not*-relation, we can think of negative facts as having the logical form, the *not*-relation relating positive particulars and positive properties (or relations). So, the truthmaker for the negative truth that Smith does not have dark hair is the fact of the *not*-relation relating Smith and the property of having dark hair.

This solution seems to be able to handle negative existential truths, for which the previous two suggested forms of negative facts are inadequate. We can say that the truthmaker for the negative truth that unicorns do not exist is the negative fact, the *not*-relation's holding between horsy things and the property of having a horn because this negative fact makes it true that horsy things do not have a horn, and this latter proposition seems to preclude the possibility of the existence of unicorns; thus, using the entailment principle, we can say that the truthmaker for <Unicorns do not exist> is the *not*-relation's holding between horsy things and the property of having a horn.

Nevertheless, the *not*-relation seems to be another artificial invention to explain negative truths, and for this reason, we should not use it to solve the problem of truthmakers for negative truths.

At this point, I am running out of suggestions for the logical form of negative facts. So, it is time for us to leave negative facts behind and move onto the next solution for the

⁷. Armstrong discusses these two objections. He further points out that the ontologically problematic nature of negative properties is highlighted by the fact that negative properties do not play any causal role (2004, 55).

⁸. Besides these problems, Dr. Linsky pointed out to me that negative particulars have this problem: if it is true that Smith is neither red nor green, negative Smith is both red and green; however, nothing is red and green at the same time.

problem of negative truths, the incompatibility solution.

4. The Incompatibility Solutions

4.1. Demos's Incompatibilism

Raphael Demos's solution to the problem of truthmakers for negative truths, known as the incompatibility solution, is, very roughly, that negative propositions are just another way of expressing certain positive propositions, and since positive propositions have their own truthmakers, these truthmakers are the truthmakers for the negative propositions (Demos 1917, 188-196). Somewhat more precisely, a negative proposition $\langle \sim p \rangle$ means $\langle \text{Some proposition } \langle q \rangle, \text{ which is incompatible with } \langle p \rangle, \text{ is true} \rangle$, where $\langle \sim p \rangle$ is a description of $\langle q \rangle$ in the sense that $\langle \sim p \rangle$ does not directly refer to any particular state of affairs. We may describe $\langle q \rangle$ by $\langle \sim p \rangle$ because $\langle q \rangle$ is related to a number of propositions by the incompatibility relation and one of the propositions that $\langle q \rangle$ is so related is $\langle p \rangle$. Furthermore, since $\langle \sim p \rangle$ is another way of saying $\langle q \rangle$, the truthmaker for $\langle \sim p \rangle$ is the truthmaker for $\langle q \rangle$.

For instance, take the negative proposition, $\langle \text{Smith does not have dark hair} \rangle$. This proposition means $\langle \text{Some proposition, which is incompatible with } \langle \text{Smith has dark hair} \rangle, \text{ is true} \rangle$, and this proposition is a description of the proposition $\langle \text{Smith has blond hair} \rangle$. When we say $\langle \text{Smith does not have dark hair} \rangle$, what is essentially happening is that instead of directly asserting the proposition $\langle \text{Smith has blond hair} \rangle$, we choose to describe it using the circumstance that $\langle \text{Smith has blond hair} \rangle$ is incompatible with $\langle \text{Smith has dark hair} \rangle$. So, we might say $\langle \text{Some proposition, which is incompatible with } \langle \text{Smith has dark hair} \rangle, \text{ is true} \rangle$, but as this proposition is rather long, we abbreviate it by saying $\langle \text{Smith does not have dark hair} \rangle$. In this sense, the proposition $\langle \text{Smith does not have dark hair} \rangle$ is merely a description of $\langle \text{Smith has blond hair} \rangle$, and the truthmaker that makes $\langle \text{Smith has blond hair} \rangle$ true, namely, Smith's having blond hair, also makes $\langle \text{Smith does not have dark hair} \rangle$ true; after all, the description is merely a different way of expressing the same state of affairs.

We may explain why the truthmaker for $\langle \text{Smith has blond hair} \rangle$ is also the truthmaker for $\langle \text{Smith does not have dark hair} \rangle$ by using the entailment. $\langle \text{Smith has blond hair} \rangle$ has Smith's having blond hair as its truthmaker. $\langle \text{Smith has blond hair} \rangle$ implies $\langle \text{Some proposition, which is incompatible with } \langle \text{Smith has dark hair} \rangle, \text{ is true} \rangle$. Furthermore, $\langle \text{Some proposition, which is incompatible with } \langle \text{Smith has dark hair} \rangle, \text{ is true} \rangle$ means and hence implies $\langle \text{Smith does not have dark hair} \rangle$. By the entailment principle, the fact of Smith's having blond hair makes it true that $\langle \text{Smith does not have dark hair} \rangle$. We may write the argument in the following scheme.

1. $T \rightarrow \langle \text{Smith has blond hair} \rangle$.
2. $\langle \text{Smith has blond hair} \rangle$ implies $\langle \text{Some proposition, which is incompatible with } \langle \text{Smith has dark hair} \rangle, \text{ is true} \rangle$.
3. $\langle \text{Some proposition, which is incompatible with } \langle \text{Smith has dark hair} \rangle, \text{ is true} \rangle$ means, and hence implies $\langle \text{Smith does not have dark hair} \rangle$.
4. $T \rightarrow \langle \text{Smith does not have dark hair} \rangle$.

The strength of this solution is its intuitive appeal. We know that Smith does not have dark hair by looking at Smith's having blond hair. Then, it seems natural to say that it is Smith's having blond hair that is the truthmaker for the negative truth that Smith does not have dark hair. However, this solution is not objection-proof either.

4.2. A Possible Objection from the Negative Facts Camp

Here is a possible objection from the negative fact camp.⁹ The proposition $\langle p \text{ and } q \rangle$ is true when the two constituent propositions, $\langle p \rangle$ and $\langle q \rangle$, are both true. Further, the condition for the truth of these two propositions is that they each have their respective truthmakers (i.e., T1 for $\langle p \rangle$ and T2 for $\langle q \rangle$). Now, consider, $\langle \sim p \rangle$. $\langle \sim p \rangle$ is true when $\langle \text{Some proposition } q \text{ which is incompatible with } p \rangle$ is true, as the latter is the meaning of the former. $\langle \text{Some proposition } q \text{, which is incompatible with } p \text{, is true} \rangle$ can be symbolized as $\langle q \text{ and } p|q \rangle$.¹⁰ Regarding $\langle q \text{ and } p|q \rangle$, we can say the following things: $\langle q \text{ and } p|q \rangle$ is true when $\langle q \rangle$ and $\langle p|q \rangle$ are both true, the condition for $\langle q \rangle$ and $\langle p|q \rangle$ being both true is that $\langle q \rangle$ is true and $\langle p \rangle$ is false, and what makes $\langle q \rangle$ true and $\langle p \rangle$ false is the truthmaker for $\langle q \rangle$ and the falsemaker for $\langle p \rangle$. Since the falsemaker for $\langle p \rangle$ is the truthmaker for $\langle \sim p \rangle$, the truthmaker for $\langle \text{Some } q \text{, which is incompatible with } p \text{, is true} \rangle$ is actually the mereological sum of the truthmaker for $\langle q \rangle$ and the truthmaker for $\langle \sim p \rangle$. It follows that the truthmaker for $\langle q \rangle$ by itself is not sufficient to make $\langle \text{Some } q \text{, which is incompatible with } p \text{, is true} \rangle$ true. Substituting $\langle p \rangle$ and $\langle q \rangle$ with $\langle \text{Smith has dark hair} \rangle$ and $\langle \text{Smith has blond hair} \rangle$ respectively, we have to say that the truthmaker for $\langle \text{Smith has blond hair} \rangle$ by itself is not a truthmaker for $\langle \text{Some proposition, which is incompatible with } \langle \text{Smith has dark hair} \rangle \text{, is true} \rangle$. Reminding ourselves that this last proposition is equivalent to $\langle \text{Smith does not have dark hair} \rangle$, we have to say that Smith's having blond hair, which is the truthmaker for $\langle \text{Smith has blond hair} \rangle$, is not a sufficient truthmaker for $\langle \text{Smith does not have dark hair} \rangle$.

Those who endorse negative facts as the solution to the problem of negative truths would go on to say the following. As it is explained in the previous paragraph, the truthmaker for $\langle \text{Some } q \text{, which is incompatible with } p \text{, is true} \rangle$ is the mereological sum of the respective truthmakers for $\langle q \rangle$ and for $\langle \sim p \rangle$. However, $\langle \text{Some } q \text{, which is incompatible with } p \text{, is true} \rangle$ means $\langle \sim p \rangle$. It follows that the truthmaker for $\langle \sim p \rangle$ is the mereological sum of the respective truthmakers for $\langle q \rangle$ and for $\langle \sim p \rangle$. This is clearly intolerable. For this reason, the incompatibility solution should be abandoned and the truthmaker for $\langle \sim p \rangle$ should be thought to be simply the negative fact for $\langle \sim p \rangle$.

Our response to this objection on behalf of the incompatibilists is that the objection seems to make an unwarranted assumption. The objection unwarrantedly assumes that the condition for the truth of $\langle q \text{ and } p|q \rangle$ is that $\langle q \rangle$ is true and $\langle p \rangle$ is false, and it concludes that what makes $\langle q \text{ and } p|q \rangle$ true is the sum of the truthmaker for $\langle q \rangle$ and the falsemaker for $\langle p \rangle$ (i.e. negative fact that $\langle \sim p \rangle$). However, the incompatibilists can just say that the condition for $\langle q \text{ and } p|q \rangle$'s being true is that $\langle q \rangle$ is true and $\langle p|q \rangle$ is true, and that the falsehood of $\langle p \rangle$ is a consequence of the truth of $\langle q \rangle$ and $\langle p|q \rangle$. Thus, from the incompatibilists' point of view, what makes $\langle q \text{ and } p|q \rangle$ true is the sum of the truthmaker for $\langle q \rangle$ and the truthmaker for $\langle p|q \rangle$, rather than the sum of the truthmaker for $\langle q \rangle$ and the falsemaker for $\langle p \rangle$.

From what has been said in response to the objection we have considered, however, we

⁹. I came up with this objection by thinking about what Armstrong might have meant when he said that Russell's objection to incompatibilism, which appears in the pages 69-70 of *Russell's Logical Atomism*, is 'that a truth of incompatibility is itself a negative truth' (Armstrong 2004, 60). Contrary to what Armstrong said, my understanding of Russell's objection in these pages is that postulating the facts whose logical form is the incompatibility relation's holding between propositions is no more desirable than postulating negative facts; in fact, it is less desirable given that propositions are not existents and non-existing things cannot be the terms of real relations.

¹⁰. I am using '| ' as the Sheffer stroke symbol.

ought to fine-tune our view of the truthmaker for the negative truth <Smith does not have dark hair>. Initially, we contended that the truthmaker for this negative truth was Smith's having blond hair. That is, we thought that Smith's having blond hair made it true that <Some proposition, which is incompatible with <Smith has blond hair>, is true>. However, we have seen that the truthmaker for <q and p|q> (i.e., <Some proposition, which is incompatible with <Smith has dark hair>, is true>) under the present theory is the sum of the truthmaker for <q> (i.e., <Smith has blond hair>) and the truthmaker for <p|q> (i.e., <<Smith has dark hair> is incompatible with <Smith has blond hair>>). Therefore, we must say that the truthmaker for the above negative truth is the mereological sum of the states of affairs: Smith's having blond hair + the incompatibility holding between <Smith has blond hair> and <Smith has dark hair>.

To summarize, <Smith has blond hair> and <<Smith has blond hair> and <Smith has dark hair> are incompatible with each other> have the respective truthmakers: namely, Smith's having blond hair and the incompatibility relation's holding between <Smith has blond hair> and <Smith has dark hair>. Now, the two propositions, <Smith has blond hair> and <<Smith has blond hair> and <Smith has dark hair> are incompatible with each other>, together imply <Some proposition, which is incompatible with <Smith has dark hair>, is true>, which is the meaning of <Smith does not have dark hair>. Therefore, by the entailment principle, the respective truthmakers for these two propositions together make <Smith does not have dark hair> true. Using the formal scheme, we may represent the revised argument as:

1. T1: Smith's having blond hair \rightarrow <q>: <Smith has blond hair>.
2. T2: The incompatibility holding between <Smith has blond hair> and <Smith has dark hair> \rightarrow <p|q>: <<Smith has blond hair> is incompatible with <Smith has dark hair>>.
3. <q> and <p|q> together imply <Some proposition, which is incompatible with <Smith has dark hair>, is true>.
4. <Some proposition, which is incompatible with <Smith has dark hair>, is true> means and hence implies <Smith does not have dark hair>.
5. T1+T2 \rightarrow <Smith does not have dark hair>.

Some may wonder what evidence we have for saying that the incompatibility relation exists. The evidence is that we cannot entertain the incompatible propositions, say, <Smith has blond hair> and <Smith has dark hair>, to be true at the same time. Because of the influence Immanuel Kant had upon me, I would like to say that the incompatibility among propositions has its ground in the mental structure of the mind. Incidentally, this view on the incompatibility has a nice consequence. It allows us to say that the incompatibility relation's holding among certain propositions supervenes upon the structure of the mind; consequently, the incompatibility relation as well as the facts about certain propositions being incompatible with each other are ontologically cost free.

4.3. Property Incompatibilism

So far, we have discussed Demos's incompatibility solution, which utilizes the idea that the incompatibility holds among propositions. At this point, I would like to introduce another type of incompatibility solution, which may be called property incompatibilism. This solution approaches the problem of negative truths in the same way as Demos's incompatibility solution does. The only difference is that instead of saying that some propositions are incompatible with each other, property incompatibilism states that some properties are

incompatible with each other.

According to property incompatibilism, since the proposition <Smith has blond hair> (which is equivalent to <Smith has the property of having blond hair>) together with the proposition <The property of having blond hair is incompatible with the property of having dark hair> imply that <Some proposition, which is incompatible with <Smith has dark hair>, is true>, by the entailment principle, the truthmaker for the last proposition is the mereological sum of the respective truthmakers of the first two propositions. Furthermore, since <Smith does not have dark hair> means <Some proposition, which is incompatible with <Smith has dark hair>, is true>, the two propositions have the same truthmaker. Hence, the truthmaker for the negative truth <Smith does not have dark hair> is the mereological sum of the respective truthmakers for <Smith has blond hair> and <The property of having blond hair is incompatible with the property of having dark hair>, namely, Smith's having blond hair + the incompatibility relation's holding between the property of having blond hair and the property of having dark hair. Let us represent what has been said by using the formal scheme.

1. T1: Smith's having blond hair \rightarrow <q>: <Smith has blond hair>.
2. T2: the incompatibility relation's holding between the property of having blond hair and the property of having dark hair \rightarrow <p|q>: <There is an incompatible relation's holding between the property of having blonde hair and the property of having dark hair>.
3. <q> and <p|q> together imply <Some proposition, which is incompatible with <Smith has dark hair> is true>.
4. <Some proposition, which is incompatible with <Smith has dark hair>, is true> means and hence implies <Smith does not have dark hair>.
5. T1+T2 \rightarrow <Smith does not have dark hair>.

4.4. Remarks on the Incompatibility Solutions

Prima facie, it seems that incompatibilism in general introduces an extra relation (i.e. the incompatibility relation) into our ontology, and if there is a theory for truthmakers for negative truths that achieves better ontological efficiency, by Occam's razor, the incompatibility solutions should be given up. However, this is, in fact, not the case.

As we mentioned earlier, in the case of Demos's incompatibilism, the incompatibility relation among propositions is not an ontological addition to the world if we think that the incompatibility among propositions have its basis in the structure of our mind. As for property incompatibilism, since it seems that the incompatibility relation is an internal relation holding among properties, this incompatibility relation (among properties) does not count as an ontological addition to the world.

What has been said here seems to suggest that incompatibilism is a better solution than the negative facts solution. For while the incompatibility solutions are ontologically cost free, the negative facts solution has to postulate some extra entity into our ontology; for it seems that the negative facts solution has to rely on some property or relation that makes negative facts negative facts (e.g., the *not*-relation).

4.4. A Difficulty with the Incompatibility Solutions

There is a proposition which it seems the incompatibilists cannot handle. It is true that some region of space is empty. Let us call such a region *R*. Then, the proposition <There is nothing in *R*> is true. Using Demos's incompatibilism, it is not clear how this proposition can be dealt with. We might try to say that this proposition is an abbreviation for <Some proposition, which is incompatible with <There is something in *R*>, is true>. Yet, the proposition that is incompatible with <There is something in *R*> seems to be the original proposition <There is nothing in *R*>. Thus, Demos's analysis of negative proposition cannot give the analysis of this proposition which yields a positive truth whose truthmaker is readily available.

Property incompatibilism does not fare well in finding the truthmaker for <There is nothing in *R*> either. The difficulty seems to arise from the circumstance that we cannot tell what property is involved in this proposition. If we cannot find the property involved in this proposition, then property incompatibilism, which requires the properties that are incompatible with each other in order to find truthmakers, is quite powerless.

5. Armstrong's Solution

5.1. General States of Affairs As Truthmakers for Negative Truths

The solution offered by D. M. Armstrong is to use general facts as truthmakers for negative truths (2004, 54-60). General facts can be used for truthmakers for negative truths because general truths imply negative truths. Suppose that a certain particular *a* has the properties, *F* and *G*, only. Then, the proposition <*a* is not *H*> is true, and this is true because in the list of the all the properties of *a*, *H* does not show up. Put differently, <*F* and *G* are the only properties of *a*> implies <*H* is not a property of *a*>. Thus, by the entailment principle, the truthmaker for the general truth <*F* and *G* are the only properties of *a*> is the truthmaker for the negative truth <*a* is not *H*>. For Armstrong, it turns out that the search for truthmakers for negative truths can be completed by finding truthmakers for general truths.

Now, what is the truthmaker for a general truth of the form, <*X*, *Y*, *Z*, etc. are the only *F*'s>? It has been said that each member of *X*, *Y*, *Z*, etc.'s being *F* cannot be the truthmaker for this general truth. This seems to be correct if we consider the circumstance that the fact of Smith's having blond hair does not make it the case that Smith is the only one that has blond hair. As a result, Russell proposed that truthmakers for general truths are general facts; however, Russell did not know the logical form of general facts (1985, 104). Following Russell, Armstrong also thinks that there are such things as general states of affairs, and he proposed the logical form of general states of affairs to be the totality relation's holding between the mereological sum of the states of affairs consisting of *X*'s being *F*, *Y*'s being *F*, *Z*'s being *F*, etc., and the property of being *F* (2004, 54, 72-75). So, in the case of the example mentioned above, the truthmaker for the general truth that *F* and *G* are the only properties of *a* is the state of affairs, the totality relation's holding between the mereological sum of *Fa* + *Ga* and the property of being *a*'s property. (And this state of affair is the truthmaker for the negative truth that <*a* is not *H*>.)

Let me illustrate what has been said using our old friend, blond-haired Smith. The truthmaker for the negative truth that Smith does not have dark hair is the totality relation holding between the sum of the states of affairs involving Smith and the property of being a state of affairs involving Smith. This totality makes it true that a certain list of properties is

all the properties of Smith, which implies the truth that the property of having dark hair is not in that list. That is, it implies that Smith does not have dark hair. Thus, by using the entailment principle, we can conclude that the totality relation's holding between the sum of the states of affairs involving Smith and the property of being a state of affairs involving Smith is the truthmaker for the negative truth that Smith does not have dark hair.

5.2. Remarks about Armstrong's Solution

It may be said that Armstrong's position runs into a trouble with the idea that there are no necessary connections among distinct states of affairs. To see why, consider the following case. As of 2005, Jean Paul Sartre and Le Duc Tho are the only people who have declined the Nobel prize. The truthmaker for the truth <Sartre and Tho are the only people who declined the Nobel prize> is the state of affairs, the totality relation's holding between the mereological sum of Sartre's declining the Nobel prize + Tho's declining the Nobel prize and the property of being the people who have declined the Nobel prize. As an abbreviation, we might symbolize this as Tot (S + Th, D). Now, John F. Nash, Jr. accepted the Nobel prize for economics so that there is a state of affairs, Nash's accepting the Nobel prize. This state of affairs seems to have some connection with Tot (S + Th, D); for the existence of Tot (S + Th, D) makes it somewhat necessary that Nash accepted the Nobel prize. I say it is only somewhat necessary because it is necessary given that there is another state of affairs that Nash was offered the Nobel prize; for if Nash had never existed or if he had not studied economics, then Tot (S + Th, D) would have existed without Nash's accepting the Nobel prize. Still, among the three states of affairs, Nash's being offered the Nobel prize, Nash's accepting the Nobel prize, and Tot (S + Th, D), a necessary connection seems to exist. This may be an unwelcome consequence of Armstrong's position, if one holds that there is no necessary connection among distinct states of affairs.

Perhaps, necessary connections among states of affairs are not all that bad. Some people think that there is a necessary connection between a cause and its effect.¹¹ At any rate, however, Tot (S + Th, D) is not a fact which is needed in our ontology. The proposition that Sartre and Tho are the only people who declined the Nobel prize has another truthmaker, the mereological sum of the following three states of affairs: (i) Sartre's declining the Nobel prize, (ii) Tho's declining the Nobel prize, and (iii) the totality relation's holding between the sum of the states of affairs which involve everyone minus Sartre and Tho and the property of being the states of affairs of everyone minus Sartre and Tho. For <Sartre declined his Nobel prize>, <Tho declined his Nobel prize>, and <Everyone else does not decline the Nobel prize> together amounts to <Sartre and Tho are the only people who have declined the Nobel prize>. Let us abbreviate this sum as S + Th + Tot (E, SofE). Now, Tot (S + Th, D), if it exists, seems to be a consequence of and supervene upon S + Th + Tot (E, SofE). Thus, for the purpose of coming up with the truthmaker for the truth that Sartre and Tho are the only people who have declined the Nobel prize, Tot (S + Th, D) is not needed. So, it seems wise not to postulate this state of affairs; for then we can avoid having the necessary connection between the distinct states of affairs.

Two things can be said against Armstrong's solution. First, the totality relation is quite artificial. Second, since my solution to the problem of negative truths, to which we now turn, can deal with negative truths as well as general truths without postulating the totality relation,

¹¹. Armstrong, in making the same point, cites C. B. Martin and George Molnar as examples (2004, 71).

my solution should be preferred over Armstrong's.

6. The World In Which Nothing Exists

6.1. My Hypotheses about Propositions

In order to talk about my solution for the problem for negative truths, I have to first present my hypothesis about propositions. My claim is that all propositions can be understood in terms of the existence of something. To begin with, all positive propositions are equivalent, in meaning, to propositions of the form <X exists, Y exists, Z exists, etc.>. For instance, the proposition <Smith has blond hair> is equivalent to <Smith exists, the property of having blond hair exists, and the state of affairs of Smith's having blond hair exists>.

Assuming that all positive propositions can be understood in terms of the existence of something, 'not' denies the existence of at least one thing, and negative propositions have different meanings depending upon which existence 'not' denies. For instance, 'Not' in the proposition <Smith does not have dark hair> denies either the existence of Smith or the property of having dark hair or the state of affairs of Smith's having dark hair or all of the above. This seems clear if we consider the use of the negative proposition <An invisible man is not standing behind Smith>.¹² Some people would assert this proposition because they think that there is no such thing as an invisible man. For them, 'not' is denying the existence of an invisible man and hence the state of affairs of an invisible man standing behind Smith. That is, the proposition is equivalent to <There is no invisible man, there is Smith, there is the relation of X standing behind Y, and there is no state of affairs of an invisible man's standing behind Smith>. (Notice here, the negative proposition implies the existence of Smith and of the standing-behind relation.) However, people who believe in invisible men might still assert <An invisible man is not standing behind Smith> if they think that an invisible man is not standing right behind Smith. In this case, 'not' is denying the existence of the state of affairs, an invisible man's standing behind Smith. So, <An invisible man is not standing behind Smith> is equivalent to <There is an invisible man, there is Smith, there is the relation of X standing behind Y, but there is no state of affairs of an invisible man's standing behind Smith>. As we can see from these examples, 'not' denies the existence of at least one thing. Moreover, the consideration of the above proposition makes it clear that negative propositions, though superficially having the same appearance, have more than one meaning, and that the meaning depends on the scope of 'not.'

I must explain what I mean by «the scope of 'not'». I say that the scope of 'not' is the entire proposition, when 'not' modifies the existence of all things that the proposition implies. If the scope of 'not' for <Smith does not have dark hair> is the entire proposition, then the proposition is equivalent to <There is no Smith, no property of having dark hair, no state of affairs of Smith's having dark hair>. Accordingly, «the scope of 'not' is partial» should be understood to mean that 'not' does not modify the existence of all things that the proposition implies. Hence, <Smith exists, the property of having dark hair exists, but the fact of Smith's having dark hair does not exist> is an example of the propositions which are equivalent to <Smith does not have dark hair> in which 'not' modifies only a part of the proposition.

¹². This argument here is basically taken from Demos, although he used the argument to reach a different conclusion (Demos 1917, 190).

6.2. The Wiwne Solution

Now that we have gone through my contention about propositions, we are ready for my thesis. No truthmaker is required for negative truths if the scope of ‘not’ is the entire proposition, but if the scope of ‘not’ is only a part of the proposition, their truthmakers are entities that these negative propositions imply exist. To see why, imagine the world in which nothing exists. I mean absolutely nothing, no particulars, no properties, no relations, and no facts. Call this world Wiwne, for short. About this world, we can truthfully say that there is nothing. Hence, the proposition about Wiwne that <Nothing exists> is true, and it is true despite the reality that there is nothing that makes this proposition true¹³. This consideration seems to show, at least, that for the truth of the proposition that <Nothing exists>, no truthmaker is needed. But about Wiwne, we can similarly assert such propositions as <No human beings exist>, <Unicorns do not exist>, <The earth does not exist>, etc., and those propositions seem to be true, again, despite the lack of things that make these propositions true. For this reason, we may conclude that a truthmaker is not required for negative existential propositions when we make an assertion about the world in which nothing exists.

We can further make the following observations about Wiwne: All positive propositions about Wiwne are false though no entity is making them false; the negative propositions about Wiwne are true if the scope of ‘not’ is the entire proposition, and their truths do not require any truthmaker; when the scope of ‘not’ is only a part of negative propositions, those negative propositions are false despite the reality that no entity is making them false.

All positive propositions about Wiwne are false. As I have said earlier, all positive propositions are equivalent in meaning to the propositions of the form <There is X, there is Y, there is Z, etc.>. The propositions of such a form about Wiwne are false, because nothing exists in Wiwne. Further, it is clear that no truthmaker is required to make those propositions false. Thus, all positive propositions about Wiwne are false and no entity makes them false.

All negative propositions about Wiwne with the scope of ‘not’ the entire proposition are true. This is because negative propositions, if the scope of ‘not’ is the entire proposition, are equivalent to propositions of the form <There is no X, there is no Y, there is no Z, etc.>. Since nothing exists in Wiwne, the propositions of this form are true, and they are true without any truthmaker. Thus, given that the scope of ‘not’ is the entire proposition, negative propositions about Wiwne are true, although there are no truthmakers for these truths.

Finally, if the scope of ‘not’ is only a part of the proposition, the negative propositions about Wiwne are false; for in that case, negative propositions imply the existence of at least one thing. When the scope of ‘not’ is partial, ‘not’ does not deny at least one existent. So, each of the propositions contains at least one ‘there is X’, because of which the entire proposition implies the existence of at least one thing. Such a proposition is false without a falsemaker because the proposition implies the existence of something in Wiwne.

In order to further elucidate what has been said, an example is in order. Consider the negative proposition about Wiwne that <Smith does not have dark hair> in which ‘not’ modifies the entire proposition. This proposition is equivalent to <There is no Smith, there is no property of having dark hair, and there is no state of affairs of Smith’s having dark hair>. This proposition is true, and it is true without a truthmaker because there is nothing in Wiwne

¹³. Armstrong reports that Bruin Christensen informed him of this insight. See Armstrong’s book (2004, 91).

to serve as its truthmaker. Intuitively, we can treat any negative proposition about Wiwne with the scope of ‘not’ being the entire proposition in a similar manner. Thus, we can conclude that all negative propositions about Wiwne are true without the need of a truthmaker as long as the scope of ‘not’ is the entire proposition.

Suppose that the scope of ‘not’ is only a part of <Smith does not have dark hair>. Then, the proposition must imply the existence of at least one thing. Hence, for example, the proposition is equivalent to <There is Smith, there is no property of having dark hair, and there is no state of affairs of Smith’s having dark hair>. This proposition is false because there is no Smith in Wiwne. Further, since there is nothing in Wiwne, nothing is making it false.

At this point, imagine adding one thing to Wiwne. It seems, then, some of the propositions concerning the newly added entity are true, and the negations of those propositions are false. For instance, if we add Smith to Wiwne, the proposition <Smith exists> is true, and the negative proposition <Smith does not exist> is false. Furthermore, the propositions concerning Smith’s internal properties are true and the negations of those propositions are false. For instance, the proposition that <Smith has blond hair> becomes true, and <Smith does not have blond hair> becomes false. (Notice, here, the scope of ‘not’ does not really matter. <Smith does not have blond hair> is false whether the scope of ‘not’ is the entire proposition or only a part of it.) As for the propositions concerning both the newly added entity and the entities that do not exist, the truth-value of those propositions may be true or false, depending on the scope of ‘not’ in the proposition. For example, the proposition <Smith does not have dark hair> is false if the scope of ‘not’ is the entire proposition. If the scope of ‘not’ is the entire proposition <Smith does not have dark hair> is equivalent to <There is no Smith, there is no property of having dark hair, and there is no Smith’s having dark hair>. This proposition is false because there is Smith in Wiwne at this point. However, if ‘not’ denies only the existence of the property of having dark hair and the state of affairs of Smith’s having dark hair, then the proposition is true; for in that case, <Smith does not have dark hair> is equivalent to <There is Smith, there is no property of having dark hair, and there is no Smith’s having dark hair>.

As we have seen above, for some propositions about Wiwne, the change in the truth-values occurs due to the newly added entity, Smith. However, for other propositions, the truth-values are not affected. The truth-values of the propositions that do not concern the newly added entity, Smith, are unchanged. For instance, <Jones wears a hat> is still false because in Wiwne, there is still no Jones, no hat, no wearing relation, and no state of affairs of Jones’s wearing a hat. Similarly, consider the negative proposition <Jones does not wear a hat>. The proposition is still true if ‘not’ modifies the entire proposition, again, because there is no Jones, no hat, no relation of wearing, and no Jones’s wearing a hat. If the scope of ‘not’ is taken to be a part of the proposition, then the proposition is still false because the proposition imply the existence of at least one thing, may it be Jones, a hat, the wearing relation, or Jones’ wearing a hat, while nothing except for Smith (and things concerning him) exists in Wiwne at this point. So, the introduction of Smith does not change the truth-value of the propositions that are not concerned with Smith and his internal properties.

Notice that those propositions whose truth-values are unchanged have the same truth-values by virtue of the same condition as before the introduction of Smith to Wiwne. Some of those propositions are true because the things that those propositions deny to exist do not exist, and the others are false because the things that those propositions imply exist do not exist. This means that the negative truths about Wiwne with only Smith in it are true without

truthmakers, granted that their truth-values are unchanged with the introduction of Smith to Wiwne. Also notice that these true negative propositions have this characteristic: the scope of 'not' is the entire proposition. Thus, it seems we can say that the negative truths about Wiwne with only Smith in it do not require truthmakers if the scope of 'not' is the entire proposition.

As for the true negative propositions whose truth-values have changed from false to true since the addition of Smith, we can make the following two observations. First, their truthmakers are the entities that the propositions imply exist. (In particular, those are Smith or Smith's internal properties.) Second, the scope of 'not' is only a part of the proposition. As we have seen before, <Smith does not have dark hair> has become true with the introduction of Smith to Wiwne, if the meaning of it is taken to be <There is Smith, there is no property of having dark hair, and there is no Smith's having dark hair>. The truthmaker for this proposition is just Smith, and the scope of 'not' is only partial. Likewise, consider <Jones does not have blond hair> taken to mean <There is no Jones, there is the property of having blond hair, and there is no Jones's having blond hair>. This proposition is false at the initial condition of Wiwne. However, with the introduction of Smith to Wiwne, this proposition becomes true; for due to the circumstance that Smith is a blond haired individual, the property of having blond hair exists in Wiwne after the introduction of Smith to it, and hence <Jones does not have blond hair> (taken in the present sense) is made true by Smith's internal property, the property of having blond hair. Also, obviously, in this example, the scope of 'not' in the proposition is only partial. By these examples, we can clearly see that for the negative propositions whose truth-values have changed from false to true by the introduction of Smith to Wiwne, their truthmakers are the entities that the propositions imply exist, and the scope of 'not' is only partial. Further, it seems that we may conclude that if the scope of 'not' is only a part of the proposition, then the truthmakers for the negative truths about Wiwne with only Smith in it are the entities that these truths imply exist.

Now, let us further suppose that we add another person, Jones, to Wiwne. It seems that we can go through a similar argument that we made above once again, and we can arrive at the same conclusion as before. That is, true negative propositions with the scope of 'not' being the entire propositions are true without truthmakers, and truthmakers for true negative propositions in which 'not' modifies only a part of the propositions are the things that those propositions imply exist. In fact, it seems that even if we add as many entities as we wish, we can always come to these two results. This is extremely important; for we can think of our actual world to be the world that we obtain by adding to Wiwne all the entities that exist in the actual world. The implication of this view of our actual world is that even in our world, negative truths are true without the need of truthmakers if the scope of 'not' is the entire proposition, and if the scope of 'not' is only partial, their truthmakers are the entities that the propositions imply exist.

To illustrate the last point, consider the following true negative propositions about the actual world: <Invisible men do not have invisible hair> and <Smith does not have dark hair>. The first proposition can be analyzed into <There is no invisible man, no property of having invisible hair, no state of affairs of invisible men's having invisible hair>. It is clear that no truthmaker is needed for the truth of this proposition and that the scope of 'not' is the entire proposition. Next, <Smith does not have dark hair>, if true, is equivalent to <There is Smith, there is the property of having dark hair, and there is no state of affairs of Smith's having dark hair>. The truthmaker for this proposition is the mereological sum of Smith + the property of having dark hair. We can also see that Smith and the property of having dark hair

are the entities that the proposition implies to exist, and that the scope of ‘not’ in this proposition is only a part of the proposition.

6.3. General Truths¹⁴

The virtue of the above account for negative truths is that it can deal with general truths in a similar fashion without postulating a totality relation. To see how this can be done, suppose that we add Smith to Wiwne, as we did so in the previous section. Then, it seems true that Smith is the only human being that exists in Wiwne, and this is a general truth. Since we only added Smith in Wiwne, the only possible truthmaker for any truth in Wiwne is, either the thin particular, Smith, Smith’s internal properties, or the state of affairs consisting of Smith and some of his internal properties. So, the only plausible candidate for the truthmaker for the general truth that Smith is the only human being is just this state of affairs, Smith’s being human. Here, the totality relation does not seem to be needed.

Next, we add Jones in addition to Smith. Then, it is true that Smith and Jones are the only human beings in Wiwne, and this time, the truthmaker for this general truth is the sum of the states of affairs, Smith’s being human + Jones’s being human. Similarly, we add all the human beings that exist in the actual world to Wiwne, then the truthmaker for the general truth that they are all the human beings in Wiwne is the mereological sum of each of those human beings’ being human. Furthermore, we may add some non-human entities to Wiwne, but the truthmaker for the general truth that all human beings in Wiwne are the only human beings in Wiwne seems to be the same as before. So, even if we add all the non-human entities in the actual world to Wiwne, the general truth that all the human beings in Wiwne are the only human beings in Wiwne is made true by the mereological sum of each human being’s being human. As I have said before, the actual world is the world that we arrive at by adding to Wiwne all the entities that exists in the actual world. Thus, the truthmaker for the general truth about the actual world <*a*, *b*, *c*, etc. are the only human beings in the actual world (where *a*, *b*, *c*, etc. are the names for all the human beings in the world)> is just the mereological sum of each *a*, *b*, *c*, etc.’s being human. In general, contra Russell and Armstrong, the truthmaker for the general truth of the form, <*X*, *Y*, *Z*, and etc. are the only *F*’s> is the mereological sum of each *X*, *Y*, *Z*, and etc’s being *F*.

If this account of general truths is correct, we do not need the totality relation in our ontology. Consequently, in view of ontological economy, the Wiwne account of negative truths should be preferred over Armstrong’s account of negative truths which uses the totality relation.

6.4. Truthmaker Maximalism and Truthmaker Necessitarianism

There are two important implications of my account to the truthmaking theory. First, the thesis that every truth has a truthmaker, called Truthmaker Maximalism, is false. This is because negative truths with the scope of not the entire proposition are true without truthmakers. Second, Truthmaker Necessitarianism, which states that truthmakers necessitate truths, is false. We have concluded that the truthmaker for general truths of the form, <*X*, *Y*, *Z*, etc. are the only *F*’s> is merely the sum of the states of affairs of each *X*, *Y*, *Z*, etc.’s being

¹⁴. I owe a lot to Dr. Linsky for the argument in this section. Dr. Linsky noticed that if we added only Smith to Wiwne then <Smith is the only human being> would be true, while I was trying to convince him that general truths do not require the totality relation by using the entailment principle, and it was at that point that I grasped my argument clearly. Though my memory may not be 100% accurate, that is more or less how I came to that argument.

F. However, the sum of each X, Y, Z, etc.'s being F does not necessitate the truth <X, Y, Z, and etc. are the only Fs>; for there might have been more things that are Fs

6.5. Modalities in Wiwne

Let me briefly talk about modalities in Wiwne. Although nothing exists in Wiwne, intuitively, it seems that something might have existed. So, the modal proposition about Wiwne that something might have existed is true. However, since nothing exists in Wiwne, this truth does not seem to require a truthmaker. Similarly, we can truthfully say that Smith might have existed in Wiwne. Again, the truth of this proposition does not seem to require a truthmaker. Further, consider the proposition <Smith might have had dark hair>. This proposition can be understood as <There might have been Smith, there might have been the property of having dark hair, and there might have been Smith's having dark hair>, and this proposition about Wiwne seems true without a truthmaker. Therefore, the proposition about Wiwne, <Smith might have had dark hair>, is true, and it is true without a truthmaker. These considerations seem to show that modal truths about possibilities are true in Wiwne without truthmakers. Further, this result seems to suggest that possibilities in Wiwne are not dependent upon entities in Wiwne, and this in turn seems to point to the direction that even in the actual world possibilities do not depend upon entities. I am not willing to discuss the truthmakers for necessary propositions here, but I trust that the strategy similar to the one that I have used with respect to possibilities can be used in the case of necessities as well. So, I say that necessities in the actual world do not depend upon the entities in the actual world.

Someone might want to say that the proposition about Wiwne that Smith might have had dark hair is false; for since nothing exists in Wiwne, possibilities do not exist in Wiwne either. But this seems to presuppose that possibilities are thingy (i.e., they are entities of some sort). For my part, I am inclined to say that possibilities are not thingy. Moreover, if we assume that possibilities are thingy and that possibilities do not exist in Wiwne because nothing exists in Wiwne, then we have to conclude that it is necessary that nothing exists in the world, which we call Wiwne. But if it is necessary that nothing exists in that world and if we think that possibilities are thingy, then we have to think that the necessity that nothing exists in Wiwne exists in Wiwne as necessities and possibilities are closely related, and if one is thingy, then the other should also be thingy. However, since Wiwne is the world in which nothing exists, it is a contradiction to think that the necessity exists in Wiwne. Thus, by *reductio ad absurdum*, we should think that possibilities are not thingy.

7. Conclusion

In this essay, I have talked about four types of solutions to the problem of truthmakers for negative truths: the negative facts solution, the incompatibility solution, the general facts solution, and the Wiwne solution. I think that the Wiwne solution should be championed. The Wiwne solution has an upper hand over the negative facts solution and the general facts solution because it does not postulate any entities in order to come up with truthmakers for negative truths, and it is superior to the incompatibility solution because it can handle the negative truth that the incompatibility solution does not seem to be able to. However, accepting the Wiwne solution creates a problem for holding the correspondence theory of truths. By saying that negative truths with the scope of 'not' the entire proposition do not require truthmakers, I am diverging from the correspondence theory of truths. So, it is appropriate for me to offer an alternative theory of truths, but at this point, I have not

developed such a theory of truths yet.^{15, 16}

References

- Armstrong, D. M. 2004. *Truth and Truthmakers*. Cambridge: Cambridge University Press.
- Demos, Raphael. 1917. A discussion of certain types of negative propositions. *Mind* 26: 188-96. <http://www.jstor.org/>.
- Russell, Bertrand. 1972. *Russell's Logical Atomism*. Ed. David Pears. London: Fontana.
- Russell, Bertrand. 1985. *The Philosophy of Logical Atomism*. Ed. David Pears. Chicago and La Salle: Open Court.

Yuki Miyoshi
Department of Philosophy
Simon Fraser University
Burnaby, B.C. Canada
V5A 1S6
yuki_miyoshi524@hotmail.com

¹⁵. I thank D. M. Armstrong for pointing this out to me.

¹⁶. This essay was originally written for my honors essay project at University of Alberta in the Fall semester of 2005 under the supervision of Dr. Bernard Linsky. Naturally, I thank Dr. Linsky the most. I thank Stephen Latta and Dr. D. M. Armstrong for comments and suggestions. Dr. Armstrong was nice enough to read a draft of this essay even though I did not have any previous acquaintance with him. I should also mention that most of the things that I know about truthmakers are introduced to me by Dr. Armstrong's *Truth and Truthmakers*. I thank Dr. Kirstie Laird for her editorial help. Finally, I thank all the people who helped me write this essay one way or another.

REFERENCE, KNOWLEDGE, AND SCEPTICISM ABOUT MEANING

by **Elisabetta Lalumera**

0. Introduction

Scepticism about meaning is the claim that there is no fact of the matter as to whether a term means something or anything else, as there is no privileged semantic relation between a term and what we would call its ‘meaning’ in ordinary parlance. In its most radical version scepticism about meaning leads to Kripke’s sceptic’s paradoxical conclusion that that there is no right or wrong in the use of words. Moderate versions involve the abandonment of the reference relation as the fundamental semantic notion, in favour of truth of sentences or coherence of web of beliefs. Resistance to scepticism about meaning is therefore mandatory for any semantic project that assigns a central role to the reference relation. This paper explores the possibility of resisting meaning scepticism by appealing to the idea that the nature of reference is to maximize knowledge. If the reference relation is a knowledge-maximizing relation, then some candidate referents are privileged among the others — i.e., those referents we are in a position to know about — and a positive reason against meaning scepticism is thus individuated. A knowledge-maximizing principle on the nature of reference was proposed by Williamson in a recent paper (Williamson 2004). According to Williamson, such a principle would count as a defeasible reason for thinking that most of our beliefs tend to be true. My paper reverses Williamson’s dialectic, and argues that (we get a defeasible reason for thinking that) reference is knowledge-maximizing from the premise that most of our beliefs tend to be true. I will therefore defend such premise on different grounds than Williamson’s, and precisely by revisiting a Naturalist argument he rejected, centred on the role of true beliefs in successful action. In the conclusion, an opposition to meaning-scepticism comes out as motivated by the knowledge-maximizing nature of reference, and backed by the plausibility of the claim that beliefs tend to be true.

1. Scepticism about meaning

Let scepticism about meaning be the thesis that

(MS) sentences of the form ‘*t* means *O*’ have no truth value, that is, there is no fact of the matter as to whether *t* means *O*.

Here, *t* is a predicate or a general term (like ‘is green’ or ‘green’), and *O* belongs to the domain of non-linguistic counterparts of predicates (whatever they are, namely, properties for metaphysical realists, or nominalist alternatives to properties). The meaning relation is taken in its broadest sense to be the semantic relation holding between a term and its non-linguistic counterpart (whatever that is). It might be called ‘the reference relation’ — my use of ‘meaning’

signals the fact that MS is not scepticism about reference and acceptance of senses, but rather, most often, rejection of both¹. The above statement of MS is general enough to include mental terms, i.e. concepts, as well as linguistic terms, i.e. words. Arguably, either MS is about both words *and* concepts, or it poses no serious threat². It is not general enough, however, to include proper names and other singular expressions together with predicates. All the traditional arguments for MS typically involve predicates or general terms, while only some of them generalize to singular terms³. Since singular expressions are generally taken to be *not* analogous to predicates and general terms in the semantic sense, arguments for MS about names cannot be obtained *by analogy* from traditional arguments. I will therefore concentrate the discussion on predicate and general terms, and the question whether the conclusion might be generalized to names will be left open.

MS allows for a radical version and a moderate one. The radical version is championed by Kripke's character of the Sceptic in *Wittgenstein on Rules and Private Language* (Kripke 1982). The Sceptic argues for MS and concludes that 'meaning just vanishes into thin air', and there is no right or wrong in using words. The position is so radical that it is unstable, or paradoxical: if the Sceptic's claim is true, then it has no meaning, therefore it cannot be true⁴. The Sceptic's case for MS is simple, at least *prima facie*. Given all the facts about an individual speaker available at a certain time to an omniscient point of view, there are infinite alternative assignments of meaning to an arbitrary term *t* she's been using before that time, that would make 't means O' true. For example, given the evidence allowed, her term 'green' could mean green, but also *grue*, where *grue* is the property of being green or being blue after the time considered. If there are infinite alternative assignments of meaning, and if they are all equivalent given the evidence allowed, there is no unique fact that would make 't means O' true, and the idea of terms having a meaning (one meaning) turns out to be an illusion.

Moderate versions of MS agree that the idea of terms having one meaning turns out to be an illusion, but still concede that complete sentences have truth conditions and truth values.

¹. It might be objected that this formulation of MS begs the question against Fregean accounts of meaning, as it does not mention senses at all. In fact, meaning sceptics tend to reject senses on independent grounds. This is the case for typical meaning-sceptics, i.e. Quinean, Davidsonian and Kripkensteinian-minded philosophers, and this is why I favour the reference-centred formulation here. Apart from this contingent fact, an interesting question would be whether a Fregean theory of meaning (with both reference and senses) is in principle immune to meaning-sceptical arguments, or not. I think it is not. As Kripke (1982) showed, if a putative meaning-relation is a satisfaction relation, i.e. it features a description that the putative referent ought to satisfy, the Sceptic can raise his doubts about the meaning of each of the terms contained in the description, thereby generating a regress. Given that the most plausible reconstruction of sense is *via* a definite description, such a move amounts to an extension of MS to senses. This, obviously, is not to say that there can be no Fregean reply to MS. Rather, the argument proposed in this paper should be taken as an alternative, not as an objection, to a possible Fregean stance on the issue.

². Four views about the relationship between concepts and words are currently debated. According to thought-theorists, concepts are the primary bearers of meaning. For them, MS about linguistic meaning only would be trivial, and MS about concepts would automatically infect language. According to language-theorists, words of everyday languages are the primary bearers of meaning. For them, the above dilemma holds as well, *mutatis mutandis*. For identity-theorists and eliminativists (concepts are words, or there are no concepts but words), MS is again either trivially true, when about concepts only, or infectious, when about words only.

³. See Davidson 1984.

⁴. The position is unstable, but it is not as naïve as being overtly self-contradictory. The Kripkean sceptic raises his challenge about past meanings and past uses, on the assumption that meanings here and now are fixed and shared; then he goes on challenging the grounds of such assumption.

The case for moderate MS is usually made by Quinean or Davidsonian arguments, that is, by arguments starting from the premise that meanings are the outputs of a theory of translation or interpretation, and directed at the conclusion that having meaning is a holistic property⁵. According to this kind of arguments for MS, to say that a term has a certain meaning is to say that a subject's behavior can be systematized by a certain kind of theory, but there is no guarantee that there is only one way of systematizing a subject's behavior. In particular, many different alternative assignments of meaning to individual terms are compatible with the same evaluation of a sentence. In the famous Quinean example, if 'here's a rabbit' is evaluated as true, 'rabbit' could mean rabbit but also rabbit-fly, provided that compensating adjustments are made to the interpretation of the surrounding context. In a similar fashion, Davidson specifies that

If some theory of truth (or translation or interpretation) is satisfactory in the light of all relevant evidence (actual or potential) then any theory that is generated from the first theory by permutation will also be satisfactory in the light of all relative evidence (Davidson 1984, 230).

With the logical device of permutation, endless assignments of meaning are generated from the first one, and they are all equivalent with respect to the evidence. There is thus no reason to pick out one of them and call it 'meaning', no fact of the matter about what individual words refer to, and no substantial sense in which it can be said that terms have a meaning. Radical or paradoxical MS is eschewed here because the *locus* of meaning is allowed to be the theory (or better, the set of theories) as a whole.

Scepticism about meaning cannot be classified among the varieties of epistemological scepticism. MS doesn't claim that we can't get to *know* the meanings of terms, it claims that there are *no facts* about meanings, no truth-grounds for meaning attributions, and no possibility of a supervenience or reduction basis for meanings⁶. Kripke is explicit about the metaphysical character of MS (Kripke 1982, 27). Quine, on the other hand, is explicit about his dismissal of epistemological scepticism in general, given his conviction that we should 'surrender the epistemological burden to psychology' (Quine 1969, 75). It can be objected that both Quinean and Davidsonian theories individuate meanings from a third-person perspective, and they both share the view that a theory of meaning is, basically, a theory of how someone can get to *know* someone else's meanings. Thus — the objection goes — their core insight is basically epistemological. Such theories, however, share the view that there is nothing more to the *nature* of meaning than what can be known about meanings⁷, so that epistemology actually collapses into metaphysics.

Moving backwards in the reconstruction of first-order and second-order reasons, one may

⁵. I'm not addressing here the exegetical question whether Quine and Davidson themselves could have accepted MS, couched in the formulation chosen here. They did surely agree on the dispensability of a strong notion of reference in a theory of meaning, as well as on the dispensability of senses — such being the upshot of an holistic theory. Quine is generally taken to endorse the thesis of the dispensability of senses with his famous rejection of analyticity (Quine 1953).

⁶. Kripke is explicit about the metaphysical character of MS (Kripke 1982, 27). Quine, on the other hand, is explicit about his dismissal of epistemological scepticism in general, given his conviction that we should 'surrender the epistemological burden to psychology' (Quine 1969, 75).

⁷. Given some (behavioural) constraints on the nature of evidence.

ask: Why opposing MS at all? Here, at least three answers are available. Firstly, MS seems to clash with common sense. It is common sense that there is one thing to which «dog» refers on a normal occasion of use. People tend to assume (at least to a certain extent) the determination of meaning when they speak, when they interpret other's speech, and when they learn a new word. So, at least, one needs to be suspicious about a revisionary claim such as MS. Secondly, there are theoretical reasons for opposing MS. Among them, the belief that meaning, formation and attribution of individual concepts have some ontological and explanatory priority *vis-à-vis* the total assemblage of thoughts at which a person arrives. In other words, it can be argued that concepts are prior to thoughts in that they are acquired first both ontogenetically and phylogenetically, and in that all thoughts are conceptually structured, but not all concept-applications, or acts of categorization, involve structured thoughts. A second theoretical reason is the belief that the meaning relation is a somewhat causal relation between a mental item and a world item, which requires that both *relata* exist. Philosophical positions sharing such beliefs are direct reference theories and informational semantics for concepts, realism about intentionality and semantic naturalism. For these positions to be tenable, MS must be false.

2. Meaning and knowledge

Let's put aside second-order reasons, and go back to the main concern, that is, strategies for rejecting MS. Generally, they take the form of constraints, i.e., additional conditions that components in 't means O' must satisfy for 't means O' to have a truth value. Anti-MS constraints divide into two broad categories: constraints on the meaning relation, and constraints on the domain of possible non-linguistic counterparts of terms. Constraints on the second category provide further specifications of what it takes to be a referent, so to rule out the greatest part of arbitrary assignments proposed by the sceptics about meaning. Constraints of this kind are not very popular. They draw on metaphysical assumptions and argue, for example, that genuine non-linguistic counterparts of predicates have to be *natural* properties, and not gerrymandered classes⁸.

Constraints of the first category, on the meaning relation, specify further conditions that a term-object relation must satisfy in order to qualify as a meaning relation. One could say, for example, that genuine meaning relations are also functional relations, that is, they connect the term to what it has been created or preserved for, by the organism or system it belongs to. Again, this move would have the effect of ruling out sceptical alternatives as spurious — typically, meaning relations obtained *via* permutation devices are not likely to correspond to natural functions. Unlike the second category of anti-MS constraints, this one is crowded: there are many other proposals about constraints on the meaning relation, and a complete survey (let alone an assessment) would far exceed the limit of this paper. My aim here is rather to try and evaluate only one of them, the *knowledge-maximization constraint on meaning* (KM from now on). Employed by Williamson in a discussion of the veridicality of so-called intuition, I think KM can play the role of a positive reason against MS⁹. Let's see how.

⁸. See Lewis 1997. More precisely, Lewis does not regard sceptical alternatives as absolutely ineligible for reference, but thinks that some properties (the natural ones) attract reference more than others (less natural ones). I have argued for a solution of this kind against Kripke's sceptic in Lalumera 2005.

⁹. Williamson uses «reference» where I use «meaning».

The core idea of KM can be stated simply as follows: terms mean what we can get knowledge about. Therefore, terms do not mean what we can't get knowledge about. For example, in Putnam's Twin-Earth scenario, Oscar's term 'water' means water on Earth, whereas it means t-water on Twin Earth. The explanation is that on Earth Oscar can get knowledge about water, but he can't get knowledge about t-water, in fact there is no t-water; and viceversa for Twin Earth. Note that the explanation bears on the impossibility of getting to know (much) about the referent, given the alternative assignments of meaning, not on the fact that the alternative assignment of meaning is not present in the environment, as traditional externalism would have it. Non-e. Since, however, there are far more putative referents in the environment than we can know about, the emphasis on knowledge allows for a more fine-grained discrimination¹⁰.

Also, it is important to distinguish between the idea that terms mean what we can get knowledge about, from the idea that terms mean what they mean through the mediation of our knowledge of their referents¹¹. In the first case, the term-meaning relation can be direct, in the second it is not. The second idea can be embodied, for example, by a descriptive theory. On a descriptive theory, 'water' means whatever substance satisfies the conditions expressed by the theory of water we possess. The meaning is fixed by the theory. Instead, the first idea can be embodied by a direct-reference theory. On such a theory, 'water' refers to that substance via causal chain. Only, the relevant causal chain (among many possible ones) is individuated as the knowledge-maximizing chain. On the quite neutral assumption that meaning relations (or reference relations) can be either causal or *not* causal, here's Williamson's statement of the view:

Roughly: a causal connection to an object is a channel for reference to it if and only if it is a channel for the acquisition of knowledge about the object. Often, a causal connection is a channel for both. Equally, a non-causal connection to an object is a channel for reference to it if and only if it is a channel for the acquisition of knowledge about the object. Sometimes, a non-causal connection is a channel for both. (Williamson 2005, 140-141).

For the present purposes, I can leave the notion of knowledge quite underspecified. Very generally, knowledge as a subject-object relation depends on facts about the subject as well as on facts about the object. Among the latter, the fact that the object exist. Among the former, facts about the subject's cognitive system, her inferential practices, her rational behaviour may be enlisted. The balance between the two classes of facts is obviously different for internalist accounts of knowledge and for externalist ones. Such a difference, however, has no bearing at this stage of the discussion. Whatever knowledge is, it may be used as a constraint on the meaning-relation.

More precisely, KM plays the role of a constraint on the *nature* of the meaning-relation. Its role is metaphysical. Though it employs the epistemic notion of knowledge, it is to be considered only secondarily as an epistemological principle about the proper methodology for the ascription of meaning, that is, only insofar as the nature of meaning determines how we

¹⁰. See Williamson's example where there are two putative assignments of meaning, and they both exist in the subject's environment, but only one is related to the subject through a reliable belief formation mechanism (perception) whereas the other is not (divination).

¹¹. This latter stance is the Fregean stance, which comes out as an alternative to the view defended here.

should ascribe it¹². Thus, given our statement of MS as a metaphysical thesis, knowledge-maximization can qualify as an anti-MS constraint. Precisely, it is a constraint of the second category illustrated above, which specify *what it takes* to be a meaning-relation. According to knowledge-maximization, candidate meaning-relations will also have to be knowledge-relations.

But what is KM supposed to do *in practice* against a meaning sceptic? The sceptic would claim that all the possible assignments of meaning to *t* are equivalent, and therefore ‘*t* means *O*’ is never true for a specific *O*. KM then allows one to reply that not all the possible assignments of meaning are equivalent, because some of them presents us with meanings we can know *more* about, and ideally one of them with meanings we know *most* about. This would be apparent in the case of multiple assignments generated by permutation devices. Take a theory of translation that individuates ‘Here comes a rabbit’ as a true sentence, and assigns rabbit as the meaning of ‘rabbit’, and then a second theory obtained by permutation from the same, which assigns the former Japanese prime minister Yasuhiro Nakasone as the meaning of ‘rabbit’. Given KM, we are in a position to say that ‘Here comes a rabbit’ means that here comes a rabbit, and not that the former Japanese prime minister is coming along, and ‘rabbit’ means rabbit, if at present the speaker is not in the position of getting knowledge about Yasuhiro Nakasone *via* perception. KM squares with the a posteriori hypothesis that, given a specific context of use, some aspects and objects in the world are in the cognitive background of speakers and hearers, while others are in the cognitive foreground, and this is a relevant fact about content individuation.

It may be objected that, even granting that KM narrows down the wild explosion of sceptical meaning assignments, it will not suffice for narrowing it down to *one*. That is, KM may help us excluding Yasuhiro Nakasone but not rabbit-flies when «rabbit» is involved, given that by hypothesis rabbit-flies can be known about when rabbits are. Not all putative meaning-relations are knowledge-relations, but there can be more than one for each term. Since, however, knowledge admits gradation whereas existence does not, KM as a constraint on meaning allows one to consider putative referents on a scale, that is, it allows to give weights to alternative choices. For example, presumably, rabbits are better referents than rabbit flies, given that they are more salient for a cognitive system shaped as ours¹³. If putative referents are no more infinite in number, and if they can be assigned different weights, then MS is wrong. Thus, KM is a reason for rejecting MS, though a *defeasible* one, for the assignment of different weights would surely involve genuinely empirical facts and considerations, and not purely a priori ones.

3. Why Knowledge-maximization

If KM can count as a reason against MS, what we need now is a reason for accepting KM. Why KM should be taken as a plausible principle about the nature of meaning? How can the claim that meaning is knowledge-maximizing be supported? In Williamson (2004) KM is the conclusion of an inference to the best explanation. What has to be explained is the fact that beliefs tend to be true or, in other words, Williamson is looking for positive reason for believing that ‘beliefs tend to be true’. Let’s call that thesis ‘Veridicality’ for short.

¹². Williamson 2005, 142.

¹³. See psychological studies on the notion of objecthood in relation to cognition, especially Spelke et al.(1993).

Williamson assesses some candidate explanation of Veridicality, and discharges them as not adequate. Among them, a naturalist argument, Davidson's principle of charity, and its spin-off the principle of humanity. Therefore, he concludes, 'we need to make a new start', which is, the principle of knowledge-maximization (Williamson 2004, 131-139). Thus, in Williamson's paper, KM comes out in absence of other candidates, as the general explanatory principle from which Veridicality is supposed to follow.

My point here is that KM can have the role of a positive reason for thinking that terms have definite meanings — here, KM comes out again as a general explanatory principle from which *that* thesis is supposed to follow. Nevertheless, I think that some independent reason for KM can be provided as well, beside establishing it as a conclusion of an inference to the best explanation. Precisely, I think that Veridicality supports KM, and provides a good (though defeasible) reason for accepting it. I am hereby reversing Williamson's dialectic — or in other words, I'm exploiting the link he individuated between KM and Veridicality, but *contrariwise*. To repeat, according to Williamson KM is a reason for accepting Veridicality, and my point is that that Veridicality can constitute a positive reason for accepting KM. And KM, on its part, is going to have the role of an anti-sceptic constraint. As I'm not going to explain Veridicality *via* KM again (as Williamson does instead), my explanation will not be circular. I need, however, to argue for Veridicality on different grounds. Before turning to that, however, let's see what the truth of most of our beliefs has to do with the idea that the nature of meaning is to maximize knowledge.

To accept that Veridicality holds not by accident, but on a regular and systematic basis, is to accept the idea that there is something about our way of forming beliefs — some general principle about the functioning of our belief-forming mechanisms — that makes it possible. It is also plausible that the general principles that make Veridicality possible are more than one, namely, a whole set including principles about our perceptual systems (optimality conditions for perception), our inferential processes (truth-preserving inferential rules), and about language and thought (knowledge-maximizing meaning assignments)¹⁴. The intuitive idea is that if our representational system didn't work the way they do, our inferential rules were not the ones we take as valid, and the meaning relation was not the one that holds, Veridicality would non be in place¹⁵. Though none of the factors accounts for Veridicality alone, if one is missing it is likely that Veridicality fails. On the other hand, that Veridicality holds *signals* that it is likely that all the general rules of functioning in the set are operating properly. This is how Veridicality can act, indirectly, as a defeasible reason in favour of KM. In other words, on the hypothesis that KM and the other general rules make Veridicality possible, that Veridicality is actual may count as a reason to think that KM and the other general rules of functioning are in place. Schematically,

Veridicality \vdash KM and other principles¹⁶

¹⁴. Why knowledge-maximizing meaning assignments rather than truth-maximizing meaning assignments? Basically, for the same reason expressed in section 3 above, that is, because meaning, formation and attribution of individual concepts have some ontological and explanatory priority *vis-à-vis* the total assemblage of thoughts at which a person arrives.

¹⁵. A similar strategy involving principles about cognition against scepticism is employed in Burge 2003.

¹⁶. Williamson's argument, as I understand it, includes a premise like 'KM $_$ Veridicality'.

Veridicality

KM and other principles

Surely Veridicality comes out merely as a defeasible reason for KM *alone*, for KM is just one of the factors that jointly suffice for Veridicality. Nonetheless, if Veridicality can be accounted for on independent grounds, KM would gain in plausibility.

4. The Naturalist argument for Veridicality

So how is it that Veridicality holds — why is it that most of our beliefs tend to be true? Not *my* beliefs in particular, but our beliefs *collectively* — if it was possible to collect them in a box and inspect them one by one, the true ones would outnumber the false one. Maybe not in any given instant of time — there surely can be times of massive error. But with big numbers, just like the probability that a tossed coin lands head equals the probability that it lands tails, the frequency of true beliefs is higher than the frequency of false ones. You can either believe that it is so or just suppose that it is so — in either case, why would *that* be?

This may seem to be a prohibitively big question. Note, however, that it is not as prohibitive as the task of providing a confutation to the traditional epistemological sceptic, who claims that beliefs, say, about the external world are false or unwarranted. The two tasks can be set quite apart. To individuate a positive reason for Veridicality would *not* be sufficient for traditional anti-sceptic aims, as the sceptic would probably run his sceptical argument on any claim you could present as a positive reason for Veridicality. Here my aim is more modest. What I would like to find is a positive reason for Veridicality that would support the second premise of the argument schema above. It need not be sceptic-proof.

As Williamson (2004) admits, it is very natural to support Veridicality with a ‘Naturalist’ argument that appeals to the idea that true beliefs are more useful than false ones:

(N) True beliefs tend to cause one to get what one wants in a way in which false beliefs do not ... on the whole, truth is more conducive than falsity to survival (Williamson 2004, 131).

Taking (N) as a premise, and developing the evolutionary hint suggested by the word ‘survival’, the Naturalist argument for Veridicality may go on as follows. According to (a simplified version of) the Evolutionary Theory, what is conducive to survival is preserved and enhanced. Therefore, Veridicality is preserved and enhanced. Here, reliable belief-forming mechanisms that account for Veridicality are identified with a phenotypic (heritable) trait, and the fact that our population came to have organisms with that phenotypic trait is explained by saying that our ancestors were selected for possessing that trait.

As it is well-known, evolutionary arguments invite objections when employed in connection with high-order cognitive capacities, such as conceptual thought, or logical reasoning. For example, it is often objected that evolutionary explanations could be plausible only for very simple beliefs about food, mating, etc., and generally only for limited subject matters. Such objections have nonetheless received convincing replies¹⁷. I do not think, however, that the Naturalist argument for Veridicality needs to be an evolutionary argument. The evolutionary

¹⁷. See Schechter (2005) for a defense of an evolutionary explanation of the validity of deductive rules.

reading is not crucial. In fact, the effect of true believing *versus* false believing can be seen within a lifespan perspective. If I am rational, I would preserve and prefer (as far as I can) those belief-forming practices that brought me to success in the past, given my personal experience. Thus, if evolution presupposes experience of the species, one may as well take experience of the individual as primary. This suggestion amounts to replacing ‘survival’ with ‘success’ in premise (N) above, thereby obtaining:

(N’) True beliefs tend to cause one to get what one wants in a way in which false beliefs do not...on the whole, truth is more conducive than falsity to success¹⁸.

The non-evolutionary version of the Naturalist argument would then have that given an agent’s rationality, that agent is likely to preserve and enhance what is conducive to success; therefore, reliable belief-forming mechanisms that account for Veridicality are preserved and enhanced. This version of the Naturalist argument involves the concept of rationality and generalizes over rational agents. It does not mention the Evolutionary theory. It is weaker than the Evolutionary version, though, for individual agents are not in a position to control all their belief-forming mechanisms, but only some of them. We can learn from experience how much to trust testimony, intuition, divination, and other belief-forming practices, and consequently decide to enhance some of them and discard others. We are not, however, in the position of discarding our own perceptual belief-forming mechanism within a lifespan (at best, we can put on glasses). This version, on the other hand, is free from the traditional objections to evolutionary explanations of high-order cognitive capacities.

What I have said so far would suffice for suggesting that what is crucial to the Naturalist argument for Veridicality is not the particular theory (evolution or individual rationality) that fills in premise (2). What is crucial is the link between Veridicality and success or survival, that is, the link between true beliefs and what is good for one. It can be expressed in a general principle about rational action. Here is Williamson’s version:

(RW) If an agent desires that p, and believes that if it does A then p, then *cæteris paribus* it acts so that it believes it does A.

Employing (RW), the role of true beliefs in successful action becomes apparent, as the following derivation shows:

¹⁸. Principle N’, like its ancestor N, should not be confused with the basic idea of success semantics. Success semantics is the view that the content of a belief is individuated (roughly) by the actions it would bring to success (given some constraints). It is a thesis about the individuation of the content of beliefs Principles N’ and N, on the other hand, simply state that true beliefs are more conducive to success than false ones, however their content (and their truth) comes to be individuated. One can hold (per absurdum) that true beliefs are read off from God One’s mind, whereas false beliefs are read off from God Two’s mind. N’ would then say that God One’s givings are more conducive to success than God Two’s ones. Thus, N and N’ are compatible with virtually any theory of content whatsoever, and they are not vulnerable to objections typically raised against success semantics. For success semantics see e.g. Whyte (1990).

p is good for S

S desires that p

S believes that if it does A then p

S acts so that it believes it does A

p

The desired good, namely p , is arrived at only provided that S 's beliefs (that if S does A then p , and that S does A) are true. This is how true beliefs explain successful action. But according to Williamson (2005), this is also how the Naturalist argument, both in the evolutionary and in the non-evolutionary version, falls prey of a fatal objection. In his words:

Unfortunately, such a derivation explains much less than it appears to. For one can show in the same way for infinitely many deviant properties true* and good* that the combination of true* beliefs and desires for what is good* for one yields (*cæteris paribus*) what is good (not just good*) for one (Williamson 2005, 133).

According to him, it is possible to prove that what is good for an agent (the desired p above) is also yielded by the combination of true* beliefs and desires for what is good*, where true* and good* are so defined:

(def1) that p is true* iff that $\wedge p$ is true

(def2) that p is good* iff that $\wedge p$ is good

where $\wedge p$ is an arbitrary structure-preserving mapping on propositions (e.g. it maps 'You are reading' with 'Snow is white'). If Williamson is right, then the Naturalist argument for Veridicality fails, for the supposed link between true beliefs and success (or survival) comes out to be spurious. Let's see his proof in details.

Assumptions include the already mentioned (def1) and (def2), as well as his principle about rationality:

(RW) If an agent desires that p , and believes that if it does A then p , then *cæteris paribus* it acts so that it believes it does A .

The proof goes as follows.

- (1) S desires that p, believes that if it does A then p, and acts so that it believes it does A (RW)
- (2) that if it does A then p is true* (assumption)
- (3) that it does A is true* (assumption)
- (4) \wedge (if it does A then p) is true (2, def1)
- (5) if \wedge (it does A) then \wedge p (4, commutation)
- (6) \wedge (it does A) (3, def1)
- (7) that \wedge p is true (5, 6, m. ponens)
- (8) that p is good* for S (assumption)
- (9) that \wedge p is good for S (8, def2)

Conclusion 9 would show that what is good for an agent is yielded by a combination of arbitrary propositions such as good* and true* ones, thereby leaving no role for *genuine* true beliefs in the rational behaviour of agents. Permutation spoils the intuitive conceptual link between Veridicality and what is good. There is no reason to claim that success or survival have to do with the truth of our beliefs more than with any other arbitrary property of them. Williamson's proof would then amount to a fatal objection to the Naturalist argument. In order to employ the Naturalist argument to support Veridicality, therefore, it needs to be rejected.

5. The Naturalist argument restored

How to block the undesired conclusion that true is no better than true* in goal-directed behaviour? I think the place to look at is Williamson's principle RW stated above. It contains a *reflective specification* — the agent acts so that *it believes* it does A — which plays an important role in the proof or the undesired conclusion, allowing for the insertion of the propositions obtained by permutation. Why this reflective specification? Why not a simpler principle, like the following?

(R) If an agent desires that p, and believes that if it does A then p, then *cæteris paribus* it does A.

This is Williamson's own explanation:

The act A is not something of which the creature has no idea. It conceives A in believing that if it does A then P. If an agent does A without believing itself to be doing so, then the natural link between antecedent and consequent in [R] is broken. For example, if you go north while believing that you are going south, your action is not explained just by your desire to reach the oasis and belief that if you go north then you will reach the oasis, [R] notwithstanding. Perhaps the explanation is that, in addition, you desire even more strongly to avoid your enemy and believe that he is at the oasis. Although such examples do not refute [R], since the *cæteris paribus* clause absorbs their shock, they indicate that the rationale for [R] takes for granted that beliefs about what one is doing tend to be true, which is a special case of the very phenomenon that we are trying to understand (Williamson 2005, 132).

Williamson points to a difficult case for a simple principle like (R). The case is one where S's non-reflective belief (that if S does A then p) is true, whereas her reflective belief (that S does A) is false. What happens to principle (R) when, say, S goes north while believing she is going south, on the hypothesis that the desired good is in the northern direction? Williamson suggests that the hypothesis be dropped in favour of a new one, according to which the *most* desired good is southwards. We would then have that S goes north while

believing that she is going south, and south is where she desires to go. In such a case, S would simply fail to accomplish action A — the action that constitutes the means to reach the good *p*. And it would fail to reach *p* (which is, on the proposed reading, southwards). On this reading, Williamson argues, (R) still holds. But S's performance in the case considered is unsuccessful; we need thereby to isolate cases like that if we want to preserve the link between true beliefs and success. The only way, according to Williamson, is to assume within the *cæteris paribus* clause that reflective beliefs (beliefs about what one is doing) are mostly true. But then (R) would beg the question in an argument for Veridicality.

Williamson's reading of the proposed case, however, is not mandatory. One may follow this alternative line of reasoning. In a case where S goes north while believing she is going south, on the hypothesis that the desired good is in the northern direction, (R) is simply fulfilled. Given that the non-reflective belief is true, S reaches the desired good. The case would only show that reflective beliefs play no role in an explanation of successful action. We can just overlook them. No matter what we think we're doing, if we act on our desires and choose the appropriate means, we get through. In fact, this is what Freudian examples show — they show us the irrelevance of (non-deep) reflective beliefs about one's actions (and desires, etc.) in the explanation of rational behaviour. Consider this other case. S believes she is posting an invitation for T. S is in fact putting the invitation for T in a dustbin. The explanation is that S desires T not to be present at her party, and S believes that if she does not send an invitation to T, T will not be present. S's reflective belief is false, but her non-reflective belief is true, and conducive to success.

This line of reasoning invites an objection. Even granted that there are many cases where we act without knowing what we do, most of the time we're reliable about ourselves. Maybe some desires remain hidden to introspection, but on many others the Ego has nothing to say, nor to censor. So, even the most enthusiastic Freudian would allow that *lapsus linguae* and missed acts are just episodes. To conclude that reflective beliefs about what one is doing are just irrelevant to action is somehow misleading, for it might suggest that agents are always passively driven by their desires, with no kind of cognitive control.

On the other hand, the Freudian case brings out an interesting point. The kind of cognitive control of agent on her actions need not be an explicit reflective belief of the form 'I am doing A now', let alone of the form 'S does A', where S is oneself. Maybe we are reliable about such beliefs when we entertain them, but it may be also true that often we act without being in the position of entertaining them, and not only for Freudian reasons. Sometimes actions are performed very quickly, on top of perceptual inputs — for example, when screening up one's eyes with a hand as a light flashes¹⁹. Some other time, an agent may lack the proper concepts that would compose the relevant belief. The more one conceives of concepts as word-like, the more one can find examples where an agent possesses the resources to behave rationally, but not the language to build up a belief about her actions (think of pre-linguistic children, of chimpanzees, etc.). For all these reasons, a conscious reflective belief seems to be too heavy a requirement for rational action.

It follows that Williamson's principle (RW), with its reflective specification, can be dropped. Dropping (RW), the proof of the conclusion that true* and good* yield what is good does not even start. The Naturalist argument for Veridicality is saved.

¹⁹. It might be objected that in such a case one would need a true belief about the position of one's head. A possible reply could be that such a content need not be a belief, but rather a non-conceptual piece of information.

A problem, however, remains to be fixed. If (RW) is too demanding, Williamson's case of going north while believing one is going south shows that (R) is too loose. It leaves out the intuitions that rational agents do not act merely randomly or passively, but rather they have some kind of cognitive control over what they do. My proposal is that the concept of *intention* would suffice for expressing the required kind of cognitive control. A principle would then have this form:

(RI) If an agent desires that p , and believes that if it does A then p , then *cæteris paribus* it intends to do A, and *cæteris paribus* it does A²⁰.

(RI) establishes a link between success, true beliefs and fulfilled intentions. According to (RI), the agent's success depends on her having a true belief, the non-reflective belief about A being the appropriate means for the desired good. This preserves the link between true beliefs and success or survival. But the agent's success also depends on her forming the corresponding intention of doing A, and of that intention being fulfilled. The first *cæteris paribus* clause in (RI) specifies that rational agents, most of the time, intend to do what they believe is useful. To assume that in the *cæteris paribus* clause is not question-begging, as we are employing (RI) in an explanation of Veridicality, and Veridicality says nothing about the formation of intentions. Finally, according to (RI), the agent's success depends on her intention being fulfilled. The second *cæteris paribus* clause in (RI) is about the fulfilment of intentions. Intentions may not be fulfilled for a variety of reasons. Sometimes the world does not cooperate, and agents get frustrated. Again, in an explanation of Veridicality, in order to show that true beliefs are conducive to success, we may assume within the *cæteris paribus* clause that intentions of acting be fulfilled. This would not amount to begging the question, because Veridicality says nothing about the fulfilment of intentions. As for the case of going north while believing one is going south, it can be redescribed. If the agent goes intends to go north, then principle (RI) holds, and the falsity of the second-order belief is irrelevant. The agent may be deluded about the fulfilment of her own intention, like when one has the impression of getting lost when in fact one is going in the right direction. If, on the contrary, the agent intends to go south, and goes north, this does not constitute a counterexample to (RI), for (RI) assumes that most of the time rational agents do form the appropriate intention of acting — (RI) absorbs the shock.

It might be objected that intentions, just like beliefs, are contentful states, so a permutation argument could be run on my new principle just like on Williamson's. An explanation of why an agent does what's good may be that the agent has true* beliefs and intends to do what's good*. My reply would be that the permutation argument in Williamson's version, relied on Veridicality, while the connection between intending to do something and doing something does not. Therefore, a new permutation argument would have to be thought up for the objection to be fatal²¹.

So far, then, equipped with an intuitive concept of intention we may preserve the idea behind Williamson's objection, as well as avoid the reflective specification that brings to principle (RW).

²⁰. The question about the nature of intentions and intending is not relevant for my purposes, as far as intending to do A is not equal to and doesn't entail believing one is doing A.

²¹. I would like to thank an anonymous reviewer of this journal for this point.

6. Conclusion

The above defence of the Naturalist argument against Williamson's object makes a case for Veridicality, that is, the claim that beliefs tend to be true. That beliefs tend to be true, in its turn, signals that certain general principles about our thought and language are in place; among them, the knowledge-maximizing principle about meaning. Veridicality is thus a defeasible reason in support of the knowledge-maximizing principle — that is what I argued in section 4. Knowledge-maximization, in its turn, may give us a good though defeasible reason against meaning scepticism, the claim that there is no fact of the matter as to whether words and concepts mean. The required fact, according to my hypothesis, is that some putative meanings, and not others, are part of a channel of knowledge that ends up in our minds — to put it simply, terms mean what we can get knowledge about.

A further question to engage in would concern the nature of the knowledge relation which constitutes the meaning relation. According to Williamson, KM makes essential use of an intentional vocabulary, and therefore cannot be employed within naturalist reductive accounts of meaning. Williamson's remark restricts the scope of KM as an anti-sceptic reason. KM would come out as available to non-naturalists only. I do not think, however, that such a conclusion is mandatory. If knowledge is naturalizable, then KM is compatible with a naturalist semantics, and its scope against MS is the widest. In particular, if the knowledge channel were a kind of causal channel, KM would be a generalization of Kripke's (1972) insight about names and general terms getting their reference via a causal channel. The truth of the antecedent of this conditional depends on the feasibility of reliabilist programmes in epistemology. As far as the present discussion is concerned, the question remains open — once established that KM can have a role against meaning-scepticism, it remains to be settled how much it can do.

References

- Burge T.(2003) *Philosophy and Phenomenological Research*, 67, 503-548.
- Kripke S. 1982 *Wittgenstein on Rules and Private Language*, Oxford, Blackwell.
- Davidson D. 1984 *Inquiries into Truth and Interpretation*, Oxford, OUP.
- Fodor J. 1994 *The Elm and the Expert: Mentalese and its Semantics*, MIT Press, Cambridge, Massachusetts.
- Kripke K. 1972 *Naming and Necessity*, Cambridge, Mass., Harvard UP.
- Lalumera, E. 2005 «A Simple Realist Account of the Normativity of Concepts». *Disputatio* 18, 2005, pp.1-17.
- Quine W. 1953 «Two Dogmas of Empiricism», *Philosophical Review*, 1951, 60, pp. 20-43.
- Quine W. 1969 «Ontological Relativity», in *Ontological Relativity and Other Essays*, New York, Columbia University Press, 1969, pp. 26-68.
- Schechter J. 2005 «Can Evolution Explain the Reliability of Our Logical Beliefs?» (Unpublished ms).
- Spelke, E. *et al.* 1993, «Gestalt Relations and Object Perception: A Developmental Study». *Perception*, 22, 1483-1501.
- Williamson T. 2004. «Philosophical 'Intuitions' and Scepticism about Judgement», *Dialectica*

58, 2004, pp. 109-153.

Whyte, J. 1990. «Success Semantics». *Analysis* 50: 149-157.

Elisabetta Lalumera
Università di Milano-Bicocca
Dipartimento di Psicologia
<elisabetta.lalumera@unimib.it>

THE SORITES Charter

· Version 2.2 ·

Copyright © 1995-2006 by the **SORITES** team

PREAMBLE

This Charter is an updated and amended version of the original «Copyright Notice and Legal Disclaimer» appeared in Issue # 01 of *SORITES* (April 1995). It is a covenant between the **SORITES** team, as a publisher, and its contributors and readers, prescribing their respective duties and rights, as well as the terms and conditions under which those rights have to be exercised.

Persons dealing with **SORITES** in either capacity, as contributors or as a readers (or users), bind themselves to abiding by these regulations, as does the **SORITES** team itself.

This Charter aims to safeguard all legitimate interests in a balanced way, so as to ensure maximum access to, and use of, the materials we produce, to serve the scholarly public at large, to enhance the prestige of our contributors and the influence of the journal and its publisher, while preventing dishonourable behaviour and academic fraud.

This Charter bears a family resemblance to various schemes: the GNU Free Documentation, Creative Commons, Creative Archive and Open Content licenses. Moreover, the current version borrows a few notions and phrases from them. Nevertheless this Charter stands on its own and must not be interpreted as adjusted to any of those models. Although all those approaches share a strong leaning towards open and free content-divulcation, a distinctive view prevails in each case.

Openness and freedom are broad and loose concepts admitting of sundry conceptions. 'Free content' has been commonly agreed upon to stand for content which readers can and may honestly use without hindrance or threat of reprisals — rather than as content they necessarily get for free — but there are several views as to the justification and implementation of those liberties. A thorough elucidation of those problems goes beyond the scope of this Preamble, though.

In order for the current scheme to stand out as a differentiated dispensation, it will be called '**the allodial approach to intellectual property**'. Four features make it unique:

- The first one is that, while it allots to everybody an unhampered access to intellectual goods **without discrimination**, it also recognizes the right to private intellectual ownership, construed — according to Civil-Law systems — as twofold: an entailed, perpetual, untransferable and non-renounceable entitlement to authorship-recognition and to the integrity of the work; and an alienable right to the exclusive use and profit from that work.
- The second one is that it regards a compilation (such as a journal Issue) as a collective work, and hence as an intellectual entity jointly authored and owned by the Editor and the contributors taken together as a group, while each article, by itself, is an intellectual entity belonging to its author alone.
- The third one is that it resorts to the legal notion of *servitude* or *easement*, which enables us to reconcile all those rights and liberties, in the same way as happens with land freeholds: certain positive or negative interests, franchises or privileges (*servitudes*) may be conferred on non-owners over an estate for certain legitimate reasons and purposes.
- The fourth one is that it rules out any registration requirement, as contrary to the rights to privacy, unobstructed access and uncontrolled dissemination.

* * *

The **SORITES** team will publish revised versions of the **SORITES** Charter, similar in spirit to the present version, but different in detail, to address new problems or concerns. Authors and readers having obtained certain rights in virtue of the current Charter will keep those rights and will never be subject to restrictions brought in by later versions.

Everybody is free to copy this Charter, and to modify or attune it to their own needs, provided appropriate authorship credit is given to the **SORITES** team.

DEFINITIONS

[D01] A **separate material** is an intellectual product which is released, broadcasted or published as an intangible entity standing on its own rather than as a section or a division of another intellectual product.

Parts of an aggregated work — such as essays gathered into a collection — are not separate materials, since they are not released as intellectual products on their own.

[D02] An **Issue of the *SORITES* journal as a whole** is either:

- (i) a single document or file (be it its official version or any unofficial rendering also released by the **SORITES** team as an undivided file, such as the PDF file or a print-out thereof); or
- (ii) a set of files produced by slicing one of the entire-file versions in order to facilitate handling, browsing or downloading (in which case the conveyor is bound to distribute the whole collection taken together at the same time).

[D03] A reader of **SORITES** is an individual who, abiding by all the terms laid down in this Charter, accesses the **SORITES** site, or a mirror thereof, or receives or uses any file or document directly or indirectly proceeding from that site.

[D04] As a rule, institutions and organizations are not readers. They are considered users only when this Charter explicitly confers a certain right on them.

[D05] Private use is personal use by an individual. Sharing downloaded or stored files within an organization is a divulgation, not private use.

[D06] An Issue of the journal *SORITES* is an aggregated work; the articles it contains are not separate materials.

[D07] A copy of a material is either a print-out on paper or a reproduced file or set of files.

[D08] A reproduced file, or a set of files, is either (i) a duplication of the original file, or set of files, as released by the **SORITES** team; or (ii) any result of converting the downloaded file or files to a different format by applying conversion tools, provided the content is not garbled or curtailed.

[D09] A different format is any modification which leaves the content intact and unaltered.

[D10] The phrase ‘copy or partial copy’ covers any print-out, extended quotation, electronic posting or display as well as any other reproduction.

ARTICLES

Art.01. — **SORITES** is both a resource site, <<http://sorites.org>> (or set of sites, such as <sorites.info>, <sorites.net>, etc), and a digital journal, *SORITES*, consisting in a series of issues. Both of them belong to the publisher, the **SORITES** team.

Art.02. — Each Issue of the *SORITES* journal as a whole is a collection, an aggregated work, jointly owned by the contributors and the publisher.

Art.03. — Individual papers remain the private property of their respective authors (the contributors)

encumbered with the following servitudes:

- (i) No part of any such paper may be printed or displayed elsewhere, or incorporated into a publication of any sort, unless *SORITES* is therein clearly and explicitly acknowledged and cited as the original source of publication.
- (ii) In particular, when a contributor posts a version of his or her paper on a home page, institutional web-site or subject repository, a link must be made to the version displayed at the **SORITES** web-site.
- (iii) When contributors merge a recast version of a substantial portion of an article published in *SORITES* into another paper, they must also give due credit to the *SORITES* source and clearly inform their audience how to access the **SORITES** site (including a reference to the official web-site, <http://sorites.org>).
- (iv) Contributors adhere to the other terms and provisions contained in this Charter, grant **SORITES** readers and users the rights provided by Art.06 and consent to abide by the publisher's claims and commitments (arts. 04, 13 and 14).
- (v) When contributors republish their papers, they are free to apply any other License, provided clauses (i), (ii) and (iii) of this article are honoured.
- (vi) Any individual who signs a co-authored article is a contributor who accedes to the above burdens being laid on the paper. Arrangements between co-authors cannot stand in the way of that undertaking.

Art.04. — By submitting a manuscript to *SORITES*, contributors declare: that it is an output of their personal research; that it has not been published yet (in the sense of appearing under an ISSN or an ISBN); that it is not currently submitted for publication elsewhere; and that it complies with established Copyright laws. Acceptance of a manuscript by the **SORITES** team is done in good faith under the assumption the declaration is truthful.

Art.05. — The authors of the included papers and the publisher — whether jointly or separately, as the case may be — hereby reserve all rights not conferred to other parties in this Charter.

Art.06. — Subject to honouring the conditions set forth in this Charter, readers are hereby accorded the following rights:

- [i] The right to freely access, view, read, search, store, and link to, any material offered at the **SORITES** site, especially any Issue of the *SORITES* journal, any paper therein included, and any other file or document displayed or supplied at the **SORITES** site.
- [ii] The right to duplicate single detached articles for a personal, private and fair use done in good faith. Fair use encompasses teaching, study, criticism and review.
- [iii] The right to reproduce and distribute on any medium *verbatim* copies of any **separate material** released at the **SORITES** site, provided the material is reproduced **in its entirety** and preserved as an Invariant Document, unaltered in its text, titles and headings.

Art.07. — When a copy of a separate material published by **SORITES** is incorporated into a volume or a package containing other materials, the content obtained from the **SORITES** site must be printed on separate sheets, clearly demarcated, by page breaks, from those accompaniments. No text or logo from other sources must be mingled with, or interpolated into, the **SORITES** material.

Art.08. — The freedom awarded to readers in Art.06 means that:

- (1) No fee will be charged or requested by the **SORITES** team.
- (2) No registration will be demanded.
- (3) No one will be locked out of accessing the **SORITES** site.
- (4) As regards issues of the *SORITES* journal and papers therein published, exercising those rights

is not conditional upon reciprocation or remuneration of any sort, whether moral or material, or upon any consideration other than respect for intellectual property claims raised in this Charter.

- (5) As regards other materials offered at the **SORITES** site, moral consideration may be demanded, such as not using those materials for any unethical purpose or activity.

Art.09. —

- [1] Private business companies are not entitled to incorporate a copy of an Issue of the *SORITES* journal into any commercial material they sell for profit without the previous conformity of the **SORITES** team.

However they are hereby allowed to combine a whole Issue of the *SORITES* journal — or any other material offered as a separate unit at the **SORITES** site — into publications they circulate at no charge, provided they observe the other provisions of this Charter, especially Art.07.

- [2] Learned societies and academic establishments are hereby authorized to sell integral copies of issues of the *SORITES* journal or any other material offered as a separate unit at the **SORITES** site.
- [3] Search engines and indexing data-bases are entitled to store, reproduce and display integral or partial copies of any material published by the **SORITES** team, including issues of the *SORITES* journal, even of detached papers, provided due credit and links to the original **SORITES** site (or a mirror thereof) are ostensibly shown.

Art.10. —

- [i] Except as provided by Art.09[3] or by fair use legal provisions, no copy or partial copy of a detached paper may be distributed without its author's acquiescence.
- [ii] The **SORITES** team hereby authorizes any such transmission subject to the author's approval, provided the copyright box on top of the paper is literally incorporated into the copy or partial copy.

Art.11. —

- [i] No part of any Issue of the *SORITES* journal may be delivered to a plurality of individuals — whether in writing or by any other means — unless the source is clearly and explicitly recognized. Furthermore, such a recognition must comply with the following stipulations.
- [ii] No part of any Issue of the *SORITES* journal or of any paper therein included may be conveyed to others by means of reproduction, quotation, copy or paraphrase, without a clear and explicit acknowledgement of the Issue of the *SORITES* journal, the author's name and the paper's full title.
- [iii] Whenever the quotation occurs within a publication, it is also mandatory to mention the date of publication and the official pages (as shown within the Copyright box on top of the paper), as well as the ISSN (1135-1349) and the official web-site: <http://www.sorites.org>, or a mirror thereof such as <http://www.sorites.info>.
- [iv] When the quotation takes place in an electronic medium, a link to the **SORITES** site (or a mirror thereof) is also imperative.

Art.12. —

- (1) Any use of copies or partial copies of an Issue of the *SORITES* journal, or of papers therein included, not permitted in this Charter will be a violation of the contributors' and the publisher's intellectual property.
- (2) Distribution of copies of detached papers without the author's consent is hereby prohibited unless pursuant to Art.09[3].
- (3) Derivative works are unlawful, except in so much as they are provided by Art.09[3].
- (4) Shortening, abridgment or truncation is only licit as provided either by Art.09[3] or else by fair use terms in accordance with intellectual property legislation.

- (5) In any case, the two following activities are absolutely forbidden: 1. change of words or any other semantic or syntactic alteration; 2. false, faulty, obscured or disingenuous attribution.
- (6) The **SORITES** team reserves the right to sue malicious or reckless transgressors of this Charter as culprits of plagiarism, forgery or other misdemeanours, requesting both compensatory and punitive remedies. It also reserves the right to publicly denounce serious copyright violations.

Art.13. —

- [i] Any Material displayed at the **SORITES** site, including issues of the *SORITES* journal, is provided as is, without any guarantee of any kind.
- [ii] While the **SORITES** team tries to ensure the correctness of all the content it publishes, it makes no representation or warranty as to the accuracy, completeness or fitness for any purpose of such a content and — to the maximum extent allowed by law — disclaims any pledges or assurances, whether express or implied, to the effect that any published material is free of defects or non-infringing.
- [iii] By furnishing those documents, the **SORITES** team does not underwrite any commitment except in so much as explicitly espoused in the present Charter.
- [iv] Any views expressed in papers published in **SORITES** are the opinions of the contributors and not those of the **SORITES** team. The publisher disowns being answerable for the contributors' ideas or arguments.

Art.14. — Under no circumstances and under no legal theory shall the **SORITES** team, or any contributor or supplier of **SORITES**, be liable to any person for any indirect, special, incidental or consequential damages for loss of good will, mistaken belief, confusion, delusion or any and all other damages or losses suffered as a result of downloading, reading, or using any materials retrieved from the **SORITES** site or included in any Issue of the *SORITES* journal. The user assumes, at their sole risk, full responsibility for relying upon any opinion or information contained in any material published or displayed by **SORITES**.

Art.15. —

- [i] By opening, accessing, downloading, storing, reading or in any other way using any material offered at the **SORITES** site or any part thereof, readers enter into a synallagmatic contract with the **SORITES** team binding themselves to full acceptance of the above stipulations. Breach of any of those provisions entails automatic forfeiture of all rights derived from this Charter and termination of the contract.
- [ii] If, upon downloading or storing a file containing an Issue of the *SORITES* journal or a part thereof, or receiving an electronic or printed copy thereof, or any other material obtained from the **SORITES** site, a reader fails to assent to the conditions stipulated in this Charter, he or she is obliged to immediately discontinue reading, storing, sharing or using the material in any way.
- [iii] In the event of such a termination, though, third-party rights derived from usage prior to the termination will survive.
- [iv] Readers to whom previous versions of this Charter bestowed rights in excess of those granted here, or subject to less stringent conditions, may continue enjoying those rights.
- [v] The **SORITES** team will revise this Charter at its sole discretion. Acquired rights will endure, though.
- [vi] In the event that there exists any inconsistency between this Charter and legal commandments, the latter will take precedence.
- [vii] If a reader cannot satisfy simultaneously the duties under this Charter and commitments stemming from custom, contract of other sources, they must either extricate themselves from those commitments or else refrain from accessing **SORITES** and from using any material released by the **SORITES** site.

[viii] This Charter is the complete and exclusive agreement between the **SORITES** team, its contributors, and its readers which supersedes any proposal or prior deal, oral or written, and any other communication relating to the subject matter of this Charter.

Art.16. — Rights granted in this Charter are contingent upon third-party intellectual-property claims.

Art.17. — In compliance with Spanish Law, all issues of the *SORITES* journal are legally registered, electronic copies being deposited with the competent authorities, namely the Legal-Deposit office of the Autonomous Community of Madrid (Registration: M 14867-1995.)

Art.18. — Although a mirror of **SORITES** is displayed at the internet host <www.ifs.csic.es/sorites/> belonging to the Spanish Academy of Sciences, or C.S.I.C., thanks to a permission graciously given by the system's administrators, the journal is not sponsored or endorsed by that institution, the only owner and publisher being the **SORITES** team.

Madrid. 2006-10-05 [V.2.2]

(Previous versions: 1995-04-10; 1996-12-31; 1998-04-27; 2002-10-03; 2004-12-31;
2005-01-18 [V.2.1])

Lorenzo Peña (The **SORITES** Team)

<<http://sorites.org/copyrigh.htm>>

Release Notice

This issue of **SORITES** (Σωρίτης) is made available in several formats, but its only official version is the WordPerfect 5.1 document released with filename:

sorite19.wp

Although each version, whether official or not — as initially released on March 18, 2008, by the **SORITES** team — is an entire seamless file, it may be splitted down into chunks in order to facilitate downloading, browsing, transferring or e-mailing. In such cases, the unity of this issue of **SORITES** as a whole must be preserved by keeping the ensuing collection intact.