Proceedings of the

Graduate Philosophy Conference on Normativity

Amsterdam, August 29 & 30, 2008

Proceedings of the
Graduate Philosophy Conference
on Normativity

Amsterdam, August 29 & 30, 2008

Theodora Achourioti, Edgar Andrade,
and Marc Staudacher (eds.)

Preface

The Graduate Philosophy Conference on Normativity was organized by the Department of Philosophy and the Institute for Logic, Language and Computation of the Universiteit van Amsterdam. The organizers were Theodora Achourioti, Edgar Andrade, and Marc Staudacher.

The conference organizers have consulted: Dr. Paul Dekker, Prof. Dr. Wolfram Hinzen, Prof. Dr. Michiel van Lambalgen, PD Dr. Benedikt Löwe, Prof. Dr. Martin Stokhof, and Prof. Dr. Frank Veltman.

The organizers are most grateful to the following sponsors for their support: Institute for Logic, Language and Computation, Research Training Site GLoRiClass, NWO-Project "The Origins of Truth, and the Origins of the Sentence," Gesellschaft für Analytische Philosophie, and the Leerstoelgroep Logica en Taalfilosofie of the Universiteit van Amsterdam.

Finally, the organizers want to thank the speakers for their contributions.

The Editors,
Amsterdam, March 2009

Contents

Naturalism and Normativity: Sellars' 'Janus-Faced' Space of Reasons

Dr. Jim O'Shea (University College Dublin)

`jim.oshea@ucd.ie`

I

Consider the following two views famously held by Wilfrid Sellars:

[1] The 'space of reasons' view:

> The essential point is that in characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says.[1]

[2] The 'scientia mensura' view:

> . . . in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not.[2]

In what follows I will not be concerned with those aspects of Sellars' *scientia mensura* view that are bound up with his notorious

---

[1] Sellars (1956, VIII §36).
[2] Ibid., IX §42.

contention that the common sense conception or 'manifest image' of the world is strictly speaking false. (Sellars bases the latter argument on the explanatory inadequacy of the manifest image in comparison with the successive replacement ontologies of the advancing scientific image of the world.) Rather, the broader naturalistic thesis upon which I will focus here is equally central to Sellars' *scientia mensura* outlook and concerns distinctions that he took to be applicable to any 'matter-of-factual' or empirical conceptual framework whatsoever. What I am taking from the *scientia mensura* passage for my present purposes, then, is a strong version of naturalism that can be seen in the following striking passage from Sellars' article from the same period, 'Counterfactuals, Dispositions, and the Causal Modalities':

> **[3]** Clearly, to use the term 'ought' is to prescribe rather than describe. The naturalistic 'thesis' that the world, including the verbal behavior of those who use the term 'ought' — and the mental states involving the concept to which this word gives expression — can, 'in principle', be described without using the term 'ought' or any other prescriptive expression, is a logical point about what is to count as a description *in principle* of the world. For, whereas in ordinary discourse to state what something is, to describe something as ϕ (e.g., a person as a criminal) does not preclude the possibility that an 'unpacking' of the description would involve the use of the term 'ought' or some other prescriptive expression, naturalism presents us with the ideal of a pure description of the world (in particular of human behavior), a description which simply says what things *are*, and never, in any respect, what they *ought* or *ought not* to be; and it is clear (as a matter of simple logic) that neither 'ought' nor any other prescriptive expression *could* be *used* (as opposed to *mentioned*) in such a description.[3]

There is much in this passage that I want to return to, but for now I want to emphasize that the 'dimension of describing and explaining the world' to which the *scientia mensura* applies is a dimension that concerns what Sellars calls 'pure empirical description', "a description which simply says what things *are*, and never, in any respect, what they *ought* or *ought not* to be." So I want to understand the *scientia mensura* thesis as the endorsement of that kind of naturalis-

---

[3]Sellars (1957, §79).

tic thesis (and of course I'll have to say some more about what that kind of naturalism *is*).

One other point of clarification. When considering Sellars' characterization of a 'purely naturalistic description' of the world as being an essentially *non-normative* description of the world, we have to make certain distinctions. Most obviously, as one who also holds the Sellarsian *space of reasons* thesis, Sellars would himself be the first to emphasize that any *describing* of the world, including a so-called 'purely naturalistic description', must *itself*, *qua* conceptual activity, be an essentially normatively characterized activity. But while it is thus essential to any act of *describing*, or to any *use* of a concept in general, that it is *as such* the occupying of a position within the normative space of reasons, the point is that the object or state of affairs thus *described* in a purely naturalistic description will be a non-normative item in the brutely natural world. Crudely put, the *concept* of a stone is an intrinsically normatively characterized item, but stones themselves know no ought's. By contrast, if I use the concept of obligation to refer to Jones's obligation to pay his debts, then what my concept in this case describes (and not merely the describing) is itself an item that is essentially normatively characterized. So that is not a 'pure description' in the sense to which Sellars' *scientia mensura* is meant to apply. I can *mention* uses of the word 'obligation' or of the concept of obligation within a purely naturalistic description; but I cannot *use* the concept of obligation within a purely naturalistic description to refer to *obligations* as basic items in the world.

So let us return to the *scientia mensura* in [2], which tells us that "in the dimension of describing and explaining the world, science is the measure of all things ... "; and let us take this for present purposes to be the view that the business of telling us what there really is in the world is a matter that pertains exclusively to purely naturalistic, non-normative descriptions of the world in the sense just explained.

Now among the 'things that there are' in the world are the conceptual activities of persons. And so strictly speaking, Sellars' *scientia mensura* view should entail that "in the dimension of describing and explaining the world," science in the form of purely naturalistic descriptions is also 'the measure' of those particular items in the world that are the conceptual activities of persons.

3

But if so, then there seems to be a *prima facie* tension between the *scientia mensura* view we have been canvassing, and the view expressed in the 'space of reasons' passage (in [1]). The space of reasons passage is an instance of Sellars' general view that it is essential to any conceptual activity as such that it is characterizable in normative terms. And this is so (as Robert Brandom in *Making It Explicit* (1994) and John McDowell in *Mind and World* (1994) have each stressed in their distinctive Sellarsian ways) for the strong reason that the *content* of any concept is the conceptual content that it is only of virtue of playing a specific normative role or having a certain normative status within the logical space of reasons. So if we take as our target of purely naturalistic description the activity of using a concept itself, then the 'space of reasons' passage might seem to be in serious conflict with the *scientia mensura* passage. For according to the 'space of reasons' passage, it seems that purely empirical or naturalistic description is simply not cut out for capturing what is *essential* to or distinctive of conceptual activity as such: namely, its essentially normative character. But if conceptual activity is a thing of which science is not the measure, then presumably the *scientia mensura* view is false.

In short, we might seem to have Sellars holding the following contradictory views:

(A) **the space of reasons view**, which seems to imply that science (or purely naturalistic description) cannot in principle explain what is essential to conceptual activity as such; and

(B) **the** *scientia mensura* **view**, which seems to imply that science (or purely naturalistic description) can in principle explain what is essential to conceptual activity as such.

For if science is the explanatory measure of *all* the things that are, surely our own conceptual activities are among the most interesting things that there are in the world.

So let us suppose that we agree with Sellars, Brandom, and McDowell in upholding the normative *space of reasons* view of conceptual content in general. Then we seem to have three options for resolving the conflict between the two Sellarsian views expressed in [1] and [2].

First, we could accept (B) as a correct description of what the *scientia mensura* view entails, but respond by happily consigning the *scientia mensura* view to the flames. As good 'left-wing Sellarsians' (Brandom, McDowell, Rorty), we could chalk up the endorsement of the *scientia mensura* thesis by Sellars to a regrettable scientistic naturalism that he could never quite manage to shake off — an unfortunate legacy from positivism that prevented Sellars himself from finally entering the true Hegelian promised land of the space of reasons. I find this an attractive outlook in many respects, but my own suggestion will be that one can in certain crucial respects *reconcile* the neo-Hegelian outlook characteristic of the left-wing Sellarsians with the particular version of naturalism that I have characterized as essential to Sellars' *scientia mensura*. That is, instead of rejecting Sellars' *scientia mensura* view one could seek to reconcile that thesis with the *space of reasons* view in either of two ways: either by rejecting the apparent entailment in (A), or by rejecting the apparent entailment in (B). It must have been one of those two reconciling options that Sellars himself had in mind.

So suppose we reject the implication in (B). That is, suppose we accept the *scientia mensura* view, but argue that this view does not entail that pure empirical description could even in principle explain conceptual activity as such, i.e. as the essentially normative phenomenon that it is. Certainly Sellars' careful restriction of the *scientia mensura* view to "*the dimension of describing and explaining the world*" was intended to contrast with *other* dimensions of human conceptual activity that are *not* in the business of explaining and describing the world — most notably, all those normative and pragmatic dimensions of human conceptual activity that have to do with intending, instituting, recommending, endorsing, and prohibiting various courses of action. One could accept the *scientia mensura* as the view that there is in principle a non-normative, purely naturalistic description of all that there is in the world, a level of explanation that is privileged in the dimension of describing and explaining the world. But one could argue that conceptual activity as a normatively structured activity — as a network of normative statuses that is instituted, maintained, and continually remade through the normative pragmatic agencies and attitudes that constitute it — is not the sort of thing that could even coherently be supposed to be

5

the proper target of a purely naturalistic, *non*-normative description as such.

This, too, strikes me as a very plausible view. In fact, I think that its real effect, in the end, is to restrict the *scientia mensura* thesis in such a way that the resulting philosophical outlook becomes pragmatically indistinguishable from the first, 'neo-Hegelian' response to the tension — the response that rejects the *scientia mensura* view and argues that there is no compelling reason why we should experience philosophical discomfort by admitting the ontological ultimacy of the normative *space of reasons*. (John McDowell's *Mind and World* is a sophisticated defense of this latter view.)

However, my suggestion is that Sellars' own position incorporates a rejection of the putative entailment in (A) rather than in (B). That is, while on the one hand defending the constitutive irreducibility of the normative space of reasons, Sellars also agreed with Hobbes and Hume, *et al.*, that the most profound task confronting the synoptic philosopher and scientific theorists is precisely (to quote Hume's memorable phrase) "to march up directly to the capital . . . , to human nature itself; which being once masters of, we may every where else hope for an easy victory" (*Treatise* XV-VI). While others do the good work of developing the scientific image *of the world*, what Sellars was interested in was 'Philosophy and the Scientific Image *of Man*'. Unlike Hume, however, what Sellars thought was required for this task would be the possibility of a purely naturalistic, non-normative explanation of the nature of *persons* as concept users as such; that is, as inhabitants of a normative space that makes it possible (as Kant saw) for them to have any contentful thoughts and rational intentions in the first place. That is why, in the passage quoted in [3] as throughout his writings, Sellars emphasizes that the most important philosophical target of a possible purely naturalistic description is not just a scientific explanation of objects but most importantly of "the behavior of those who use the term 'ought'."

My own view is that Sellars attempted to boldly go in this omnivorously naturalistic direction in a way that also allows us to maintain, and in fact seeks to *explain*, our neo-Hegelian comfort with the irreducibly normative space of reasons. In fact, that is precisely the philosophical aim of what Sellars calls a 'synoptic vision' that would finally unite the manifest image of persons with the scientific image of nature. In what remains I will try to sketch briefly just a few

central aspects of Sellars' particular way of reconciling the two Sellarsian views, taking my lead on this occasion from the passage cited in [3].

II

The overall project that Sellars undertook may be viewed as having had two main parts, which I will call the 'analytic' task and the 'explanatory' task:

[4] **The Analytic task:** to show that *extensions are limiting cases of essentially normatively characterized intensions*:

> My ultimate aim is to argue that extensions are limiting cases of intensions and cannot be understood apart from them.[4]

[5] **The Explanatory task:** to show that the essentially normatively characterized *intensional* conceptual activities of persons, while conceptually irreducible and pragmatically ineliminable, *can in principle be given an ideal explanatory account in purely extensional, naturalistic terms.*

Put in Humean terms, the Analytic task is to articulate the essentially normative character of the "sciences of *Logic* [and] *Morals*"; while the Explanatory task is to storm the citadel of norms *qua* norms, "human nature itself," and sketch out the possibility of a purely naturalistic explanatory account of the essentially normatively characterized activities of persons themselves. The Analytic task is the one set by the space of reasons thesis; and the Explanatory task is the one set by the *scientia mensura* thesis.

One conveniently brief way to convey what I take to be the Analytic task in Sellars is by referring to Robert Brandom's much discussed attempt to execute a similar project in *Making It Explicit*. In his attempt to rigorously carry out the task of analyzing representationalist and referential idioms in ultimately normative pragmatic terms, I think Brandom does capture and extend much that lies at the heart of the Analytic aspect of Sellars' views. (John McDowell's Davidsonian take on the space of reasons takes a very different path toward some of the same broadly Sellarsian Analytic ends.)

---

[4]Sellars (1968, ch. III, §43).

At any rate, the crucial breakthrough for Sellars in this Analytic dimension turns out to be his well-known development (already by 1950) of a normatively characterized functional role or conceptual role semantics. The key notion in Sellars' analysis is the concept of *pattern-governed linguistic behavior* as a species of *norm-conforming regularity*. (Note that since Sellars was in the process of inventing a normative *functionalist* philosophy of mind during the 1950s, all this talk of linguistic behavior should be understood as affording a model for the interpretation of genuinely 'inner' conceptual episodes or mental events.) Pattern-governed linguistic behavior is activity that both conforms to, and is generated by, the normative 'ought-to-be's of social linguistic practice, resulting in the Sellarsian inferential role account of meaning and intentional content of the sort that will be familiar to readers of Brandom.

In what remains I will attempt to articulate briefly the nature of Sellars' Analytic task and in particular show how it is ultimately undergirded by an account of 'purely naturalistic descriptions' in the form of a certain theory of mental and linguistic representation; and then I will conclude with a few remarks concerning the naturalistic explanatory task itself.

The Analytic task begins with socially instituted norms of linguistic practice; in particular, with social-linguistic norms prescribing what linguistic behaviors 'ought-to-be' exhibited in 'language-entry' responses to objects, in 'intra-linguistic' patterns of inference, and in 'language-exit' intentions and volitions. The implicit endorsement of the normative ought-to-be rules of the language generates corresponding causal uniformities in the actual pattern-governed linguistic behavior of users of the language. This aspect of Sellars' view is captured in what I have elsewhere called Sellars'

[6] Norm/nature [or rule/uniformity] **meta-principle:**

> Espousal of principles is reflected in uniformities of performance.[5]

The key to Sellars' various analyses of meaning and knowledge, and of truth and representation as well, concerns the complex relationships of pragmatic presupposition that result from the

---

[5]Sellars (1962, p. 216), page numbers from reimpression in Sellars (1963).

various normatively constrained causal uniformities in which pattern-governed linguistic behavior consists. For example, perceptual knowings, on Sellars' account, as normative standings in the space of reasons, systematically presuppose various corresponding norm-generated reliable causal connections in the world-to-mind causal direction, in our habitual language-entry transitions from the presence of objects to the corresponding correct linguistic-conceptual responses. Similarly, reason-based intentions and volitions, in the form of (or modelled on) language-exist transitions, presuppose various norm-generated, rule-governed reliable causal connections in the mind-to-world causal direction, from our 'intention to do A *now*' to the *doing of A* (other things being equal). These normative standings in the space of reasons and the particular natural-causal uniformities that they presuppose are the two mutually conditioning sides of the same coin throughout what I have elsewhere called Sellars' *naturalism with a normative turn*.[6] As Sellars sums it up in one place:

> [**7**] Thus, the fact that the uniformities (positive and negative) involved in language-entry, intra-linguistic and language-departure transitions of a language are governed by specific ought-to-be statements in its meta-linguistic stratum, and these in turn by ought-to-bes and ought-to-dos concerning explanatory coherence, constitutes the Janus-faced character of languagings as belonging both to the causal order and the order of reasons. This way of looking at conceptual activity transposes into more manageable terms traditional problems concerning the place of intentionality in nature.[7]

Sellars' Analytic task crucially bottoms out in a normatively constrained causal theory of linguistic and mental representation. This is Sellars' notorious 'Tractarian picturing' account of the purely naturalistic descriptions of matter-of-factual states of affairs that, he contends, are essential to telling any empirical 'world story' at all. However, this bottom-level, naturalistic representationalist story in Sellars is not as implausible as it has often been taken to be by neo-pragmatists and neo-Hegelians otherwise sympathetic to Sellars' views on the space of reasons. The naturalistically conceived representings of matter-of-factual nature that figure in all of Sellars' writ-

---

[6]Cf. O'Shea (2007).
[7]Sellars (1979, V §64).

ings (including passage [3]) are essentially normatively constrained by the same Analytic package of ought-to-bes as any other level of linguistic behavior on Sellars' account. There is indeed a purely causal, naturalistic account of the representational isomorphism that obtains (or *ought* to obtain) between mind and world, on Sellars' account. But on Sellars' account such 'purely world-descriptive' linguistic and mental representations are always strictly normatively generated and constrained; mental representations are always *Janus-faced* on Sellars' view, simultaneously backed by norms and fronting the world; or to switch metaphors, the causal-representational regularities are the slaves of our rule-governed reasons, and it is essential to the resulting 'purely descriptive' pictures of the world that they are the causal shadows of norms.

Contrary to popular wisdom, then, on Sellars' account there is nothing 'normatively blind' about our matter-of-factual representings or 'picturings' of those brutely non-normative complexes and sequences of events that make up the natural world as it is in itself.

That, in a nutshell, is the normative-pragmatic or 'Analytic task' as Sellars conceived it; and it thus bottoms out in an account of the normatively constrained, 'purely naturalistic description' or representation of the complexes and sequences of events that make up the world. But among the most interesting things that there are in the world, as we have seen, is "the behavior of those who use the term 'ought' " [from passage [3]]. And that, finally, sets up the most important Explanatory task issuing from Sellars' *scientia mensura*.

Since on Sellars' view perception, knowledge, and action are all species of conceptual activity that is essentially normative, the key explanandum in Sellars' Explanatory synoptic vision of 'man-in-the-world' is a 'purely naturalistic descriptive' account of the *nature of normative activity itself*. Unlike the eliminativists, however, Sellars conceived this Explanatory task in a way that he thinks preserves and explains the pragmatic or conceptual irreducibility of the normative as such. So I'll end with just a few remarks on how I think Sellars conceived the Explanatory task.

Here is one remark by Sellars on the key Explanatory task:

> [8] ... unless and until the 'scientific realist' can give an adequate explication of concepts pertaining to the recognition of norms and standards by rational beings his philosophy of mind must remain radically unfinished business. Chapter VII

on objectivity and intersubjectivity in ethics is consequently,
the keystone of the argument, for the lectures have stressed at
every turn the normative aspects of the concepts of meaning,
existence, and truth.[8]

At its core, Sellars' explanation of the nature of normative ac-
tivity as such consisted in a behavioral-functionalist theory of the
nature of intentions and volitions, and by extension of those *com-
munal 'We'-intentions* by which 'ought-to-be' norms are instituted
and have their causal efficacy. Sellars attempted to sketch an 'in-
ternalist' account of normative 'ought's as communal 'we'-intentions
that would capture both their intersubjectivity and their essentially
motivating character. In an early article Sellars had characterized
his naturalistic account of normative rule-following as follows:

> **[9]** The historically minded reader will observe that the con-
> cept of rule-regulated behavior developed in this paper is, in
> a certain sense, the translation into behavioristic terms of the
> Kantian concept of Practical Reason. Kant's contention that
> the pure consciousness of moral law can be a factor in bring-
> ing about conduct in conformity with law, becomes the above
> conception of rule-regulated behavior. However, for Kant's
> conception of Practical Reason as, so to speak, an intruder
> in the natural order, we substitute the view that the causal
> efficacy of the embodied core-generalizations of rules is ulti-
> mately grounded on the Law of Effect, that is to say, the role
> of rewards and punishments in shaping behavior.[9]

In all this, however, it remains the case that on Sellars' account
the relevant behavioral-functional roles are *normatively* construed;
the relevant regularities cannot be picked out in merely statistical
terms, as it were. The relevant causal patterns involved in "the be-
havior of those who use the term 'ought' " remain essentially *Janus-
faced* to the end. With both the Analytic and Explanatory dimen-
sions in view, however, we now have two ways of looking at the same
behavioral sequence of events in nature. Using the broader Sellarsian
distinction that is presupposed in passage [3], there is a difference
between *mentioning* or describing how 'ought's figure in the pattern-
governed behaviors in which our normatively constrained activities

---

[8]Sellars (1968, 'Preface', p. X).
[9]Sellars (1949, note 3).

are realized in nature; and someone's actually *using* or *having* an 'ought'-laden sentence or thought. There is thus an ineliminable *normative pragmatic* dimension to Sellars' Janus-faced naturalism. The normatively characterized things that we do and say cannot be done and said in any other way than by doing and saying them in normatively characterizable ways. And any ideal Explanation that would refer to or *mention* those sayings and doings would be a 'purely naturalistic *describing*' of the world that is itself normatively constrained. So while there *is* (according to Sellars) a legitimate sense in which the patterns of uniformity mentioned in such an Explanatory account would be describable without *using* any normative 'ought's to refer to any property of obligatoriness in the natural world, the normative vocabulary is nonetheless pragmatically and conceptually ineliminable. One last passage, with my own added explanatory gloss:

> [**10**] Whatever users of normative discourse may be *conveying* about themselves and their community when they use normative discourse,
>
> [(J.O'S.:) the *information conveyed* in such uses concerns the particular patterns of natural causal uniformity that are appropriate to, and hence are *presupposed* by, that particular normative discourse]
>
> what they are *saying* cannot be said without using normative discourse. The task of the philosopher cannot be to show how, in principle, what is said by normative discourse could be said without normative discourse, for the simple reason that this cannot be done.[10]

Robert Brandom once expressed the hope that the disagreements between the 'left-wing' and 'right-wing' Sellarsians might be addressed in a manner more satisfying than was historically case with the disputes between the left-wing and right-wing Hegelians — which were settled at the Battle of Stalingrad. My suggestion has been that Sellars' own account arguably involves a more plausible systematic integration of its strong scientific naturalism and its irreducibly normative pragmatics than has sometimes seemed to be the case.

---

[10]Sellars (1953, §66).

Dr. Jim O'Shea (University College Dublin)

References

Brandom, R., 1994. Making it Explicit: Reasoning, Representing, and Discursive Commitment. Harvard University Press, Cambridge, Mass.

McDowell, J., 1994. Mind and World. Harvard University Press, Cambridge.

O'Shea, J., 2007. Willfrid Sellars: Naturalism with a Normative Turn. Polity Press, Oxford.

Sellars, W., 1949. Language, Rules, and Behavior. In: Hook, S. (Ed.), John Dewey: Philosopher of Science and Freedom, a symposium. The Dial Press, New York, pp. 289–315, reprinted in Sicha (ed.) (1980).

Sellars, W., 1953. A Semantical Solution to the Mind-Body Problem. Methodos 5, 45–82, reprinted in Sicha (ed.) (1980).

Sellars, W., 1956. Empiricism and the Philosophy of Mind. In: Feigl, H., Scriven, M. (Eds.), The Foundations of Science and the Concepts of Psychoanalysis, Minnesota Studies in the Philosophy of Science, Vol. I. University of Minnesota Press, Minneapolis, MN, pp. 127–196, reprinted in Sellars (1963).

Sellars, W., 1957. Counterfactuals, Dispositions, and the Causal Modalities. In: Feigl, H., Scriven, M. (Eds.), Minnesota Studies in the Philosophy of Science, Vol. II. University of Minnesota Press, Minneapolis, MN, pp. 225–308.

Sellars, W., 1962. Truth and Correspondence. Journal of Philosophy 59, 29–56, reprinted in Sellars (1963).

Sellars, W., 1963. Science, Perception and Reality. Routledge & Kegan Paul Ltd, London.

Sellars, W., 1968. Science and Metaphysics: Variations on Kantian Themes. Routledge & Kegan Paul Ltd, London.

Sellars, W., 1979. Naturalism and Ontology. Ridgeview Publishing Co., Atascadero, CA.

Sicha (ed.), J. F., 1980. Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars. Ridgeview Publishing Co., Atascadero, CA.

Assessment-sensitivity and the Naturalistic Fallacy

Joshua Beattie (University of California, Berkeley)

jsbeattie@yahoo.com

The notion of normativity looms large on the contemporary philosophical scene, and a major reason is that normativity is seen as having important anti-naturalist implications. That is, it is widely — though by no means universally — accepted that there is a "naturalistic fallacy," that the normative is *sui generis* and so cannot be identified with, reduced to, or derived from the non-normative. Among those inclined to agree with this claim, however, there is still no firm consensus regarding exactly why it is true, what its precise scope is, or how it is best argued for. It would be nice to be able to provide some clarity on these matters, to explain and justify the naturalistic fallacy.

But let's set that concern aside for a moment, and just assume that the fallacy is, for whatever reason, well-defined and well-grounded. That leaves us with a powerful conditional: if there is ineliminable normativity in a certain domain, then any strong form of naturalism in that domain is ruled out. It then becomes a matter of *locating* normativity, so to speak, to see where the anti-naturalist verdict applies. This, too, is far from straightforward. It has been argued that normativity is present, and that the anti-naturalist verdict applies, not only in value theory, but also in epistemology, theory of meaning, and philosophy of mind. These arguments, though, are quite controversial. As Rosen (2001) convincingly maintains, the

notion of normativity is in many ways too ill-defined, too jumbled, to allow for clear-cut answers to basic classificatory questions involving it. Improving upon this situation (to at least some degree) constitutes a second task it would be nice to perform.

In this paper, I want to suggest a way of accomplishing both of these tasks: taking the source of the naturalistic fallacy to be *assessment-sensitivity*. That is, for a claim to be normative in a sense that has anti-naturalistic implications,[1] the truth-value of the claim must depend not (or not only) on the context in which it is produced, but on the context from which it is assessed. For convenience, I will sometimes refer to this main proposal as the assessment-sensitivity thesis.

This proposal has many theoretical benefits. First and foremost, it provides an *explanation* for the naturalistic fallacy, showing precisely *why* some normative claims resist naturalistic treatment. In doing so, it applies in a simple and unified manner to the two traditional paradigms of normativity: *prescriptives* (as focused on by Hume (1739/2000)) and *evaluatives* (as focused on by Moore (1903)).

Second, the proposal offers a principled way of understanding the open question argument, the interpretation of which has been a matter of great dispute. Understood in this way, the argument has the broad scope that is often desired for it (but that has proved difficult to secure): it rules out *a posteriori* identities as well as definitional ones.

Third, the proposal avoids the problems that attend the other main ways of accounting for the naturalistic fallacy. On the one hand, it maintains the truth-aptness of normative claims and so does not have the embedding problems, for example, that plague noncognitivist approaches. On the other hand, it does not rely on motivational internalism, the controversial idea that normative judgments are inherently motivating.

Fourth and finally, the proposal can be smoothly applied to a related issue in the philosophy of mind: it suggests a plausible construal of the much-discussed Davidsonian claim that there can be no psychophysical laws because of the "disparate commitments" of the two realms involved.

---

[1]The qualification matters. The term "normative" is used in various ways, some of which may be compatible with naturalism. The necessary condition I'm proposing is only for being normative in the stronger, anti-naturalistic sense.

So the proposal can be supported by a kind of inference to the best explanation.[2] Of course, the explananda — the naturalistic fallacy, the open question argument, Davidson's argument against psychophysical laws — are questionable themselves, and this makes the situation a bit complicated. For those who find the explananda intuitively compelling, as I think many do, there is good reason to accept the assessment-sensitivity thesis, since it provides the clearest and most effective explanations of them. This at least puts pressure on those who are inclined to accept the explananda but deny assessment-sensitivity — are the considerations telling against assessment-sensitivity really strong enough to warrant giving up its explanatory benefits, particularly when there are no good alternatives?

On the other hand, if one finds the explananda antecedently implausible, then the assessment-sensitivity thesis in fact *cannot* be accepted. This result might be relevant for some people who are independently attracted to a kind of relativism about certain normative matters. More importantly, though, I think it calls for a closer look at why the explananda are actually found implausible. If it is really that one feels *no* intuitive pull toward them at all, then there is no more to be said, and the argument presented here for the assessment-sensitivity thesis doesn't get started.[3] But if in fact it is underlying naturalistic commitments that lead one to reject the naturalistic fallacy (and claims such as Davidson's), then perhaps the current proposal could lead to a reassessment. That is, the proposal might be seen as revealing that naturalistic commitments don't in fact require that rejection after all.[4]

So far, the focus has been on the proposal's potential for meeting the first goal I mentioned, that of accounting for the naturalistic fallacy and specifying its strength. But it also helps with the second goal, that of determining where the naturalistic fallacy applies, or, put another way, where there is normativity of a kind that rules out

---

[2] I think it is difficult to argue for the thesis more directly. Evidence deriving from how normative discourse works in practice does not *rule out* assessment-sensitivity, and so leaves it as a live option, but it does not fully establish it either. I will discuss this further in the next section.

[3] Though, again, if one were to find *other* grounds for the assessment-sensitivity thesis, then one would turn out to be committed to these non-intuitive claims.

[4] I will not be able to address this issue directly here, but I plan to do so in future work.

naturalism. This can be seen from the following important example. If the proposal is correct, then the notion of *semantic correctness*, often thought to be the important normative element precluding naturalism about linguistic meaning and mental content, does *not* have that implication after all. This is because judgments concerning semantic correctness are not assessment-sensitive, so the naturalistic fallacy is not brought into play. This means the naturalistic fallacy cannot be used as a principled argument against naturalism with regard to semantic content, though of course there may be other reasons why naturalism is not viable in that context.

I will now explain the assessment-sensitivity thesis in more detail, before turning to a discussion of the explanatory benefits it offers.

I   Assessment-Sensitivity

Assessment-sensitivity is a semantic framework that John Mac-Farlane has developed in a series of papers (see his MacFarlane (2003, 2005, 2007, Forthcoming)), applying it to a number of theoretically troublesome kinds of assertion. The basic idea of assessment-sensitivity is this: the truth-value of a single claim, produced in a particular context, can shift as a result of relevant differences between those assessing the claim. MacFarlane argues that future contingents (e.g. "there will be a pop quiz tomorrow"), knowledge attributions (e.g. "Alfred knows that his car hasn't been stolen"), claims of personal taste (e.g. "Indian food is delicious"), and epistemic modals (e.g. "Alice might be at the library") are all assessment-sensitive. He attempts to show in each case that this interpretation provides the only satisfying account of various data concerning such claims — e.g. how we treat disagreements involving them, the circumstances under which they are retracted, etc.

In order to get a fix on this basic idea, let's look briefly at claims of personal taste. On MacFarlane's model, when someone says, "Indian food is delicious," they do not say something that is, as he puts it, "covertly about the speaker" (MacFarlane, 2007). In other words, they do not say something more or less equivalent to "Indian food tastes good *to me.*" This view — the one MacFarlane rejects — could be formulated as follows: a term like "delicious" applies truly to something only relative to a *standard of taste*, and a particular speaker's claim involving that term implicitly involves his or her standard. This amounts to a standard form of contextualism, since

the truth of the claims at issue depends on features of the context in which they are made (in this case, on a key feature of the speaker).

MacFarlane argues that this interpretation simply doesn't do justice to the disagreements that can arise concerning claims of this kind. After all, if we were to reply to the original utterance by saying, "no, Indian food is not delicious," we would take ourselves to be saying something incompatible with it, such that only one of the claims could be correct. But that isn't the case on the contextualist view sketched above; the two claims would be essentially unrelated, semantically speaking, since they were produced in different contexts (i.e. by different speakers), and they could quite easily both be true (if those speakers have different standards of taste). So this leaves two options. Either (a) our sense that the reply is incompatible with the original utterance is completely mistaken, or (b) predicates like "delicious" do not behave in the suggested contextualist manner.

MacFarlane's approach is a way of taking option (b), allowing us to hold onto our common-sense picture of disagreement and incompatibility. That said, the approach does adopt a key component of the contextualist view, for it accepts the idea that predicates like "delicious" apply truly to things only relative to particular standards of taste. The key difference is that it is not the original speaker's standard that matters to the evaluation of a claim as true or false, but that of the person(s) *assessing* the claim. So, "Indian food is delicious," as uttered by A, might be true as assessed by B (given his standard of taste) and false as assessed by C (given hers). Or it might be true as assessed by A at the time of utterance and false as assessed by A later on, if his or her standards have changed in the meantime. Importantly, though, if D says "Indian food is not delicious," then no matter who is doing the assessing (and when), only one of the two claims — A's or D's — could be assessed as true. In that way their incompatibility is preserved, and we can make sense of A and D as disagreeing with one another.

Some will surely think that the problem with this proposal is precisely what it shares with the original contextualist view: the contention that predicates of this kind apply truly to something only relative to a certain standard. The obvious alternative is to say simply that they apply truly or not — Indian food either is or is not

delicious — and that's the end of it. And this objectivist reaction[5] will be all the more common/compelling if the issue is predicates like "right," "good," and "rational," rather than "delicious."

I readily admit that the proponent of assessment-sensitivity is in a tricky spot here. MacFarlane, in a very brief discussion that opens his (2007), suggests that there are features of how we use terms like "delicious" that tell against an objectivist construal — such as that we make judgments involving those terms without the circumspection and deference that is characteristic of other, paradigmatically objective judgments — just as there are other aspects of use — e.g. how we go about disagreeing and issuing retractions — that provide evidence against contextualism. I'm not sure how convincing the anti-objectivist considerations he offers really are, though.[6] For one thing, it's pretty clear that circumspection and deference are not *entirely* lacking even when it comes to our judgments of taste; many people are more cautious issuing verdicts around so-called experts in such domains, often retreating to safer, overtly subjective judgments instead (e.g. "well, *I* like it, at least"). And to the extent that circumspection and deference *are* relatively lacking, this could simply be because it is less well-established who is authoritative about such matters, and we all provisionally take ourselves to be (though some of us must of course be wrong).

The underlying trouble is that if these terms really do behave in a non-objectivist manner, then that makes the original account of disagreement less compelling — or rather, it makes it less clear that disagreement is a datum that needs to be accounted for in the first place. Do we really take two claims to be incompatible if hearing one of them (even from many different people) leads to no hesitation whatsoever in uttering the other? Moreover, MacFarlane speculatively suggests that the point of assessment-sensitive discourse is to "foster controversy" as a means to the coordination of our various

---

[5]Though the terms are by no means perfect, I will throughout the paper use "objective" as a label for expressions whose semantics is more standard, i.e. neither contextualist nor assessment-sensitive, and "obectivist" as a label for views that take certain terms to be objective in this sense.

[6]I should note that MacFarlane himself in no way presents these considerations as fully decisive. His discussion is primarily meant to set the stage for his paper's main focus, which is the comparison of contextualism and assessment-sensitivity, and he simply leaves the objectivist aside with the suggestion that there is "more to say" against that option.

standards. But this too depends on disagreement being uncomfortable — otherwise there would be no pressure to coordinate — and non-objectivist behavior seemingly belies the presence of any real discomfort. The prospect of unpleasant disagreement would presumably make us at least a little bit cautious in our judgments.

It seems, then, that a viable picture of at least some assessment-sensitive expressions is this. First, the way they are *used* is compatible with both objectivism and assessment-sensitivity. This condition is not trivial, since use-facts plausibly rule out assessment-sensitivity by themselves in many cases (as with ordinary observational judgments, for example, concerning which our disagreements are relatively infrequent and more easily settled). But second, this similarity at the level of use is coupled with a fundamentally different underlying *semantics*. It is simply a fact about claims involving these expressions that their truth is relative to the background standards of those assessing them, even though that fact does not manifest itself in practice in any clear and decisive way.[7]

If that's right, though, then how does one establish that a term of this kind is assessment-sensitive rather than objective? As far as a general answer to this question goes, I can do no more here than provide a brief and rather sketchy suggestion; when one focuses on the particular case of normative expressions, however, a certain line of argument presents itself. I will spell that out shortly.

First, with regard to the general question, I think Harman (1985) nicely describes the underlying framework for an answer. Harman says there are two basic approaches we can take toward an area of inquiry.[8] On the one hand, we can engage in it from within, taking the conceptual resources it offers and the relevant beliefs we find ourselves with and attempting to achieve reflective equilibrium — maximal coherence, a set of general explanatory principles, etc. On the other hand, we can step outside that area of inquiry and ask how its basic conceptual ingredients fit into our overall picture of the world — this, Harman says, might result in reductionism, or

---

[7]Assessment-sensitivity need not always fail to show up in practice, though; a plausible case can be made that epistemic modals, for example, are used quite differently from standard objective expressions.

[8]Harman is primarily concerned with moral inquiry, but he mentions religious belief and our thinking about the mind as other areas toward which these two approaches can be taken. I take it that the point could be extended to various other areas as well, judgments of personal taste included.

nihilism, or non-reductionist ontological commitment, or it might call for a revisionary semantic proposal like expressivism or assessment-sensitivity.

For our purposes, the important thing to note is that this distinction between two approaches provides a way of understanding the split between use-facts and semantic-facts that I mentioned above. It is perfectly feasible, on this view, for the engaged use of a set of terms or concepts to exhibit objectivist characteristics, while the detached or external approach calls for a non-objectivist semantic treatment.[9] As such, establishing assessment-sensitivity will crucially depend on the detached or external approach, and will not come from an examination of our engaged practice alone. I'm inclined to think that, to the extent we find it plausible that terms like "delicious" are assessment-sensitive rather than objective, this indeed results not from differences in how those terms are used so much as differences in how they strike us from something like an external perspective.

To be truly informative, this framework would need to be worked out in greater detail, showing precisely what the external approach involves and how a particular semantic treatment can get established by means of it. My aim here, though, is to defend the specific thesis that assessment-sensitivity is the source of the naturalistic fallacy, that it is the feature of paradigmatically normative discourse that explains why it resists naturalization. And my argument for this thesis, though compatible with the Harman-style "two-approach" model, does not depend on it.[10] What it depends on is just this: when we reflect on the paradigmatic cases of normativity and ask how they fit into our overall picture of things, there is a reasonably strong intuition supporting the naturalistic fallacy and its anti-reductionist implications. As such, it is of clear theoretical benefit, other things being equal, if that intuition can be backed up and the fallacy accounted for.[11] My claim is simply that taking those norma-

---

[9]Harman thinks this is the case with moral terms and concepts, for instance.

[10]Nor does it depend in any way on the suggestion that some assessment-sensitive discourse is indistinguishable from objective discourse at the level of practice; my argument would simply supplement whatever case could be made on the grounds of differences in use.

[11]As I said before, if someone has *no* such intuition, no sense that there is a conceptual gap in the vicinity, and so sees no advantage in accounting for such a gap, other things equal, then this argument will have no force. But if they are

tive expressions to be assessment-sensitive does the best job of this, explaining the naturalistic fallacy (and the associated Open Question Argument) far more effectively than an objectivist construal can, while at the same time avoiding the semantic problems that beset the main noncognitivist alternatives.

In a sense, the argument for the main thesis boils down to this. We face two separate, apparently conflicting demands. On the one hand, the naturalistic fallacy, as a deep theoretical intuition, calls for explanation. On the other hand, we need to make sense of the behavior of normative claims. The first task tends to push one away from objectivism, but, as the example of noncognitivism shows, that can lead to trouble with the second. A return to objectivism makes the second task straightforward, but renders the first one completely baffling. The claim that the naturalistic fallacy stems from assessment-sensitivity is uniquely equipped to handle both of these tasks simultaneously.

I now turn to the positive theoretical explanations the assessment-sensitivity thesis offers.

II   Explaining the Naturalistic Fallacy

According to the naturalistic fallacy, there is a principled division between normative properties (and judgments involving them) and natural properties (and judgments involving them).[12] These two classes are autonomous with respect to each other in (at least one of) the following senses: (1) there are no property identities between members of the two classes; (2) no property of one class can be reductively explained in terms of properties of the other class; (3) the presence of a normative property cannot be inferred from the presence of natural ones alone.[13] Fortunately, I don't have to adjudicate

driven to deny the fallacy on the basis of general naturalistic commitments, then I hope my argument might be seen as allowing them to have their cake and eat it too. See footnote 4.

[12]These two classes are mutually exclusive by terminological stipulation. They need not be jointly exhaustive, though — supernatural properties, for instance, presumably fall into neither category. And while it is difficult to say precisely what it is that qualifies a property as natural, the ensuing discussion depends on no more than a rough sense of it. Of course, a large part of my aim in this paper is to go some way — though admittedly not all the way — toward saying what it is that qualifies a property as normative.

[13]Unlike the other two, this third characterization of autonomy is asymmetric. This is due to the plausibility of principles like "ought implies can," which would

between these different construals, since the assessment-sensitivity of normative properties would nicely account for all of them.[14]

Assessment-sensitive properties exhibit quite distinctive behavior, since the instantiation of such properties varies along with the standards (of the relevant kind) of those assessing the given situation. An object might have a property as assessed by one person, but lack the property as assessed by another, and neither person would have to be making a mistake (from his or her point of view). This is very much unlike most natural properties, which are objective in nature — their instantiation is a simple (though not always easily decidable) yes-or-no matter.[15]

It should be clear, then, why the assessment-sensitivity of normative properties would underwrite (1) above: two properties cannot be identical if they engage in such different behavior. Property identity requires sameness of extension at any possible world, but that requirement will not be met in these cases. With a standard natural property, a given object (at a given world) is either in the extension or it isn't; with an assessment-sensitive property, that same object (at the same world) might be in the extension on some assessments but not on others. Put another way, a standard natural property can be thought of as a function from worlds to extensions, but an assessment-sensitive property is a function from world/standard

---

allow certain descriptive conclusions to be reached on the basis of normative premises.

[14]I have shifted here from talk about normative *expressions* to talk about normative *properties* primarily because it is easier to put the metaphysical points currently at issue in the latter terms. I am simply working with a *deflationary* notion of property that basically allows there to be a property expressed by any grammatically appropriate expression. I don't believe anything crucial turns on this shift, however; the treatment of assessment-sensitivity is strictly parallel and the metaphysical points could be reformulated in terms of expressions, logical constructions in terms of other expressions, etc.

[15]It is important here not to confuse assessment-sensitive properties with very closely related objective ones. For instance, even if *delicious* is assessment-sensitive, *delicious according to so-and-so's standards* (or $delicious_{so-and-so}$) is not. A single object can be in the extension of *delicious* as assessed by A, and not in the extension of *delicious* as assessed by B — that's the distinctive behavior just mentioned. At the same time, the object will simply be in the extension of $delicious_A$, and not in that of $delicious_B$, with no qualification needed. But those are different properties. Properties of that kind might play a key role in a contextualist view, but the problems with that approach have already been discussed.

pairs to extensions. This difference ensures that a property of the first kind will be distinct from a property of the second.

However, that only shows that an assessment-sensitive property cannot be identified with any objective property. Presumably, most natural properties are objective, but there could be some properties that are both natural *and* assessment-sensitive.[16] If MacFarlane (2003) is right, for instance, claims about the future (if the future is undetermined) are assessment-sensitive, but those claims undoubtedly still concern fully natural facts. For that reason, establishing (1) requires the following supplementation: two assessment-sensitive properties *that depend on different aspects of the context of assessment* will also engage in differing behavior, precluding their identity.

This supplementation is exceedingly plausible — the fact that different aspects of the context of assessment are involved in determining the instantiation of two properties means there will always be room for slippage between the two. For any given object, and a given assessment, the object could be in the extension of one of the properties and not in the extension of the other. Once again, full coextension is guaranteed to fail. It is also exceedingly plausible that natural properties, if assessment-sensitive, will depend on the context of assessment in ways that are different from normative properties. It's difficult to see how a property would be considered natural if it depended on something akin to a standard of taste.

Thus, the assessment-sensitivity of normative properties explains (1). Moving on to (2), the situation is not quite so straightforward. The problem is that it's not entirely clear what is required for "reductive explanation" of a property. In most cases, this explanatory requirement is treated as *stronger* than the requirement of identity, permitting one to claim irreducibility even while (empirical) property identities are granted.[17] If that's right, then (2) simply follows from (1), so no further argument is necessary. But is there some way to see the explanatory requirement as instead a weaker one, such that a property could be reductively explained by others even though it is not identical to any of them (or, what is more likely, to all of them in conjunction)?

There is good reason to think this is not possible. After all, it seems to be a minimal requirement of property explanation that one

---

[16]This is why assessment-sensitivity is at best necessary for normativity, not sufficient.

[17]So-called non-reductive naturalists stake out this sort of position.

say (in other terms) when and where it is instantiated. There may, of course, be ways of doing this that do not suffice for reductive explanation — noting that *salt* is instantiated whenever NaCl is, for instance, in which one makes use of an already known empirical identity — but that's beside the point. What we're looking for is a way of explaining a property that doesn't entail (let alone presuppose) any identity claim at all. But if property explanation necessarily involves saying when and where it is instantiated, it looks like identity cannot be avoided. It seems that a property either cannot be explained at all — because it is primitive, perhaps — or it can be shown to be identical to some other cluster of properties.

What about functional definitions? As applied to mental properties, for instance, definitions of this kind are often taken to provide reductive explanations that can nonetheless be coupled with claims of non-identity — that's why functionalism is an alternative to identity theory. However, it is important to see that while such definitions can be coupled with *certain* claims of non-identity — generally between mental properties and physical ones — they cannot be coupled with the claims of non-identity that would matter in this context. This is because the mental property being functionally defined *is* taken to be identical to the cluster of properties that is doing the explaining. So, for instance, *believing that p* just is *being caused by a, b, and c* plus *causing d, e, and f* plus whatever else goes into the functional definition. Once again, reductive explanation brings identity along with it. And if so, then (2) follows directly from (1).

Still, there might remain a nagging suspicion that the very case at hand provides an example of explanation without identity. Let's look again at the property *delicious*, accepting for the sake of argument that it is in fact assessment-sensitive. Aren't we able to explain this assessment-sensitive property in terms of objective ones — e.g. *having such-and-such a standard of taste, delicious according to such-and-such a standard*, etc.?[18] I think we do in some sense make use of these other properties in giving an account of *delicious* and saying when and where it is instantiated. In a way, *delicious* is explained as being constructed from these other properties, each of which captures one portion — one assessment-context component — of the whole.

That said, I think the demand for reductive explanation ultimately cannot be met in this case. The reason is that reductive

---

[18]See footnote 8 for further discussion of this point.

explanation of a property has important implications for our judgments concerning it. In particular, our judgments concerning that property can be *directly guided* by our judgments concerning the other, explanatory properties. For instance, we can find out that something is a college freshman simply by finding out that it is a student and that it is in its first year in school. But things don't work that way with *delicious*. Even if we determine both (a) that a certain standard of taste is currently in play and (b) that a particular object is delicious according to that standard, that doesn't enable us to say that *delicious* is instantiated. It allows us to say that $delicious_{current\ standard}$ is instantiated, but that's not quite the same thing.

The basic problem here is that judgments concerning *delicious* are judgments made *from* contexts of assessment and *with* certain standards of taste, not judgments *about* contexts of assessment and standards of taste. The determination that can be made on the basis of (a) and (b) is one that anyone else could make too, from any other context of assessment. The property involved is, like those proposed by the contextualist, covertly about the judger, and so varies from one judgment to the next. And that means that the single assessment-sensitive property *delicious* hasn't been fully captured.

If an assessment-sensitive property were to be reductively explained, it would enable us to determine *from within any particular context of assessment* that the property was instantiated, simply by determining how things stand with other properties. Perhaps that can be done in some cases, but it won't be by means of objective properties or by means of assessment-sensitive properties that depend on the context of assessment differently — either way, judgments will potentially diverge as the context of assessment shifts. If normative properties are assessment-sensitive, then, (2) also immediately follows.

The case for (3) is much simpler, and builds on this last point. If one property can be inferred from another, then the two have to be of the same basic kind — either assessment-sensitive (and similarly dependent on context of assessment) or not. If they are of different basic kinds, then their differing behavior across contexts of assessment will break the inferential connection between them. It will not simply follow from the presence of one that the other is instantiated as well.

There are two brief points worth noting here. First, it is not that assessment-sensitive properties can be inferred *exclusively* from other assessment-sensitive properties; they can be inferred from a group of properties as long as just one member of that group is assessment-sensitive (and similarly dependent on context) as well. This is because that one assessment-sensitive property in the inference base will allow for the whole pattern to be maintained across contexts, the others in the group (presumably standard descriptive ones) serving as a constant background. So this fits nicely with the intuitive judgment that a normative conclusion is viable as long as there is at least one normative premise.

Second, the asymmetry of condition (3) also makes sense on this account. The instantiation of an assessment-sensitive property may well depend (and, I would say, usually does depend) on various objective properties being instantiated as well. Of course, as the context of assessment changes, the instantiation of the former depends on different sets of the latter, which is why the original property cannot be identical with or reductively explained by any particular set. Nonetheless, there might be some properties that are found in all such sets. This makes it possible that, for a given assessment-sensitive property, whenever it is instantiated — i.e. from whatever context of assessment — certain other properties are instantiated too. And this would allow one to infer from the assessment-sensitive property to one that is not (though not vice versa). That, again, fits with common intuitions about normative matters.

It can be seen, then, that the basic claims of the naturalistic fallacy are well accounted for by the thesis that normative properties (and judgments concerning them) are assessment-sensitive. I will end this section by quickly noting how the thesis bears on the two classic paradigms of normativity and the naturalistic fallacy.

Hume was one of the earliest to suggest the fallacy; he remarked that *prescriptive* judgments express a "new relation" — one involving *ought* or *ought not* — that cannot be deduced from straightforwardly natural properties and relations "which are entirely different from it." Moore made a similar claim about evaluative judgments, saying that they concern properties — like *good* and *bad* — that cannot be defined in terms of natural ones. What isn't made fully clear in either case is exactly how the one sort of property differs from the other, such that the attempted deduction or definition is bound

to fail. Assessment-sensitivity provides an answer that can apply to both cases. If the truth of these two sorts of judgment depends on certain standards that vary with contexts of assessment, then the claims made by Hume and Moore are backed up and accounted for. The operative standards might be the same in the two cases or they might not — that's a matter for further inquiry, regarding the relationship between evaluation and prescription — but the fact that both are assessment-sensitive in one way or another would be enough to underwrite the naturalistic fallacy as applying to them.

In the next section, I'll focus on Moore's famous argument for his conclusion, and show how it can be profitably understood in these terms.

III   Explaining the Open Question Argument

I should make clear at the outset that I will not be attempting to show that Moore's *actual* open question argument (OQA) should be thought of as involving, even implicitly, the notion of assessment-sensitivity. My claim is only that the notion of assessment-sensitivity can play an effective role in an argument *of that basic kind*. This updated version of the OQA establishes the same conclusion but in a much clearer way, and gets around a crucial limitation of the original.

Moore argued, very roughly, as follows. For any normative property, X, and natural property, Y, the question, "this thing is Y, but is it X?" remains open. In other words, the answer to such a question is never *trivially* "yes," as it would have to be if X were reducible to Y. (Consider the parallel question, "this thing is a first-year college student, but is it a freshman?") This shows, then, that there is a fundamental gap between the normative and the natural — the conclusion is the naturalistic fallacy.

This argument has struck many philosophers as being onto something important. Not only has the conclusion seemed basically right, but the key argumentative device — the open question test — has seemed like an effective way of reaching that conclusion, a way of revealing something crucial about the relationship between normative properties and natural ones. That said, there are two problems with the argument that are often pointed out.

First, there is the matter of the argument's scope. As Moore presents it, the argument only rules out *analytic* or *definitional* naturalism. That is, it shows that the *meaning* of a normative expression

or concept cannot be fully captured in non-normative terms. This is presumably all Moore himself wanted the argument to accomplish, and it is by no means a trivial result. However, it provides no reason at all for rejecting naturalism understood, as it often is nowadays, as a metaphysical claim about what properties there are — namely, as the claim that all properties are natural ones. In other words, it is perfectly possible, given Moore's argument, for normative properties to be identical with natural ones in the same way that salt is identical to NaCl; the OQA may show that the concepts are different in some way, but empirical investigation is still capable of showing that the two concepts pick out just one property, and a perfectly natural one at that.

Second, even when one focuses on the narrower construal of the argument, it is not clear how we know that the key question will *always* remain open. The claim of openness isn't meant to describe our reaction to naturalizing proposals made so far, but to project to all possible proposals. But how can it do that? If the argument is to do more than express a mere hunch, it needs to be supplemented with an explanation of what it is about normative properties that results in this guarantee of openness.

The assessment-sensitivity of normative properties supplies a solution to both of these problems. I'll begin with the second. Assessment-sensitive properties give rise to Moorean open questions because it is always possible for the context of assessment to shift. So even if an assessment-sensitive property is tightly linked with a certain natural property in the current context, the two might come apart if a shift were to occur. I contend that it is our implicit awareness of this possibility that accounts for why Moorean questions strike us as open in the relevant (i.e. projective) sense.

A few clarifications are in order here. First, it is important to note that I am not saying that assessment-sensitivity is the *only* way to account for Moorean open questions. Normative properties could be semantically just like straightforward natural properties — as objective as you like — and still give rise to them. It might simply be a brute fact, something we can give no explanation for, that we take normative properties to be irreducible, or there might be an explanation different from the one I'm offering. (I'll consider one popular proposal in the next section). My claim is just that assessment-sensitivity is one quite plausible way of accounting for

Moorean open questions, and that that tells in favor of the thesis that normative properties are assessment-sensitive. The strength of this consideration increases, of course, the bleaker the alternatives look.

Next, I want to say a bit more about the *implicit awareness* of assessment-sensitivity that I take to account for Moorean open questions. This might seem to be in tension with the discussion up to this point, in which I've emphasized the practical indistinguishability of assessment-sensitive and objective discourse. For notice that the mere *fact* of assessment-sensitivity, if not in any way recognized by us, might be able to explain metaphysical truths like those canvassed in the last section, but would be incapable of explaining the OQA. That argument relies on our *sense* that these questions remain open, and where would that sense come from if assessment-sensitivity lies beneath the surface? Indeed, one might even think that the picture I've sketched so far has the opposite implication, suggesting that someone caught up in a particular context of assessment will judge (or might easily judge) that questions of that kind are closed.

To see why there is no such tension, let's look again at the Moorean question: "this thing is Y, but is it X?" Yes, the latter clause is an engaged question, one which someone will answer from within a particular context of assessment and with a certain standard, as I've said before. The answer to the question may well be "yes," full stop, and if a different person says otherwise it will be treated as a disagreement. In other words, that question and its answer will be treated in a more-or-less objectivist manner. Moreover, a person's answer may well depend quite directly on whether various objective properties are instantiated, and may even depend in that way on property Y. (Incidentally, I actually take it to be an advantage of my view that it has all this in place to help make sense of the fact that some people believe, at least at certain times, that these questions are not open and that reduction is possible; that may be more difficult to explain on other views).

The important thing to see is that a definitive "yes" answer has nothing to do with whether the question is taken to be open or not. The key question, really, is a second-order one: "was that answer trivial or, in a way, automatic?" At that point, I think, we are no longer being asked to judge whether X applies, but to reflect on the relationship between X and Y, to step back and assess where our

original answer came from. Returning to the suggestion made by Harman, this second-order question requires (or at least inclines) us to take an external perspective rather than an engaged one.

It is at that point that an implicit awareness of assessment-sensitivity comes in. We recognize that, even if a thing is both X and Y, there is more to being X than just being Y. That additional factor, I'm suggesting, is the potentially shifting standard of evaluation on which judgments regarding X depend. And since that potential for shifting is ever present, we sense that Moorean questions cannot be closed simply by modifying or adding to Y; we sense that those questions will *always* remain open.

This point leads to one last clarification: it is awareness of assessment-sensitivity, and not of standard-relativity more generally, that accounts for the openness of Moorean questions. Standard-relativity, recall, is a feature of contextualism as well as assessment-sensitivity; the difference is whether the operative standards are determined by the context of application or the context of assessment. One might think that the real explanatory work in this and the previous section is done by the more general feature, not assessment-sensitivity specifically.

There is something to this thought. Dreier (1992), for example, argues that a contextualist treatment can make sense of what he calls "the ineliminability of the normative." However, taking normativity to be assessment-sensitive turns out to provide a far superior explanation of the OQA.[19] This is because, according to the contextualist, for any normative claim *made in a particular context* it is settled once and for all whether the normative property at issue is instantiated or not. As Dreier notes, once a standard of evaluation is fixed by the context of application, a normative assertion is equivalent to one involving perfectly natural properties. Dreier: "Remember that it is not the proposition expressed that is especially [normative], on the [contextualist] view, but the *way* it is expressed" (original emphasis).

This brings us back to the first criticism commonly leveled against the OQA, the one related to its scope. In effect, the OQA has a weak construal and a strong construal. On the weak construal, Moorean questions are open in this sense: someone who is fully competent with terms "X" and "Y" will feel that the answer to the question

---

[19]This is in addition to the other drawbacks of the contextualist view, such as the problem of disagreement mentioned toward the beginning of this paper.

is not trivially "yes." As already noted, if the OQA is understood this way it has no force against a strictly metaphysical naturalism, for which empirically discovered property identities are perfectly adequate. Contextualism can give a fine account of this version of the OQA. That is because full competence with "X" can be accompanied by lack of knowledge concerning a given context of use. If someone doesn't know enough about the context to be sure what standard of evaluation is operative, they will be unsure about the answer to the Moorean question as well. So X doesn't *immediately* or *trivially* follow from Y, since additional information is required to establish the link.[20]

Many people, though, have felt that the OQA has a broader reach, demonstrating that the search for *a posteriori* identities involving normative and natural properties (on the model of the identity between salt and NaCl) is just as misguided and hopeless as the search for naturalistic definitions.[21] This would require a stronger construal of the OQA on which Moorean questions are open in something like this sense: someone can be fully competent with terms X and Y *and know all the natural facts related to them* and still feel that the answer to the question is not trivially "yes."

Contextualism cannot underwrite this stronger version of the OQA, but assessment-sensitivity can. On the contextualist view, as long as we know all the natural facts pertinent to a given normative judgment - in particular, the facts about the context that establish a certain standard of evaluation - then we'll know which fully natural proposition the judgment expresses in that context. Given that knowledge, a Moorean question pairing the two will seem quite trivial indeed. (It would be like the question, "this is NaCl, but is it salt?" asked of someone with minimal knowledge of chemistry).

If normative properties and judgments are taken to be assessment-sensitive, on the other hand, then no matter how much we know about the context we're in and any relevant empirical correlations

---

[20]Rip Van Winkle examples provide the model here, as Dreier makes explicit. For instance, the question, "this happened on June 10, but did it happen yesterday?" is open on the weak construal because someone who has lost track of time would be unable to answer it, even if they are competent with all the terms involved.

[21]In any case, many have the sense that this search is hopeless, whether or not the OQA is the source, and it would be useful to have an explanation for that sense.

that have been discovered, a Moorean question will still strike us as having a non-trivial answer.[22] This is because we implicitly recognize that standards might change — *our* standards might change — such that the answer to the original question will, assessed from a new context, be different. We implicitly recognize that there is a factor involved in the normative judgment that is not involved in the natural one paired with it, and that factor is not simply additional information that would need to be grasped in order to forge a link between the two.

With that, I hope it is clear that the assessment-sensitivity of normative properties puts the OQA on solid footing — so solid, in fact, that the argument secures the most ambitious result that could be hoped for it. Again, that serves as a reason for accepting my main thesis, since it is of theoretical benefit to have this intuitive argument backed up and clarified. In the next section I will run through some of the advantages this approach has over its main competitors.

IV Two less satisfactory alternatives: noncognitivism and motivational internalism

Up to this point, I have tried to show how the assessment-sensitivity of normative properties and judgments can explain the naturalistic fallacy and the open question argument (in their strongest forms). The need for such explanation has been widely noticed, of course, and other proposals have been offered in an attempt to supply it. I take the two most popular options to be noncognitivism and motivational internalism, and I now want to make clear why the assessment-sensitivity thesis is preferable to them.

The naturalistic fallacy says that normativity cannot be identified with or reductively explained in term of purely natural properties. A noncognitivist approach accounts for this in a very simple way: normative judgments, despite superficial appearances, are not factual judgments *about* properties or relations at all, so there just isn't anything there to be identified with or reductively explained by something else.[23] Instead, according to noncognitivism, normative

---

[22]Again, this is best thought of as a fact about our second-order reflection on the question, not a description of our first-order judgment; if we're just interested in whether property X is instantiated or not, then all that knowledge of natural facts will lead us directly to an answer (within our context of assessment).

[23]This is, I think, the standard way to think of noncognitivism, but Gibbard (2002) argues that even a noncognitivist approach is compatible with the claim

judgments are declarations of approval/disapproval.[24] As it is some-times put, normative judgments do not express beliefs, like factual judgments do, but other, more desire-like mental states. This would nicely explain the OQA as well: since the first clause of a Moorean question expresses a factual belief, while the second asks how things are with some other, independent mental state(s), the answer to the question will never seem trivial.

A full discussion of noncognitivism is well beyond the scope of this paper. I will simply highlight two well-known criticisms of that approach and show that the thesis I am proposing is not vulnerable to them.

The first problem noncognitivism faces is one that came up be-fore in another context — making sense of disagreement. Because noncognitivism denies that normative judgments issue in truth-evalu-able propositions, it is unable to account for disagreement as a mat-ter of two judgments that cannot both be true. Attempts have been made to work out a different notion of incompatibility, but they have run into notorious difficulties.[25]

I won't assess these attempts or the objections that have been made to them; I will simply note that the assessment-sensitivity thesis allows one to take the easy way out on this issue. Normative judgments, if assessment-sensitive, are still truth-evaluable; the ques-tion of their truth-value is just slightly more complicated. But, as was discussed earlier in the paper, that added complication doesn't prevent one from giving a more-or-less straightforward account of disagreement in terms of truth-value incompatibility.[26]

Noncognitivism also confronts what is known as the embedding problem. If a normative claim is not a factual claim but, say, an expression of approval, then it is far from clear how it can be embed-ded in larger contexts, such as conditionals, in the same way that factual claims can. Embedding is standardly understood by taking the larger context to be truth-functional, i.e. a function from the truth-conditions of the (embedded) parts to the truth-conditions of

---

that a natural property is picked out by normative concepts. In that case, the view is much more similar to what I've called non-reductive naturalism in that it allows for *a posteriori* property identities.

[24]Or, in any case, some kind of act that is not truth-evaluable.

[25]See Unwin (1999) and Schroeder (Forthcoming), among many others, for discussion.

[26]See MacFarlane (2007) for a much fuller discussion.

the whole. That option obviously isn't available to the noncognitivist, since an embedded normative claim has no truth-conditions, and trying to see the larger context as a *different* kind of function is, to say the least, challenging. Consider an example where the antecedent is normative and the consequent is not. What does the conditional function yield — a (different) expression of approval? Could *modus ponens* be accounted for on that model?

Once again, I've done no more than sketch these worries. Still, accepting that normative judgments are assessment-sensitive (rather than noncognitive) permits one to avoid the worries entirely. Assessment-sensitive claims, because they express truth-conditional propositions, will embed just like any other factual claim. Assessment-sensitivity affects the precise nature of those truth-conditional propositions, but not the basic framework.

So here is the overall picture. It is commonly, intuitively recognized that normative judgments are *somehow* different from run-of-the-mill factual judgments. This pushes many toward some form of noncognitivism. But noncognitivism is a semantic proposal that is quite radically revisionary, and its implications seem to go too far. In other words, it can be objected to on the grounds that normative judgments are *not so* different from run-of-the-mill factual judgments. Assessment-sensitivity occupies the required middle ground. It is in important ways a less radical semantic revision than noncognitivism, and this allows it to account for the mostly ordinary behavior of normative claims while at the same time capturing a crucial element of difference. That element of difference doesn't show up much (or at all) in practice — where it would seemingly have to if noncognitivism were right — but it does show up when we reflect on underlying metaphysical questions, i.e. precisely when the naturalistic fallacy and the open question argument come to the fore. So assessment-sensitivity provides the explanations we need without the unwelcome implications.

The second major alternative is to use the thesis of motivational internalism as a basis for explanation. Motivational internalism maintains that normative judgments have a special, direct connection with motivation to perform actions. So, just to give one simple example, a judgment that something ought to be done is specially and directly connected with (some) motivation to do it. Now, a more precise formulation of this special connection is obviously needed —

at a minimum, it must not be mediated by contingent background motivations, since that would allow all judgments to have it — but at this point it is enough to have the basic idea.

Motivational internalism, as stated, is perfectly compatible with noncognitivism, so much so that it is often seen as evidence for it. This is because noncognitivism, since it treats normative judgments as expressions of desire-like states, accounts quite straightforwardly for the connection with motivation. However, since noncognitivism has already been addressed, I will be focusing on motivational internalism as an *alternative* proposal. That is, I will treat motivational internalism as the general claim above (regarding the special, direct connection) *combined with* the claim that normative judgments are, semantically speaking, objective. In effect, the thesis at issue holds that there are objective normative properties, and that judgments concerning them are inherently motivational.

This thesis, if correct, could explain the naturalistic fallacy in the following way: since judgments about normative properties are inherently motivational, while judgments about natural ones are not, reductive explanation is hopeless; there will always be a key feature of the former that cannot be derived from the latter. Similarly, the open question argument makes perfect sense: the additional motivational feature is relevant to answering the Moorean question, but it clearly goes beyond whatever is stated in the first clause of it; as a result, the answer to the question will never seem to follow trivially.

Two drawbacks immediately present themselves, though. First, and very briefly, many people take the core idea of motivational internalism (as I am interpreting it) to be highly dubious. Judgments concerning objective properties, it is thought, simply don't entail motivations; such judgments interact with our desires and goals, but they do not *produce* desires or goals by themselves. This is an aspect of Mackie's well-known argument that normative properties are too *queer* to be countenanced. At the very least, then, the thesis of motivational internalism is a contentious one. So if it can be shown that assessment-sensitivity does as good or (as I think) a better job explaining the naturalistic fallacy, there will be no pressure to get by, so to speak, with motivational internalism, simply because it is so clearly acceptable in any case.

The second drawback is relevant to this comparison of explanatory capacity: motivational internalism cannot be used to support

the stronger version of the naturalistic fallacy, that on which empirical property identities as well as reductive explanations are ruled out. The connection between judgment and motivation might show that there is something distinctive and irreducible about normative *concepts*, but they could pick out fully natural *properties* nonetheless. As is well known, two coreferential concepts can play very different cognitive roles; in this case one would play a motivational role in a way the other does not.

Setting that point aside, though, and sticking with the slightly weaker version of the fallacy, motivational internalism still has a pretty serious problem. This has to do with the task I mentioned a moment ago: specifying the nature of the special connection with motivation.[27] The strongest formulations of that connection conflict with basic psycho-behavioral facts. For instance, if the claim is that there is a *necessary* link between normative judgment and motivation, then instances of weakness of will should not be possible.[28] Since they are possible, that formulation cannot be right. Yet it's not clear how one can weaken the thesis without losing its ability to explain even the weaker version of the naturalistic fallacy.

One possibility is to suggest, as Dreier (1990) does, that there is *normally* a direct link between normative judgment and motivation. This is almost certainly true, but it seems to be true of some naturalistic judgments too. Take the judgment that some item will provide pleasure. That judgment will normally be connected with a motivation to pursue or acquire the item (though of course there will be exceptions).[29] How exactly can one distinguish this case from the normative one? If this case is said to depend on the contingent presence of a background motivation to pursue the pleasurable, the normative case could be said to depend on the contingent presence of a background motivation to pursue the good (or whatever). Notice that the stronger formulation, by contrast, licensed the claim that

---

[27]This discussion draws upon Dreier (1990) and Hattiangadi (2007).

[28]One might respond by saying that there is still *some* motivation on the losing side in cases of weakness of will; the point I am making would then have to rely on more extreme cases in which there is *no* motivation to do what aligns with one's normative judgment (e.g. "desiring the bad" from Stocker (1979)). These cases are certainly quite unusual, but they seem conceptually possible at the very least.

[29]Is this the same notion of *normal* at work? Perhaps not, but in that case the challenge is to say what is distinctive about the sense of "normal" used in the thesis.

we are *necessarily* motivated to pursue the good, something clearly not true of pursuing the pleasurable. But if we look merely at what we're *normally* motivated to pursue, all sorts of things appear to qualify.

Here's another try, due to Smith (1994), Wedgwood (2007), and others: it is necessarily the case that, *if one is rational*, there is a direct link between normative judgment and motivation. This seems to help with the counterexample I just offered, since it is plausible that one can lack the motivation to pursue the pleasurable without being irrational. But the same kind of problem can arise for this proposal. Take the judgment that doing Z will destroy the universe. Presumably, it is necessarily the case that, *if one is rational*, there is a direct link between this kind of judgment and motivation (in this case, one hopes, motivation *to avoid* the action). This kind of judgment may not offer much naturalizing potential, but what about the judgment that doing Z will minimize aggregate suffering? Is this judgment directly linked with motivation, assuming one is rational? The answer is not obvious; some people have certainly thought there is such a link, and simply denying that there is runs the risk of begging the question with regard to the naturalistic fallacy.

These problems exhibit a pattern. The "if one is rational" and "normally" clauses serve to cut down the number of cases to be examined, thereby excluding certain clear counterexamples (like weakness of will) to the strongest formulation of motivational internalism. But by doing so, they also open up the possibility that other judgments will meet the restricted condition. This works against the alleged explanation of the naturalistic fallacy (weaker version), since there is no longer any conceptual guarantee that normative judgments have a feature that naturalistic ones lack. So while these weaker construals of motivational internalism may be correct — the arguments here do not suggest otherwise — they cannot be used to perform the explanatory job that is of interest to us.

In conclusion, the thesis that normativity is assessment-sensitive has clear advantages over both noncognitivism and motivational internalism when it comes to explaining the naturalistic fallacy and the open question argument. Unlike noncognitivism, it allows us to see why there is a naturalistic fallacy while also allowing us to maintain various common-sense views about normative judgments: that they are truth-evaluable, that we can have real disagreements about them,

that they embed completely normally in larger contexts, etc. Unlike motivational internalism, there is no obscurity in how it accounts for the irreducibility of normative properties, and it is also capable of explaining, as motivational internalism is not, the strongest version of the naturalistic fallacy, which rules out *a posteriori* property identities.

At this point, I will shift gears slightly and discuss the bearing of the assessment-sensitivity thesis on a well-known argument in the philosophy of mind.

V Explaining Davidson's argument that there can be no psychophysical laws

Davidson (1970; 1974) claims that there can be no laws connecting phenomena understood in physical terms (hereafter: physical facts) with phenomena understood in intentional terms (hereafter: intentional facts). The reason he offers, very roughly, is that the attribution of intentional facts is governed by the ideal of rationality, while the attribution of physical facts is not. So, because these two domains of discourse involve "disparate commitments," the central claims and expressions of one cannot be neatly lined up with those of the other.

Though Davidson doesn't explicitly put this argument in terms of normativity and the naturalistic fallacy, it is exceedingly natural to interpret it along those lines.[30] Understood that way, the claim is that because rationality is a *normative* matter, and plays a constitutive role in the attribution of intentional facts, judgments concerning intentional facts cannot be derived from judgments concerning physical facts (which are presumably non-normative).[31] This argument seems to rely on something like formulation (3) of the naturalistic fallacy that was addressed above. As such, an explanation is needed

---

[30] And of course, many have done so, e.g. McDowell (1985) and McLaughlin (1985), among others.

[31] There is an added wrinkle here, common in the application of the naturalistic fallacy to the philosophy of mind, in that the judgments immediately at issue — e.g. the judgment that Sam believes Thailand is a monarchy — are not, *on the face of it*, normative judgments. Part of what is being argued is that these judgments do in fact concern normative matters, since they constitutively depend on other judgments that are clearly normative. In this case, it is judgments about rationality that are clearly normative, and judgments about intentional facts are, as it were, normative by association.

for why that claim is true, and assessment-sensitivity can do the trick.

I should note that the caveat applying to the discussion of Moore's open question argument applies here as well. That is, I am not claiming to give an interpretation of what *Davidson* meant by his argument, but to give a construal of that argument that is clear and effective. The basic problem is just that Davidson didn't say nearly enough, so there are important gaps in his argument that need to be filled.

For instance, Davidson says at one point that "there cannot be tight connections between the realms if each is to retain allegiance to its proper source of evidence." It is obvious, of course, that there cannot be tight connections between realms if each is to retain *sole* allegiance to its proper source of evidence, but why should they have to do that? One of the great things about establishing tight connections, after all, is that it allows a concept or set of concepts to *retain* allegiance to one source of evidence while *gaining* allegiance to another. It would seem, then, that there must be something distinctive about the proper source of evidence for intentional judgments, such that allegiance to that source precludes allegiance to any other (or at least to certain others). That, presumably, is what is meant by saying there are *disparate* commitments — the sources of evidence are not just different, but somehow irreconcilable. The clear next step is to say what the distinctive feature is that has this result.

Noting that the evidence involves normative considerations, or an ideal of rationality, isn't enough by itself. For consider a putative psychophysical law: intentional property I is instantiated if and only if neural property N is. Like all laws, this one is provisional; there is a persistent possibility of falsification — if, say, normative considerations imply I while the physical evidence says N isn't around.[32] But

---

[32]This seems to be enough to avoid the problems that Kim (1985) says psychophysical laws have, according to Davidson. Kim says that a law such as this, because it "would permit us to attribute intentional mental states independently of the rationality maximization rule," would thereby allow mental states "to escape the jurisdiction of the ruling constitutive principle of the mental." But this isn't so. We would constantly be checking on the law, using the rationality constraint to identify mental states in order to see if they line up with the law's predictions. Yes, if the law continues to work, we will be able to make predictions without consulting rationality constraints, but *we only know the law is working* by using those rationality constraints to check on it. So rationality remains

suppose that doesn't happen. Suppose that every time we get good physical evidence for neural property N, normative considerations simultaneously imply intentional property I, and vice versa. In this case, it is granted that there are two separate sources of evidence, that normativity is implicated in one of them, and that intentional judgments remain answerable to that normative evidence, but for all that a lawlike connection still seems possible. What guarantees that this is not, in fact, a law?[33]

The assessment-sensitivity of normative properties and judgments can account for this guarantee. If judgments about intentional matters are normative, and therefore assessment-sensitive, then the truth-value of those judgments (and the instantiation of the relevant intentional properties) will vary as different contexts of assessment establish different standards of evaluation. No proposed law involving objective physical properties can track this behavior, so there can be no laws connecting the two.[34]

Consider again the proposed law connecting intentional property I and neural property N. It has the right form, and, up to a certain point in time, it has no falsifying instances. So how do we know it cannot be a full-fledged law? Well, we implicitly recognize — again, from a reflective standpoint — that I is instantiated, at a given time and place, only relative to a context of assessment. So far, perhaps, we have occupied contexts of assessment from which it is true that I is present when and only when N is, but this correlation is guaranteed to fail for some context of assessment. That is precisely what it means for a judgment or property to be assessment-sensitive — it varies with context of assessment even as all other facts remain fixed.

---

an "essential constraint," and maintains "jurisdiction" over the intentional; the physical properties used in the law just don't flout that jurisdiction.

[33]One possible Davidsonian answer is interpretational indeterminacy. McLaughlin (1985) takes Davidson to rule this option out explicitly (p. 353), and Kim (1985) suggests that the main argument doesn't depend on it (p. 381). I am not so sure. If this is his answer, though, it is quite unclear how exactly it is supposed to work. I don't have time for a full discussion of this tricky issue here, so I will content myself with offering an alternative construal of the argument.

[34]As before, the same conclusion applies to physical properties that are not objective but assessment-sensitive, on the assumption that they will depend on different aspects of the contexts of assessment.

This explains why a proposed psychological law cannot *project* in the required way. It is not just that it might run into a falsifying instance, as is true with any law, but that such falsifying instances are in a certain sense ensured. It is possible, of course, that our actual context of assessment will never change in the relevant respect, and that the generalization will continue to be reliable. But that is not something we can depend on. One can think of it this way: physical laws, if correct, latch onto basic physical relationships that ensure there will be no exceptions.[35] But no putative psychophysical law can be correct in that sense; the built-in possibility of contextual shift means that exceptions are always possible, not just epistemically but metaphysically.

So we can now make sense of the commitments of the intentional and physical domains as "disparate." The former are assessment-sensitive, allowing for variance with context of assessment, while the latter are objective, precluding any such variance.

Two related objections could be made to the interpretation I have offered. First, the interpretation might seem to be incompatible with the fundamental Davidsonian claim that there are token identities between intentional and physical events. Second, it might seem to rule out the supervenience of intentional facts on physical facts, which Davidson also clearly endorses.

The response to the first objection is that token identities are fine, as long as we realize that the very notion of a token event, where that event involves assessment-sensitive properties, makes sense only if a particular context of assessment is fixed. In other words, while from one context of assessment there may be an event I_1, from another there will be no such event (because property I is not instantiated). There is no sense to be made of *the* token intentional event, independent of context of assessment.

As a result of this, claims concerning token identity are claims made from within a context of assessment. And those claims are completely viable on the view I've outlined. Notice Davidson's own reasoning on the matter: a given intentional event stands in causal relations, and since causal relations require a law-based physical underpinning, that intentional event must be identical with a physical

---

[35]This is tendentious as an interpretation of physical laws, of course, but I don't think the point I'm making really depends on it. It just allows the contrast to be drawn in a particularly clear way.

event. This begins with the assumption that there is an intentional event, so we are clearly starting from a certain context of assessment. So, if the rest of the argument works, the conclusion will follow, *keeping the context of assessment fixed*. It does not follow that a particular event, picked out objectively, that is identical with an intentional event from one context of assessment, must be identical with that intentional event from all contexts of assessment. But nothing Davidson says requires that to be so.

The issue of supervenience is somewhat trickier, precisely because it involves a determination relation going in the physical-to-intentional direction. As such, it seems to make a claim that should hold across contexts of assessment, which the assessment-sensitivity of intentionality would not allow. In the end, I think there is a version of supervenience that is ruled out on the assessment-sensitivity model, but it's one that Davidson, on closer inspection, cannot maintain either. At the same time, there is a weaker version that both can accept.

Davidson himself formulates the supervenience claim in at least a couple different ways. One says that "an object cannot alter in some mental respect without altering in some physical respect." This is perfectly compatible with the assessment-sensitivity of intentional properties. Since an alteration of a single object is at issue, the example concerns either a single ongoing context of assessment, or two temporally distinct contexts that are occupied by a single assessor. This is because a judgment concerning alteration is a single, unified judgment that the object was one way, then became another. Two assessors occupying distinct contexts of assessment might make different judgments about the intentional properties instantiated, but that would not amount to a judgment that an alteration had taken place.

Now, in the first case, i.e. that of a single ongoing context of assessment, the judgment that the intentional properties of an object have changed cannot be the result of a change in standards of evaluation. So something must have changed in what those standards are applied to. I've already noted that the instantiation of assessment-sensitive properties (from a given context of assessment) very likely depends on the instantiation of various objective properties, and it is exceedingly plausible in this case that those would be physical properties. If that is right, then the only explanation (within this

context) for the change in intentional properties would be a change in physical properties.

The second case is one in which a single assessor occupies two different contexts of assessment in succession. We could imagine someone gazing at an object, assessing its intentional properties, when suddenly their standards of evaluation change. Wouldn't that result in the judgment that the intentional properties had changed even though nothing physical had? No, it would not. When the standards of evaluation change, remember, the assessor interprets his or her previous assessment as *incorrect*. That is an important way in which assessment-sensitivity differs from contextualism. So the assessor would judge not that the intentional properties had changed, but that they had been one way all along. Supervenience is still preserved.

A second way Davidson formulates supervenience is to say that "there cannot be two events alike in all physical respects but differing in some mental respect." This seems to me to allow two readings. On the weaker one, it is implicit that a single, fixed context of assessment is at issue. From a single context of assessment, two physically identical events cannot be judged as having different intentional properties, since a single standard of evaluation is operative. On the stronger reading, however, the claim is that, given two physically identical events, there is *no sense* in which they can have different intentional properties. This clearly conflicts with the assessment-sensitivity thesis, which holds that the two physically identical events can have different intentional properties if they are assessed according to different standards.

This is a form of supervenience that does have to be given up on the model I am proposing. It is worth pointing out, though, that Davidson himself seemingly cannot accept it either, due to his commitment to interpretational indeterminacy. Consider this slight modification of one of Davidson's examples. There are two physically identical occurrences of someone uttering, "that's a shooting star." One observer's perfectly legitimate interpretation is that this person has the intentional property *believes that some stars are small and cold* (and that they mean "star" literally). Another observer's equally legitimate interpretation is that the person does not have the intentional property *believes that some stars are small and cold* (but

that they mean meteorite by "star").[36] It follows, then, that two physically identical events *can* have different intentional properties if they are assessed according to different interpretational schemes.

What Davidson can say, presumably, is that the physical facts fix the range of interpretational options open to one, or the interpretational landscape. Given two physically identical events, it will follow that both can be interpreted as having either intentional property I_a (according to interpretational scheme A), or intentional property I_b (according to interpretational scheme B), and so on. The assessment-sensitivity thesis is compatible with a similar kind of claim. That is, the physical facts can be seen as fixing the range of property/context pairs, or the property-instantiation landscape. Given two physically identical events, it will follow that both have intentional property I_x (as assessed from context X), intentional property I_y (as assessed from context Y), and so on. This amounts to a kind of supervenience that holds across contexts of assessment, though it falls short of the strong formulation above.

I have argued that taking normative properties to be assessment-sensitive provides a way of filling in Davidson's argument that there can be no psychophysical laws. It gives substance to the idea that the two domains involve commitments so disparate that they preclude even provisional *a posteriori* laws connecting them. This proposal can be objected to on the grounds that, whatever else may be true of it, it is not *Davidsonian* since it entails the rejection of closely related Davidsonian theses. I considered two forms of that objection, and found that the assessment-sensitivity thesis is, on the contrary, compatible with everything Davidson commits himself to on those topics.

In the next section, I will bring the paper to a close by considering two important implications of the thesis.

VI Implications

To this point, my aim has been to provide support for the thesis that assessment-sensitivity is a necessary feature of normativity. I have done so by showing that the acceptance of that thesis has substantial theoretical benefits. In particular, it validates and explains

---

[36]Davidson doesn't rest his indeterminacy claim on this particular example, of course; even if it is not convincing that both of these interpretations are equally legitimate, he says there will be such examples.

the intuitively compelling naturalistic fallacy and open question argument, and it does so more effectively and with fewer associated drawbacks than the main explanatory alternatives. It can also be applied in the philosophy of mind to validate and explain the Davidsonian claim that there can be no psychophysical laws. I now want to examine one way in which the thesis can be put to further use.

We can begin by remarking that a certain more direct defense of the thesis is not feasible. Namely, we cannot confirm the thesis by simply getting a list of all the normative properties and judgments and showing that they are all (plausibly) assessment-sensitive. This is because it is a matter of some dispute precisely which properties and judgments should go on that list in the first place. There are two paradigmatic examples of normativity — prescriptive judgments involving *ought* or *ought not*, and evaluative judgments involving *good* or *bad* — to which the naturalistic fallacy has historically been taken to apply, and it would indeed tell against my thesis if these judgments were clearly not assessment-sensitive. I don't think there is this disconfirming evidence, but I will not pursue the matter here. I will just assume that an assessment-sensitive construal of these paradigm cases is viable, and that it provides the best explanation for the *sui generis* character they seem to have.[37]

But if the thesis is accepted — on the basis of the explanatory benefits I've laid out combined with its plausibility as applied to the paradigm cases — it can then be put to use in determining which *other* properties and judgments are normative. If it is questionable whether or not a certain property or judgment is normative, the thesis can serve as a kind of test, offering one clear way to reach a negative verdict. That is, if the property or judgment in question is not assessment-sensitive, then it fails to meet a necessary requirement for being normative.

I believe this is precisely how things stand with regard to the property of *semantic correctness*. It is quite commonly claimed that semantic correctness is normative, and that, consequently, a natu-

---

[37]Similarly, in the discussion of Davidson's argument I assumed that *rationality* was indeed normative and that it could be plausibly construed as assessment-sensitive. If that turned out not to be true, the argument obviously wouldn't go through. My purpose in that discussion, though, was to show only that the argument could be made sense of in those terms, not that it had to be correct.

ralistic account of content is not possible.[38] I think, on the contrary, that semantic correctness fails the assessment-sensitivity test for normativity, so the proposed argument against content naturalism is unsound. This is not to say that a naturalistic account of content *is* possible, just that normativity cannot be the reason why it isn't.

Why is semantic correctness often thought to be normative? For one thing, it seems to involve a prescriptive dimension, in that a given concept, if it applies *correctly* to Xs, *ought* to be applied to an X and *ought not* to be applied to a non-X. Relatedly, it seems to involve an evaluative dimension as well, since correctness is clearly something positive, an achievement, while incorrectness is clearly something negative, a defect. Both of these claims are problematic as they stand and would need to be refined at the very least, but I will not address them further. I will instead move on and apply the test provided by the assessment-sensitivity thesis. This will show that semantic correctness cannot be normative.[39]

Consider first an ordinary, objective, naturalistic judgment, such as, "Earth's tallest mountain is less than 10,000 meters tall." The common view, which I've been assuming, is that this judgment is not assessment-sensitive, for its truth value does not and cannot vary as the context of assessment changes. Rather, the judgment expresses a proposition that is, in this instance, true (at this world), and its truth holds for anyone assessing it. Someone who (wrongly) disagrees with this judgment denies that very proposition. Given my framework, anyone who takes there to be a naturalistic fallacy is committed to this common view, since otherwise there would be no contrast between naturalistic judgments on the one hand and assessment-sensitive — in particular, normative — judgments on the other. The question is whether someone can accept this common view about judgments like this and not also accept it with regard to judgments about semantic correctness. I believe they cannot.

---

[38] This claim is generally taken back to Kripke (1982); it is endorsed by Boghossian (1989), Brandom (1994, 2001), McDowell (1984), and many others.

[39] It is perhaps worth issuing a reminder here that I am concerned in this paper only with a notion of normativity that carries with it anti-naturalist implications. There may be some other sense in which semantic correctness is normative, but if that sense entails no naturalistic fallacy it is beside the point. See Hattiangadi (2007) for a distinction along these lines between two senses of normativity.

Joshua Beattie (University of California, Berkeley)

Of course, some disagreements we find ourselves in arise due to differences in how the terms involved in the judgment are understood (or, equivalently, differences in the concepts taken to be applied). For example, someone might disagree with the above judgment because they take the concept *mountain* to apply to distinctively shaped masses of rock *plus* 1,000 vertical meters of the air above them (something like a nation's airspace).

This might give the superficial impression of assessment-sensitivity — can't both disputants be correct, given where they are coming from? — but it is important to see how this case is different, on the common view. On that view, either two separate propositions are expressed — one involving the ordinary concept *mountain*, the other involving an idiosyncratic variant — which is why they can both be true, or there is a single proposition at issue — perhaps because facts about social usage establish that the ordinary meaning is the correct one — in which case the person denying the judgment is plainly wrong. Either way, a disagreement due to this kind of meaning discrepancy is *not* a case of there being a single proposition that is true from some contexts of assessment and false from others.

How, then, do things stand with judgments concerning semantic correctness, such as the following: "the concept *mountain* applies correctly to that thing" (said while demonstrating Mount Everest, let's say)? Once again, we cannot take this judgment to be assessment-sensitive simply because disagreements might arise as a result of different ways of understanding the concept being judged. If there is a difference of this kind, then either two separate propositions are expressed by the two judgments, or there is a single proposition expressed and one of the disputants must be wrong. Either way, once the proposition is fixed, the default view looks to be correct: that proposition is simply true or not — either the thing demonstrated is in the extension of *mountain* (as it figures in that proposition) or it isn't — and no change in the context of assessment will affect that.[40]

The support for this default view goes beyond surface appearances too. When it comes to more plausibly assessment-sensitive judgments — we can return to our original example, "Indian food is delicious" — there seem to be good reasons for introducing *stan-*

_____

[40]The meaning of a term might change, of course, in which case later judgments of the same form could be assessed differently, but the truth value of any particular proposition remains always the same.

dards of evaluation that can vary without constituting changes in meaning of the kind we've been looking at.[41] But no such standards can be introduced with regard to semantic correctness, as a result of the very tight connection between correctness conditions and the meanings of terms. Any variance in standards of correctness will either (a) collapse into a difference in meaning at the ground level, which, as we've seen, means there is no assessment-sensitivity anywhere, or (b) push assessment-sensitivity down to the ground level as well, which, as I'll argue, leads to an unsustainable picture.

In some cases, meaning can remain constant even while correctness conditions (extensions) change in systematic ways. This is obviously true of classic indexicals like "I" and "here," and is perhaps true of assessment-sensitive terms like "delicious" and "good" as well. Standardly, though, a term's meaning and its correctness conditions go more closely together. And since the concept of *semantic correctness* is used in stating such meanings, a variance in standards of correctness will result in differing extensions, and therefore different meanings, for the terms involved. That just brings us back to the previous example of a disagreement about the claim, "the concept *mountain* applies correctly to that thing there." One cannot say that this expresses a single proposition, and that the involvement of *correctness* allows for different truth values when it is assessed from different contexts — there would no longer be any sense in which the initial noun phrase contributed the same thing to the proposition in both cases. Rather, as we saw, this case needs to be interpreted as involving two separate expressed propositions, each incorporating a slightly different concept. So here the differing standards of correctness seep down into different first-order meanings, and the assessment-sensitivity model fails to apply.

But perhaps there are shifting standards that would not have this consequence. The most likely candidate might be varying standards of *strictness*. So, to vary a classic example, "the concept *hexagonal* correctly applies to France" could be thought to be false as assessed from some contexts, which involve relatively strict standards of correctness, and true as assessed from others, in which standards are relatively lax. If this were right, though, the same situation would have

---

[41]Dreier, in his contextualist treatment of normativity, explains this in terms of Kaplan's character/content distinction, and something similar can be adduced here as well.

to apply to the judgment, "France is hexagonal." In other words, allowing varying standards of strictness with regard to *semantic correctness* requires that one allow varying standards of strictness with regard to the first-order concepts as well. In which case, if *semantic correctness* is assessment-sensitive, *all* concepts are.

So, in the end, really maintaining that semantic correctness is assessment-sensitive would require adopting a position that is, at the very least, quite revisionary. More importantly, from the current perspective, that position is incompatible with the acceptance of a naturalistic fallacy applying to semantic correctness, for it leaves no distinction between the two types of judgment — normative and naturalistic — that the fallacy presupposes.[42] This is a serious problem, since making use of the naturalistic fallacy was the whole point of arguing that semantic correctness is normative. The clear implication is that semantic correctness is not assessment-sensitive, and so not normative. The naturalistic fallacy simply does not apply as a result of that feature of terms and concepts.

I want to end by noting, very briefly, a second implication of the thesis that normativity necessarily implies assessment-sensitivity. As I've presented it, that thesis serves to underwrite the naturalistic fallacy — it helps to establish the claim that normative properties and judgments cannot be naturalized. However, by shifting the explanatory burden from normativity *per se* to one necessary (but not sufficient) aspect of it, the thesis in a way mitigates normativity's anti-naturalistic importance. A way of seeing this point is to note that *delicious*, for example, cannot be naturalized either, and for precisely the same reasons. In other words, though there is indeed something *special* about normativity that keeps it at arm's length from naturalistic explanation, it is not something particularly *unique*.

References

Boghossian, P., 1989. The rule-following considerations. Mind 98, 507–49.

---

[42]But couldn't the two types of judgment still be accounted for in terms of the different ways in which they depend on the context of assessment? Strictly speaking, yes. However, the way in which *semantic correctness* would depend on context of assessment seems to be the same as that in which "naturalistic" concepts would. So the distinction still could not be made to apply in the case at issue.

Brandom, R., 1994. Making it Explicit: Reasoning, Representing, and Discursive Commitment. Harvard University Press, Cambridge, Mass.

Brandom, R., 2001. Modality, normativity and intentionality. Philosophy and Phenomenological Research 63 (3), 611–23.

Davidson, D., 1970. Mental Events. In: Foster, Swanson (Eds.), Experience and Theory. Duckworth, reprinted in Davidson (1980).

Davidson, D., 1974. Psychology as Philosophy. Reprinted in Davidson (1980).

Davidson, D., 1980. Essays on Actions and Events. Clarendon Press Press, Oxford.

Dreier, J., 1990. Internalism and Speaker Relativism. Ethics 101 (1), 6–26.

Dreier, J., 1992. The Supervenience Argument against Moral Realism. Southern Journal of Philosophy 30 (3), 13–38.

Gibbard, A., 2002. Invoking Rationality to Explain Happenings. In: Bermudez, J., Millar, A. (Eds.), Reason and Nature: Essays in the Theory of Rationality. Clarendon Press, Oxford.

Harman, G., 1985. Is there a Single True Morality? In: Copp, D., Zimmerman, D. (Eds.), Morality, reason and truth: new essays on the foundations of ethics. Rowman & Allanheld, Totowa, NJ, reprinted in Harman (2000).

Harman, G., 2000. Explaining Value and Other Essays in Moral Philosophy. Clarendon Press, Oxford.

Hattiangadi, A., 2007. Oughts and Toughts. Oxford University Press, Oxford.

Hume, D., 1739/2000. A Treatise of Human Nature. Oxford University Press, Oxford.

Kim, J., 1985. Psychophysical Laws. In: LePore, E., McLaughlin, B. (Eds.), Actions and Events: Perspectives on the Philosophy of Donald Davidson. Blackwell, Oxford.

Kripke, S., 1982. Wittgenstein, Rules and Private Language. Harvard University Press, Cambridge, Mass.

MacFarlane, J., 2003. Future Contingents and Relative Truth. The Philosophical Quarterly 53, 321–36.

MacFarlane, J., 2005. The Assessment-sensitiviy of Knowledge Attributions. In: Gendler, S., Hawthorne, J. (Eds.), Oxford Studies in Epistemology 1. Oxford University Press, Oxford.

MacFarlane, J., 2007. Relativism and Disagreement. Philosophical Studies 132 (1), 17–31.

MacFarlane, J., Forthcoming. Epistemic Modals are Assessment-sensitive. In: Weatherson, B., Egan, A. (Eds.), Epistemic Modality. Oxford University Press, Oxford.

McDowell, J., 1984. Wittgenstein on Following a Rule. Synthese 58, 325–63.

McDowell, J., 1985. Functionalism and Anomalous Monism. In: LePore, E., McLaughlin, B. (Eds.), Actions and Events: Perspectives on the Philosophy of Donald Davidson. Blackwell, Oxford.

McLaughlin, B., 1985. Anomalous Monism and the Irreducibility of the Mental. In: LePore, E., McLaughlin, B. (Eds.), Actions and Events: Perspectives on the Philosophy of Donald Davidson. Blackwell, Oxford.

Moore, G. E., 1903. Principia Ethica. Cambridge University Press, Cambridge.

Rosen, G. A., 2001. Brandom on Normativity, Intentionality and Modality. Philosophy and Phenomenological Research 63 (3), 611–23.

Schroeder, M., Forthcoming. How Expressivists Can and Should Solve their Problem with Negation. Noûs.

Smith, M., 1994. The Moral Problem. Blackwell, Oxford.

Stocker, M., 1979. Desiring the Bad: An Essay in Moral Psychology. The Journal of Philosophy 76, 738–53.

Unwin, N., 1999. Quasi-realism, Negation, and the Frege-Geach Problem. The Philosophical Quarterly 49 (196), 337–52.

Wedgwood, R., 2007. The Nature of Normativity. Oxford University Press, Oxford.

The Sources of Epistemic Normativity

Anthony Booth (Utrecht University)
arbooth@fastmail.fm

I Introduction

If the normative question in ethics is why should we do what we believe to be the moral thing, then the normative question in epistemology is why should we act or believe according to what we think is the epistemic optimum. Why should we seek to have epistemically justified beliefs, for instance? This question is somewhat related to what is called the epistemic value problem, which is the question of why knowledge is better than mere true belief. However, the question I wish to address here is broader, since it is not just about why and whether knowledge is *more* valuable that mere true belief, but concerns why and whether knowledge (necessarily consisting of true belief) is valuable at all. As such it is closer to what Linda Zagzebski (2003) has called "the deeper value of knowledge problem": what is at issue here is why we should care about knowledge, that is, why should we do and believe things that will get us more knowledge. It is commonly claimed that the source of epistemic normativity is special and autonomous to other sources of normativity, such as moral normativity. Following this line of thought, one may claim that epistemic normativity is derived, perhaps, from our role *qua epistemic* agents, or from our valuing certain *epistemic* ends. Allied to, and supported by, this claim is the thesis that epistemic normativity is irreducible to other kinds of normativity. I shall call the view that the source of epistemic normativity is autonomous and that *thereby*

epistemic normativity is irreducible and *sui generis*, the Epistemic Independence thesis. In this paper, I seek to show that the epistemic independence thesis is false. In order to do this, I use Christine Korsgaard's (Korsgaard, 1996) taxonomy of views regarding what the possible sources of moral normativity are. These are: voluntarism, realism, the reflective endorsement view, and the appeal to autonomy. For each of these views about the source of moral normativity, I attempt to construct an analogue view concerning the source of epistemic normativity. I do so with a view to evaluating whether the epistemic independence principle would be viable under each picture.

II Voluntarism

According to the voluntarist, the source of normativity is some powerful authority who legislates over us. This authority can be thought of as a deity (God), a person (such as a King) or a group of people (such as an oligarchy). The thought here is that 'morality' is really just the legislative will of such an authority; a will that we are somehow bound to comply with. So, why 'be moral'? Because the will of our legislators commands it, and we have reason to comply with that command. Granting that it is, indeed, the will of our legislators to have us act morally, the view gives us an explanation as to why morality *feels* normative, since it is the *command* of someone who has authority over us (let's call the feeling that one is obliged to do something, normative force). But in order for the voluntarist to explain the actual normativity of morality, he needs to tell us why we should heed to the commands of these authorities, that is, explain why (and whether) we are bound to the will of these authorities. One, unsuccessful, way to answer that question is by stipulating that the authorities we are bound by are only ever *legitimate* authorities, such that we are bound to do what these authorities will, because their authority is legitimate. This is the move, according to Korsgaard, that belongs to Pufendorf. The move is clearly unsuccessful, however; relying as it does on an account of normativity that is different to that advocated by the voluntarist. The voluntarist can either give us a circular account of the legitimacy of the authorities in question, or else owes us an account of this legitimacy in other terms than those of his own theory. And then we'll wonder why this other account doesn't supersede voluntarism. So another strategy the voluntarist can take is to say that

we are bound to do as the authorities bid because of their power to punish us. This is the view that Korsgaard attributes (perhaps controversially) to Hobbes. But as Korsgaard points out, this strategy is hardly unproblematic either, given that not all immoral acts get punished. But I'm not here to criticise voluntarism, but rather put it on the table as a possible theory about the source of epistemic normativity.

Or rather, I want to see if the epistemic independence thesis could be true if voluntarism is the correct view about the sources of normativity. So what would have to be the case in order for voluntarism to be the correct view about the source of epistemic normativity? Well, if the voluntarist claim is that the source of normativity is derived from an authority (or certain authorities) who has legislative power over us, then in order for the epistemic independence thesis to hold, this authority will have to differ from the authority who is the source of moral normativity (or indeed any other sort of *sui generis* normativity). In other words, the voluntarist would have to claim that there is a special epistemic authority who legislates over epistemic matters only. The relevant candidates here might be a God of Epistemology, an Epistemological King or, more plausibly, a diverse group of cognitive authorities such as Eminent Professors, Doctors and Lawyers. It is the will and power of these authorities that is the source of epistemic normativity. For instance, that we ought to believe that the earth is not flat is thus explained by the existence an epistemological king's will to have us to believe it.

I think there are some rather obvious defects with this view. First of all, it is pretty metaphysically extravagant (and *ad hoc*, to boot) to postulate Gods and Kings of epistemology for the cheer reason of saving the epistemic independence thesis. Have we ever encountered and dealt with such entities? At the very least, we are owed some independent reasons for thinking such authorities might exist. But, of course we do (clearly!!) have reason to think that people such as Eminent Professors, Doctors and Lawyers exist. But now I think we need to explain why we must comply with these people's commands, since there does not seem to be anything, at least *prima facie*, that compels us to do so in a way that complying with the will of Gods and Kings does. It won't do to appeal to the *legitimacy* of these authorities because, first, as with the ethical authorities, appealing to their legitimacy is to appeal to an antecedently given

notion of normativity. But if the normativity underlying their legitimacy is more than merely epistemic, then we have compromised the epistemic independence thesis. Second, even if we could appeal to these authorities' legitimacy, that would not give us a good explanation as to why we *feel* compelled to do their bidding, even when we recognize that we ought to — in other words it would not explain normative *force*. Voluntarism in ethics does this via the power of the ruler that commands us. But what power do professors and lawyers have? Arguably considerable, when one takes them collectively; but isn't this power over how we lead our *practical* lives? In other words, what could such people do if I refuse to believe that the earth is not flat, or refuse to go about forming beliefs in a cognitively responsible manner? Nothing, I would say; not unless there is a practical end at stake (such as getting access to a particular conference, or having a court case turn out a certain way). However, if their power is only evident when practical ends are at stake, then we surely ought to question whether their power is purely epistemic, and can as such be the source of purely epistemic normativity. Now, this criticism might seem like it is the same one Korsgaard made against Hobbes, and as such is just a criticism against voluntarism altogether, not against voluntarism's ability to accommodate for an independent source of epistemic normativity. However, the criticism is slightly different: what I am questioning is not the putative epistemic authorities' power to deal with *every* case of epistemic wrongdoing, but rather their ability to deal with *any* case of purely epistemic wrongdoing. The epistemic voluntarist, I'm suggesting, thus has the problem of explaining what I've called normative *force* in a way that the ethical voluntarist does not.

A further problem with this view is that is would fail to give us an answer to the normative question. For the view allows there to be normative dilemmas between doing what our epistemic authority commands and what our moral authority commands. In order for the dilemma to be resolved we could appeal to a supreme authority, but this move would belie the claim that the source of epistemic normativity is autonomous. I am later going to develop this line of thought as a general strategy to show why the source of epistemic normativity cannot be independent, but for now let me deal with a possible rebuttal to this criticism. One may be minded (as is Mills (1998)) to appeal to evidentialism as a solution to these kinds of dilemmas. Or, rather, appeal to evidentialism as a means of show-

ing why these dilemmas are not genuine. Evidentialism is (among other things) the thesis that beliefs, and only beliefs, can be up for epistemic scrutiny, *i.e.* are eligible for epistemic justification. If evidentialism is true, then it seems like there can never be a choice between doing the epistemically right thing and doing the morally right thing, since epistemic justification and moral justification apply to different domains, namely belief and action. One need only consider epistemic matters when one is deliberating over what to believe and only moral or practical matters when one is deliberating over how to act; thus moral and epistemic concerns are never in competition with one another.

However, I don't think the appeal to evidentialism helps here; even if it were true, it wouldn't be of any help in dissolving these dilemmas. This is because even if there is never a choice between believing for either moral or epistemic reasons (or a choice between acting for moral or epistemic reasons), there can be cases where there is a mutually exclusive choice between believing that p at t and ϕ-ing at t, *i.e.* there can be a choice, at a particular time, between acting and believing. For instance, suppose I can't bring myself to ask someone out on a date if I believe that they will turn me down. The evidence, however, strongly indicates that the person I will ask will indeed turn me down. But I want to ask them out. This might, incidentally, lead one to think that I do have pragmatic reason (given my desire to ask this person out) not to have the belief in question. But we don't need to say this (*i.e.* deny evidentialism) for there to be a problem: that is, the epistemically right thing to do is believe that p, the pragmatic thing to do is to ϕ. If there is a choice between either believing or acting, then the choice (if evidentialism is true) is a choice between doing the epistemically right thing or the pragmatically right thing. Variations on the example will show that one can have a problematic choice between doing the epistemically right thing and doing the morally right thing. Evidentialism only solves the problem here if we think that one can never have a problematic choice as to whether believe or act. There is, however, no reason to suppose that this is the case.

III   Realism

I move now to discuss the realist approach to the normative question. According to the realist, the source of normativity is the world, or certain *facts* about the world. The belief that murder is wrong is

59

something (*qua* propositional attitude) that is prone to truth conditions. So, according to the realist, there are *moral* facts which act as truthmakers to propositions about ethics. For realism and the epistemic independence thesis to be both true, there have to be autonomous epistemic facts that act as truthmakers to propositions such as 'S is epistemically justified in believing that p'. So one might say, then, that the sources of epistemic normativity are epistemic facts, and the sources of moral normativity are moral facts. At first blush, it seems that this need only commit us to a benign pluralism about truthmakers (defended for instance by Jonathan Lowe (2006)). Contra Frege, it does seem plausible to say, for example, that whatever it is that makes a moral proposition true, is a different truthmaker to that which makes an epistemic proposition true, in the same way that what makes it true that Utrecht is a city in Holland is different to what makes it true that I am in this room right now. However, I think that the realist needs to be committed to a stronger sort of pluralism consisting in the claim that there is more than one reality. To see this, consider that most moral realists are also naturalists; that is, they think that moral facts are also natural facts (the sort of things investigated in the natural sciences), as opposed to non-natural facts with intrinsic 'to-be-pursuedness' built into them. However, if naturalism is true, then a benign pluralism about truthmakers can't support the epistemic independence thesis. This is because, according to naturalism, there is no discreet part of the world that makes true moral statements, rather moral facts are a continuation of, or at most supervene on, the natural world. So it is not as if (to use an analogy) it is like moral facts are Africa and epistemic facts South America; they are not two discreet parts of the world. But if this is true, then there will be situations where moral and epistemic truthmakers overlap. My seeing evidence that John committed murder might both make it true that I (epistemically) ought to believe that John committed murder and that I (morally) ought to report John to the police. Now, unless we want to deny that there is fundamentally only one reality, we will be forced to say that there is only one fact (or one piece of the world, depending on your theory of truthmakers) at play here. But the epistemic independence thesis needs there to be more here, or else the source of normativity would be a single one. What the epistemic independence thesis needs here is for there to be two contingently converging

60

realities which make the separate propositions true independently of each other.

To affirm that there is more than one fundamental reality is to deny what Jonathan Schaffer has called 'priority monism'. According to Schaffer there is good reason for a truthmaker theorist to be a priority monist, in that it helps deal with the problem of negative facts (*e.g.* is there are truthmaker for the proposition that blue monkeys don't exist?). In fact, as Ross Cameron (forthcoming) has pointed out, the truthmaker theorist *must* be a priority monist, since allowing there to be more than one fundamental reality entails a problematic commitment to necessary connections between contingent existents. Further, the claim that there is more than one fundamental reality smells of radical relativism in a deep sense, and as such, I would suggest, is *anathema* to any realist project, especially one that is allied to naturalism. But perhaps there is an interesting reply one could make here. For someone who holds the epistemic independence thesis is only committed to the claim that the source of epistemic normativity is autonomous, he does not thereby think that the source of moral normativity is too. He could then respond that the realist account of normativity shows us that there is only a source of epistemic normativity, and therefore that epistemic normativity is all the normativity there really is. Moral normativity is, in the end, just a kind of epistemic normativity. After all, that ethics is threatened to turn into epistemology is a common complaint against realist views in ethics. He could even say that a realist conception of the source of epistemic normativity solves this problem, since it gives us an account of why we should do the epistemically right thing.

Unfortunately, however, I think that if this is the best response available, then the realist has only enjoyed a somewhat pyrrhic victory. Despite its Platonic overtones, the thesis that moral normativity is reducible to epistemic normativity is highly counterintuitive. If it were true, then there would always be some epistemic reason to do the moral thing, but this is clearly not the case. For example, imagine that conducting research in eugenics would provide us with lots of very interesting and vital results, *i.e.* lots of important true beliefs. Clearly, conducting research in eugenics is unethical. But the immorality of eugenics is not derivable from epistemic considerations, since these considerations strongly pull is in the opposing direction. So, unless we want to eschew the intuition that eugenics

is wrong, then we have to maintain that whatever makes it wrong is not an epistemic matter. One could reply that conducting research in eugenics is *prima facie* epistemically wrong, but that this consideration is trumped by stronger epistemic reasons. However, it is hard to see in what sense it is even *prima facie* epistemically wrong to conduct research in eugenics, given that conducting it would have all these fruitful results. Further, even if we could make sense of this claim, it would still not explain why we think that the consideration that eugenics is unethical trumps any epistemological advantage we might think carrying out this research has.

There is perhaps a different way the realist could reply, however. The objection regarding pluralism about truthmakers, is only an objection to a naturalist realism. An option the realist has, then, is to follow G.E. Moore, reject naturalism, and accept the existence of such things as intrinsically normative entities. This move would get round the objection that there are times when one 'piece' of the world can act as a truthmaker for moral and epistemic propositions, since we could, after all, carve the world into intrinsically normative and non-intrinsically normative parts of the world, such that the two are different *parts* of the world. Further, in order to accommodate the epistemic independence thesis, we could say that the world is divided into intrinsically epistemically normative entities and intrinsically morally normative entities, such that the epistemic domain is like Africa and the moral domain is like South America. In other words, we can, after all, say that the parts of the world that act as truthmakers for epistemic propositions are different parts of that world than those that act as truthmakers for moral propositions. However, even if we are happy to accept the existence of such 'queer' things as intrinsically normative entities, the epistemic independence thesis would not thereby be adequately defended. This is because for the latter to hold, we need such things as intrinsically *morally* normative entities and intrinsically *epistemically* normative entities, but these things look like a contradiction in terms. If an entity is only normative from a certain point of view, or with respect to certain ends (here moral or epistemic ends), then I think we have counterfeited the right to call it intrinsically normative. If these entities are not intrinsically normative then we are back to the problems I've levied with regards to the realist naturalist's accommodation of the epistemic independence thesis.

IV   Reflective Endorsement/Appeal to Autonomy

In this section, I will discuss the reflective endorsement view and the appeal to autonomy as theories about the source of normativity. I will discuss the two together, since the latter is a form of the former. According to both theories, normativity comes from Human nature, and not, as the realist would have it, in our apprehension of truths about certain aspects of the world. According to the emotivists, for example, the sentence "eugenics is wrong" is just an expression of a human sentiment, rather than something that has propositional content and is subject to truth-conditions. If this is true then the normative question becomes one where we question whether we are right to have these sentiments. As Korsgaard puts it:

> Since morality is grounded in human sentiments, the norma-tive question cannot be whether its dictates are true. Instead, it is whether we have reason to be glad that we have such sentiments, and to allow ourselves to be governed by them. The question is whether morality is a good thing for us.[1]

There are a number of proposals on the table as to how we can go about answering the question of whether morality is a good thing for us. The first two that Korsgaard discusses belong to Hume and Bernard Williams respectively, the third is the appeal to autonomy she attributes to Kant. The Humean method is to show whether or not it is in our self-interest to be moral, Williams's method is to harp back to Aristotle and show whether or not being moral is conducive to 'human flourishing' or 'eudemonia'. The epistemic ana-logue to the question of whether morality is a good thing for us, I guess, is the question of whether doing the epistemically right thing is good for us. And there are analogous answers to the question available in the epistemology literature too; what is known as 'virtue epistemology', whose advocate is most prominently Ernest Sosa (*cf.* Sosa (2007)), also harps back to Aristotle. According to the virtue epistemologist, an epistemically justified belief is (roughly) one that has been acquired by an epistemically virtuous agent. The relevant normative question, however, then becomes: why behave like an epistemically virtuous agent? And unfortunately for the epistemic independence thesis, the answer to that question in the end is the

---

[1]Korsgaard (1996, p. 50).

answer to whether or not it is conducive to human flourishing. So even if there is, necessarily, an epistemic component to human flourishing, we cannot gauge its importance outside, or independently of, further non-epistemic considerations relevant to our being human. So the epistemic independence thesis is incompatible with Virtue Epistemology.

Further, a number of philosophers, such as Hilary Kornblith (1993), Larry Laudan (1990), and Peter Railton (1997), think that epistemic normativity is hypothetical or instrumental normativity.[2] In other words, they think that epistemic normativity is contingent on our desiring or valuing certain ends such as the acquisition of true beliefs and the avoiding of false ones. In so far as it looks like having our desires satisfied is compatible with self-interest, then the strategy resembles the Humean one. But, of course, sometimes it is in one's interest that one's desires are not satisfied. So what's to say that is in our interest to desire to have true beliefs? The hypothetical/instrumental view of epistemic normativity cannot, in principle, answer the question, since epistemic normativity is contingent on our having the relevant desires. So this view really turns out just to be a way of saying that the normative question in epistemology cannot be answered without recourse to another sort of normativity, that epistemic normativity is not *sui generis*.

In any case, Korsgaard objects to Hume's and Williams' method of answering the normative question in ethics, on the grounds that although they can establish that we have reason to be moral in general, they do not explain the normative force that doing the right thing has in particular situations. This is because though we might acknowledge that morality in general is good for us, there will be times when we know that doing an immoral act won't in the long run make any difference to my flourishing or self-interest (*i.e.* morality will survive this one immoral act) but, as Korsgaard puts it: "the normative question is one that arises in the heat of action. It is as agents that we must do what we are obligated to do, and it is as agents that we demand to know why."[3] We can raise the analogous objection in the epistemic case too: sometimes we know, for example, that not meticulously checking the references while reading a particular philosophy paper (by a famous philosopher, for example)

---

[2]See Kelly (2003) for some criticisms of this view.
[3]Korsgaard (1996, p. 91).

won't compromise a commitment to meticulously checking references in general. It is this line of thought, claims Korsgaard, that leads us to Kant and the appeal to autonomy.

For Kant thought that we can know whether any particular act of ours is right or wrong by knowing whether it can pass a particular kind of reflective endorsement test. For Kant this test is, of course, whether we have a motive for committing that particular act that can be willed as a universal law (the famous categorical imperative in the formula of the universal law). This is an appeal to autonomy because, according to Kant, the categorical imperative just *is* the law of a free will. The reasoning behind this idea is something like this: if the free will can cause action then it must have a law (this is something like the principle of the nomological character of causation), but if the will is to be free then it must be its own law. And, as Korsgaard puts it:

> The categorical imperative merely tells us to choose a law. Its only constraint on our choice is that it has the form of a law. And nothing determines what the law must be. *All it has to be is a law.* Therefore the categorical imperative is the law of a free will ... It describes what a free will must do in order to be what it is. It must choose a maxim it can regard as a law.[4]

Now, (I continue to follow Korsgaard's reading of Kant) the act of choosing a maxim is at once to identify oneself with (or give expression to) a particular identity. This is because when we are reflecting as to whether (for example) to act upon one of our desires, we do so by taking *ourselves* to be an arbiter, as it were, 'above' such desires, such that the maxim we choose to regulate over whether or not we commit the action is an expression of ourselves. So it is my autonomously defined identity as a father, for example, that obliges me to take care of my child. For if I fail to do so I loose a sense who I am and why my life is worth living. It is, in a deep sense, to kill myself. Generally speaking, we are obliged to be moral because we autonomously define our selves as moral agents, and so failing to abide by what being moral agents entails is, in essence, to stop being human.

Now, of course, nobody has a very tidy conception of his identity. Nobody is *just* a father, or just a moral agent; rather, we are a com-

---

[4]Korsgaard (1996, p. 98).

plex amalgam of infinitely different conceptions of ourselves. One such conception of our identity might be as inquirers or epistemic agents. So if we have an epistemic identity which is independent to our moral identity, then perhaps it seems that the source of epistemic normativity could be *sui generis* under this view. We could be minded to say, for instance, that moral normativity springs from our moral identity, and epistemic normativity springs from our epistemic identity. Since my epistemic identity is independent to all my other identities, then the source of epistemic normativity must be independent also. According to epistemologists Richard Feldman (2000) and David Owens (2000), it is, indeed, our identity or role *qua* epistemic agents that is the source of epistemic normativity. However, an appeal to an independent epistemic identity won't do the work to establish the epistemic independence principle if this Kantian approach is true. To see this, I think we need to have a look at some issues with the Kantian strategy.

The most obvious of these is the threat of relativism. For if there are many contingent conceptions of my identity I can identify with, then what's to stop me from choosing to identify myself as a murder, for example? Curiously, if I have identified myself as a murderer, not only are the crimes I commit that are constitutive of being a murder permissible, they are obligatory! So since whether or not I am obliged to be a good epistemic or moral agent is contingent on my identifying my self in a certain way, then if I can freely eschew that identity then I am no longer subject to its demands. Further, what if two conceptions of my identity are in competition with each other and are issuing me with mutually exclusive obligations, how do I decide what to do? But Kant, or at least Korsgaard's Kant, has an answer to this problem. The answer is to deny that our identity as moral agents is contingent. This means that all our contingent identities are somehow governed by this necessary identity, and thus that conflicts between them can be resolved via appeal to our fundamental moral identity. The necessity of our moral identity is established with a transcendental argument, which goes something like this: In order for us to have any reason to act at all we must have *some* conception of our identity, be it father, philosopher, friend or lover. But it is not the fact that we have defined ourselves in terms of one or other of these identities that makes this the case. Rather the reason why we must have some conception of our identity is because we are human

and as such are reflective animals who need reasons to act. Thus, in order for us to have reasons for action, we must value our humanity. Finally, to value oneself as a human being is to have a moral identity.

Arguably, this last step of the argument can be resisted. However, I won't do so now, since what I want to do is see if the resulting view is compatible with the epistemic independence principle. And it does not look like it is. First, it looks like we are going to need a transcendental argument like the above to establish that our epistemic identity is also necessary. For my present purposes, I am just going to assume that there is such an argument available. The reason that such an argument is needed is because without it our epistemic identity is merely contingent and as such governed by a necessary moral identity. As Korsgaard notes, the necessity of our moral identity does not 'trump' all of our other identities in the sense that we are not thereby moral beings *and nothing else*, so our epistemic identity can stay intact. However, the governing role that the necessary moral identity takes entails that the *normativity* of obligations that arise via our having an epistemic conception of ourselves is parasitic, and thus not independent, to our moral identity. It would mean that if our epistemic identity were ever to come into conflict with our moral identity, our moral identity would take over. This means that epistemic normativity could not be *sui generis*, since it would be evaluable in moral terms. So in order to block this objection, one needs to say that our epistemic identity is not regulated by our moral identity, since it too is necessary. But this too is going turn out problematic, for say that we could establish that our epistemic identity, like out moral identity, was necessary; what would happen when such identities come into conflict? For example, should a mother's moral obligations towards her son override her epistemic obligations when she is deliberating over whether to believe her son innocent of a crime in the face of evidence to the contrary? If both her moral identity and epistemic identity are necessary, then there is no way for her to decide what to do, since whatever she does will mean loosing a necessary part of herself, in other words she cannot avoid, in a sense, killing herself. One could reply that this just shows us that sometimes life is hard, but there is more going on in the objection. First, the thought that we can have two utterly diverse, incommensurable yet necessary identities, threatens the unity of the self, it threatens, to use Korsgaard's terminology in a different

context, to tear us apart. Second, to allow that we can be faced with genuine, irresolvable dilemmas is just another way of saying that the normative question cannot be answered. For when, in the heat of action, we are faced with such dilemmas we do ask ourselves the questions "why be moral?" or "why should I do the epistemically right thing?" Here it does not help me in the least that I can give a transcendental argument that establishes the necessity of my moral and epistemic identity. So it won't do to reply that the strategy is a way of answering the normative question, just not the type question such as why I should do the epistemically right thing *over* the morally right thing. This former question just is the normative question. For in the heat of action, we are never really asking the normative question in the abstract, rather, we want to know why we should spend our time and energy doing the epistemically right thing *rather* than something else. I think this point generalizes against any attempt to establish the epistemic independence thesis. Unless we want to say that epistemic normativity is primary and from which all other kinds of normativity are derivative, then we will have to admit the possibility of there being genuinely irresolvable dilemmas. This is because if epistemic normativity is *sui generis* then there can be no further non-epistemic standpoint from which we can commensurate the advantages of doing the epistemically right thing over the advantages of doing something else. This means that the normative question in epistemic terms or otherwise admits, in principle, of no answer. So to insist that there is a special sort of epistemic normativity is ultimately self-defeating; it entails denying that there is any such thing as normativity at all.

V   Conclusion

To conclude, then, I think we ought to say that the epistemic independence thesis is false. I have to admit that I have not quite established this, however. For there remain a few avenues the defender of the epistemic independence thesis can take. The first is to argue that though the sources of moral and epistemic normativity could not be structurally similar if we are to preserve their independence from each other, they could be structurally different such that the correct view concerning moral normativity is voluntarism, while the correct view concerning epistemic normativity is realism, for instance. The second is to claim that what I have presented is

vulnerable, like many such merely negative theses, to the rebuttal that I have not shown that there could not be a theory of normativity outside of Korsgaard's taxonomy. Further, it could be pointed out that perhaps some of the permutations inside Korgaard's taxonomy are immune to the arguments I have given. Modern meta-ethics is after all currently rife with finessed theories, from Cornell Realism, to Moral Fictionalism, to Norm Expressivism, so I need to show that all these theories are incompatible with the Epistemic Independence principle. To these charges I simply reply as follows: First of all, I think there is something strange in saying that there can be two different theories that best describe, respectively, the sources of epistemic and moral normativity. At the very least, the proponent of, say, realism about epistemic normativity would end up owing us an explanation as to why realism can *only* be a thesis about epistemic normativity. I don't see any *principled* (*i.e.* non *ad hoc*) reason why this should be the case. Second, even if there were such a reason, the ensuing position would still be vulnerable to, indeed would only exacerbate, the worry concerning dilemmas, since there would still, in principle, be no way of deciding between doing the epistemically right thing or the morally right thing. And, as I have argued, this is a way of saying that the normative question cannot be answered. Further, I take this to be a problem *any* theory of normativity will have to confront, be it a refinement of an existing view, or a new view altogether. I thus take myself to at least have identified certain constraints a theory of normativity would have to meet if it is to be consistent with the epistemic independence principle. I suggest that the probability that these constraints can be met is very low.

References

Feldman, R., 2000. The Ethics of Belief. Philosophy and Phenomenological Research 60 (3), 667–95.

Kelly, T., 2003. Epistemic Rationality as Instrumental Rationality: A Critique. Philosophy and Phenomenological Research 66 (3), 612–40.

Kornblith, H., 1993. Epistemic Normativity. Synthese 94, 357–76.

Korsgaard, C., 1996. The Sources of Normativity. Cambridge University Press, Cambridge.

Laudan, L., 1990. Normative Naturalism. Philosophy of Science 57, 44–59.

Lowe, J., 2006. The Four Category Ontology. Oxford University Press, Oxford.

Mills, E., 1998. The Unity of Justification. Philosophy and Phenomenological Research 58, 27–50.

Owens, D., 2000. Reason Without Freedom: The Problem of Epistemic Normativity. Routledge, London.

Railton, P., 1997. On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action. In: Cullity, G., Gaut, B. (Eds.), Ethics and Practical Reason. Clarendon Press, Oxford.

Sosa, E., 2007. A Virtue Epistemology: Apt Belief and Reflective Knowledge Volume 1. Oxford University Press, Oxford.

Zagzebski, L., 2003. The Search for the Source of Epistemic Good. Metaphilosophy 34 (1–2), 12–28.

Role of Normativity for the Explanation of Norm-Conformity

Catherine Herfeld (Witten/Herdecke University)

catherine.herfeld@uni-wh.de

I   Introduction

Nowadays it is widely accepted that social norms play a significant role in the explanation of cooperative behaviour and the maintenance of social order. People use social norms such as fairness, reciprocity, and promise-keeping to interact with each other (Henrich, 2000). While explaining human behaviour, economists have realized that the concept of social norms can contribute to the understanding of (economic) behaviour (Basu and Weibull, 2003; Conlin et al., 2003; Fehr et al., 1998; Kahneman et al., 1986; Lindbeck et al., 1999). Despite the problem that social norms are a) hard to measure and b) only vaguely understood in terms of their actual influence on people's behaviour, economists increasingly integrate social norms into their frameworks (Akerlof, 1982; Bicchieri, 2006; Conlin et al., 2003; Fehr et al., 1998). The framework I am mainly referring to here is the theory of rational choice (RCT).

When explaining action by using RCT, economists focus on observable behaviour. Such observed behaviours are then attributed to desires and beliefs as their causes, usually at the expense of other possible motivations. Models are based on the behavioural assumption of instrumental rationality and agency is determined by self-interest. If the agent is rational, his sole motive for action is assumed

to be personal welfare-improvement (Davis, 2004; Sen, 1985).[1] Consequently, the performed action is seen as a form of optimizing behaviour, taken to maximize individual utility and to satisfy personal preferences (Elster, 1989a). Transferring this conception unto the issue of social norms draws a picture of an internal motivating source for norm-conformity that is purely instrumental in character (Opp, 1999; Coleman, 1990).

Empirical results of ultimatum-, dictator-, or public-good-games show that economists cannot (always) predict cooperative behaviour. Instead, individuals will often follow social norms of cooperation at the expense of self-interest (Bicchieri, 2006; Fehr and Fischbacher, 2004). Thus, for the economist, sometimes the motivational force remains obscure. Now there are at least two ways of accounting for these 'behavioral anomalies': One way is to simply reduce them to irrational behaviour. This does not seem to be reasonable (Dawes and Thaler, 1988). If all actions are reduced to mere self-interest, even defined in a very broad sense (Opp, 1999), RCT becomes analytical. In this case, its status as a theory is generally questionable. A second way would be to refer to preferences for social norms as a solution for this explanatory dilemma. But, as claimed in this paper, to make social norms work, some additional motives are required to be in place. Different motives such as commitment play a crucial role in motivating norm-conformity itself and hence in explaining cooperation. This is because norm-conformity is often motivated by commitment. In particular, commitment motivates conformity to social norms of cooperation, *i.e.* to the norm of fairness, reciprocity and promise-keeping, whereby cooperative norms are used to interact with others. They enable or enhance cooperation among members of a given group or society (Henrich, 2000). Social norms are key elements in the provision of public goods. They prescribe solutions to social dilemma situations, *e.g.* to collective action problems that would otherwise exacerbate the provision and the maintenance of social order. If conformity to these norms is motivated by commitment, commitment is then a crucial motive in enabling cooperation.

---

[1]Welfare here is the satisfaction of preferences and thus understood in the neoclassical sense. This view is also called "preference satisfaction view" (Hausman and McPherson, 2006, p. 120). It implies that maximizing utility leads to personal welfare improvement, although both notions are not considered as the same and not synonymous from the philosopher's point of view.

To overcome the explanatory shortcomings of the behavioural foundations of economic theory by preventing RCT from analyticity and in order to find adequate explanations for norm-conformity, we have to distinguish between different types of motivation which must be investigated beyond self-interest. To differentiate these different types of motivation, it is useful to analyse the normativity involved in norm-conformity. This will create an understanding of the role normativity plays in the explanation of norm-conformity. To tackle the issue, this paper is concerned with finding an answer to the question 'what kind of normativity is involved in norm-conformity, *i.e.* why do people think that they ought to follow a social norm?'

II Social Norms, Commitment, and Rational Choice Theory

Norm-conformity is often explained by economic frameworks, such as RCT. However, such frameworks are mainly based on the assumption of self-interested agents. If a person is modelled to act in his or her self-interest, this entails that he or she is driven by a reason consisting of a desire to achieve a certain end and a belief, that a specific action will achieve this end. Being motivated in this way implies that the performed action is an instrument for reaching a further end and therefore excludes the view that an action can be performed for its own sake. Applying this view to norm-conformity, the assumption that an agent is motivated by self-interest implies that the agent's conformity serves solely as an instrument for him to achieve a certain further end. Holding a reason of obedience to a social norm because one aims at some further end, whereas the conformity would serve as a means to achieve that end, is what I will call holding an 'instrumental reason' for norm-conformity. The normativity involved in such norm-conformity, *i.e.* the reason why people think they ought to follow a social norm, is purely instrumental in character. In contrast, the motive of commitment provides people with reasons for obedience that are fundamentally different in nature. As I will argue in this paper, the motive of commitment provides practical reasons that are "independent of the gains and losses for the person in case he or she acts on that reason" (Pauer-Studer, 2007, p. 75). This is because these commitments are not based on the agent's desires (Schmid, 2007). They do not express what the agent *wants*. Rather, if a commitment presents a reason for action this is because commitments, as Searle puts it, create

desire-independent reasons for action (Searle, 2001). If we accept commitment to be a relevant motivation for people to follow social norms, the normativity involved in the resulting norm-conformity differs from that of the instrumentally-driven norm-conformity described above. This difference in the type of normativity involved in conformity then depends on the motivation of the people.[2] As I argue in this paper, this is why commitment as a motivational force cannot be analysed by a framework based on the idea of instrumental action and the motive of self-interest. If this is the case, classical economic theory cannot explain all types of norm-guided behaviour.

There are three ways to deal with the ensuing explanatory dilemma: Firstly, many economists simply dismiss the motive of commitment as an important motivating force for norm-conformity. But this is problematic, because commitment seems to offer a good explanation of different normative phenomena such as work motivation, solidarity bonds *etc.* (Pauer-Studer, 2007). And it provides a useful account of cooperation observed in experimental settings (Kerr et al., 1997).[3] A second possibility is to ignore the special nature of commitment and just assume that people have a preference for commitment; these approaches explain anomalies with as-if models based on long-term calculations and self-interest. This seems plausible at first sight. Economists have highly elaborated upon their concepts of self-interest and preferences thanks to the openness / flexibility of the utility-function. However, because all the motives covered by economic models are ultimately traced back to self-centred agency, which fundamentally contradicts with what is understood as an action motivated by commitment, this does not present a solution either. Additionally, defenders of this view are confronted with the objection of economic explanations then becoming merely analytic. Assuming a preference for commitment to explain committed behaviour seems to lead to a tautological explanation which provides very little information. The *explanandum*, *i.e.* the event or fact to be explained is at the same time used as part of the *explanans*, *i.e.* the

---

[2]Whether or not this leads to relativism of normativity will not be discussed here in greater detail but will be, at least indirectly, referred to later in the paper.

[3]I will not tackle the issue of providing empirical evidence for the existence of a motive of commitment in this paper. However, I am aware that this is a very delicate and important issue where a lot more work needs to be done. Interesting empirical literature is provided by Kerr et al. (1997).

mechanism which does the explanation. A third possibility open to economists is to take commitment seriously as a motivational force, to admit the explanatory shortcomings of the current framework and try to adapt it. In this case, the assumption of individual-welfare maximization and self-interested agency would have to be relaxed or even rejected (Sen, 1982). I am going to argue for this last possibility.

III   Four questions

To develop my argument I shall address the following four questions:

I. How does classical economic theory explain why people think they *ought* to follow a social norm, *i.e.* what makes individual norm-conformity normative from the economist's perspective?

II. What is the nature of this 'ought' or the type of normativity involved in norm-conformity and how does this 'ought' or type of normativity relate to motivation?

III. What is this 'ought' or type of normativity grounded upon? Does it differ between people/cultures and, if so, why?

IV. What role do motives and normativity play in explaining norm-conformity?

By addressing each of these questions in turn, I hope to show the methodological implications of analysing the normativity involved in norm-conformity. I shall leave the conceptual implications for later analyses.

I.

To address the first question of how classic economic theory explains why people think they *ought* to follow a social norm, one must first understand how actions in general and norm-conformity in particular are explained by economists and how they define social norms. Looking at the key features of social norms provides us with an understanding of the motivation people are assumed to have when conforming to social norms.

Why people act in the way they do is a difficult question. We cannot observe people's motivation for performing a certain action.

What goes on in people's heads is often taken to be inaccessible. People's minds are black boxes for the economist. However, a lot of empirical research has been done to find out why people conform to social norms. In general, the economist's explanation is based upon what he generally assumes to be characteristic for a motive of action. This view of motivation is based on the concept of intentional action used in folk psychology. An intentional action is a piece of behaviour which derives from an intention of the individual exhibiting it (Elster, 1994). This is what distinguishes action from mere behaviour, namely that it is done with a purpose or reason (*e.g.* Anscombe, 1963; Rosenberg, 1995). This means that every individual, when performing an action has a reason to do so.

Reasons are modelled as being provided by mental states, *i.e.* the desire to do something and the belief that a certain action serves as a means to satisfy this desire (Elster, 1994; Bicchieri, 2006; Hausmann, 2005). The motive for action is thus taken to be identified by providing a reason for an action. In this line of thought, the reason for an action represents the intention of the agent. If the person acts rationally, his intention finally motivates, *i.e.* causes, the action. Thus, by giving the reason for an action, the motivational structure of an action is seen as being causal in nature (Davidson, 1963). And the action of the agent is always assumed to be the intended action, based on what the agent desired and what he believed to be the means for fulfilling the desire.[4] This implies that the reason given by an intentional explanation is assumed to be the reason which ultimately motivated the agent to perform the action.[5] In this sense, this view supposes that an action is always done for a purpose; it is purely instrumental in nature.

In the economic language, the desires of the person are translated into a preference-set. The beliefs a person holds are taken to be the full set of information available to the agent. Given the end the agent aims to achieve, his preferences and information motivate him to perform an action. The action in question is seen as the

---

[4]I will not consider exceptions such as *akrasia* here.

[5]Note that it remains opaque and also uninteresting what the ultimate reason of a person is. The person might not be able to identify it and thinks he acted because of reason X but what really but unconsciously motivated the agent was reason Y. Or there might be a chain of reasons which jointly caused the action in question (Rosenberg, 1995).

best means to achieve his end (which is mostly the satisfaction of his personal preferences). Assuming this motivational structure, the action performed can be explained causally.

Preferences and beliefs are required to be internally and mutually consistent to call a choice instrumentally rational (Elster, 1994). However, consistency alone does not imply any kind of normativity of the agent's intentions in the sense that desires, the beliefs or the end pursued have valuable content for the agent or that the end is perceived as *right* in a more profound sense (Raz, 2005) — *right* in the sense that the agent believes or values it as a good or right thing, independent of what he is inclined to do or what would satisfy his preferences. There is also no space for the agent considering the action itself as worth pursuing, independently of his personal preferences.[6] What the agent evaluates or chooses as the right action goes back to the agent's ordering of his preference set. The agent orders his preferences on a scale. The measure is 'utility' and the ordering is represented by the utility offered by different (material or immaterial) goods. According to this order, the agent evaluates the outcomes of his actions; he chooses an action and acts accordingly. The judgement is assumed to be exclusively based on the criterion or principle of utility maximization. Because the outcome of the actions are ideally the fulfilment (or exceeding) of the actor's desires, the satisfaction of desires is assumed to always be the person's main end and the action is therefore said to be in his or her self-interest (Heap Hargreaves et al., 1992). In short, "desires [or preferences] are the unmoved movers" for human action in economic theory (Elster, 1989b, p. 4). The limited empirical basis for unobservable preferences or desires does not present a problem in this context. Instead, it is simply claimed that the observed action reveals the preference. And preferences reveal the motivations involved. In this understanding, different preferences (re-)present different motives; and preferences are given to economic theory.[7]

---

[6]To say that somebody prefers one state of affairs X to another Y "is to describe a mental state of that person which disposes her to choose actions which lead to X rather than actions which lead to Y" (Sugden, 2000).

[7]This relates to the idea of revealed preferences where it is said that the action taken reveals the preference which motivated the action (Samuelson, 1938; Sen, 1977). This is not necessarily true. Somebody who is going to church might be acting on purely selfish terms but reveals the same preference as a firm believer. However, there is still an explanation missing why some people are motivated

If one transfers this general picture of action-explanation outlined above to the case of social norms, conformity is seen as purely instrumental. An instrumentally rational agent has a prudential or — as I would call it — 'instrumental reason' to obey the norm because norm-conformity presents the best strategy in this particular situation, given the strategies of all other players involved.[8] Thus, the economist has a clear response to question I: norm-conformity is normative for an instrumentally rational agent because it is — in one way or another — in his self-interest to do so. In the following I will talk about 'instrumental normativity' in this context. Does this view of 'instrumental normativity' really fit the concept of what we think a social norm is or represents? Let us first clarify what is generally understood among economists to be a social norm.

Within various ontological views and explanatory accounts, there is no unified definition of a social norm used by social scientists (Hechter and Opp, 2001; Stout, 2001; Okruch, 1999). However, there are some key features or properties generally considered to be common to social norms. First, social norms are characterized as informal rules of behaviour "that people follow for some reason other than the fear of legal sanctions" (Stout, 2001, p. 6). This means that they are not enforced by a formal agency in contrast to formal institutions, *e.g.* legal norms. Instead, they are enforced by informal sanctions executed by third parties, for example through social forces such as the exclusion from a social group by its members (Mantzavinos, 2001).

The informal character of social norms leads to the second feature, their social aspect. They are social in the sense that they are assumed to exist in every society, every subgroup of society and any kind of organisational structure (Kirchgässner, 2006). In economic approaches, social norms are often assumed to be localized in people's preference set. The agent is modelled to have a preference for a social norm. This can be either by being (a) accepted by the members of a group or society or (b) through the internalization of the social norm by the individuals through the process of socialization

---

differently than others (Kurzban and Houser, 2005).

[8]Bicchieri defines a prudential reason for action as "if you have a goal X and the best available means to attain X is a course of action y, then you ought to adopt Y" (2006: 14). This is similar of having what I called 'instrumental reason' as an instrumentally rational agent.

(Hechter and Opp, 2001; Fehr and Fischbacher, 2004). In (a), the social norm is generally realized and consciously followed by the people, *i.e.* conformity involves choice on the part of the agent, based upon a preference for the end for which norm-conformity is the best means. Thus, social norms serve to as means for the individual to achieve or satisfy certain ends that *directly* affect its personal welfare. In case (b), the social norm can be followed unconsciously. Here, the social norm itself is considered to be the dominant preference of the agent. Thus, internalized norms *indirectly* affect the personal welfare of the individual through (positive or negative) emotions or external but informal (positive or negative) sanctioning.

There has been plenty of discussion about this 'narrow' behavioural basis of economic theory when using the 'preference-approach' to explain norm-conformity. The discussion is mainly centred on the assumption of the self-centred or selfish agent. It is claimed that many possible motives cannot be accounted for in this framework (Sen, 1985). As a possible solution economists invent a very broad notion of self-interest. The concept of preference has been extended to all kinds of tastes that people can have: material, immaterial, pro-social/altruistic/conditional, emotive, *etc.*[9] Thereby, self-interest becomes definitional to actions and social norms are also allowed to be motivated by altruism or other-regarding preferences. Even in these cases, conformity can be explained by classical economic theory as long as individual choices are made consistently with people's set of preferences (*e.g.* Bicchieri, 2006; Stout, 2001; Coleman, 1987; Elster, 1989a; Woodward, 2009). But what matters for the argument in this paper is that — in these explanations outlined above — all action based on preferences always goes back to personal welfare-improvement and self-interest. Why is that problematic?

---

[9]The economic framework is used to accommodate 1) all types of social concerns, altruism, emotions, and informal sanctioning (reputation, esteem, social disapproval, disesteem, *etc.*). It 2) accommodates completely unaware, automatic and blindly norm-conformity. It 3) can explain social norms as means to internalize externalities such as for example littering (*e.g.* Coleman, 1987). 4) Norm-conformity is considered as instrumentally rational when motivated by direct sanctions and indirectly by emotions. And rational choice theory accounts 5) for norm-guided behaviour based on the idea of conditional cooperation, thus when we believe that other people conform to social norms and also expect us to do so in a specific situation.

Let's take the case of an altruistic action. To explain altruism, the economist equips the agent with an altruistic preference. Altruism can be defined as "costly acts that confer economic benefits on other individuals" (Fehr and Fischbacher, 2003, p. 785). The agent acts altruistically because enhancing other people's welfare is what he is aiming at, *i.e.* his preference. His ultimate desires, *i.e.* preferences, are completely other-regarding (Bicchieri, 2006). But is this picture of the economist really what we understand to be altruistic? The agent prefers acting altruistically and caring about other people's welfare *because* his welfare depends on the outcome. His welfare is ultimately affected through present or future immaterial rewards. By acting altruistically, the agent improves his personal welfare to the extent of, for example, 'feeling good' or being socially esteemed. Altruism lies in his self-interest because of other than altruistic reasons, *i.e.* it is not his intention to be an altruist in the first place. It is simply the outcome of an action from which he benefits. This means that he does not care about the intrinsic value of altruism in the first place. He does not judge altruism itself as the right thing to do, as there is no judgement involved independent of personal welfare-improvement. It is only *because* he has a preference for altruism, which allows him to ultimately improve his personal welfare that he acts in the way that he does. Preferring altruistic behaviour is contingent and conditional on the outcome received by acting altruistically. But wouldn't we expect a Mother Theresa to be somehow unconditionally altruistic? Do we not think that, because she judges the intrinsic property of an altruistic action as being the right thing to do, that, if this judgement is based on her values, the performance of altruistic action becomes unconditionally necessary for her, *i.e.* independent of any rewards? And isn't this because she committed herself to certain religious values of a community which can be represented and implemented by the performance of an altruistic action? One could argue that an act may be utility-maximizing for a person because it satisfies the person's preferences yet at the same time reduces the personal welfare. Especially in the aforementioned example it seems hard to see how an unconditional and altruistic act benefits Mother Theresa, *i.e.* her level of personal welfare, in any real sense, although we could at the same time say that it satisfies her altruistic preference. Could we not simply say that the action in this case is consistent with utility-

maximization yet welfare-decreasing acts? This argument is valid when one considers welfare and utility-maximization to be fundamentally different. And we already made clear that we do not, as formal RCT equals them (Hausman and McPherson, 2006). The point I want to make here is that the concept of preferences is often misleading when it comes to explaining norm-conformity. Although it does not completely dismiss the normative dimension of agency, it gives a particular and overly narrow interpretation of this dimension. The capacity of a person to accept obligations and duties as motivating, to undertake judgements about right or wrong, good or bad, independent from immediate and self-centred desires, the capacity to base or adjust one's actions on these judgements and the capacity of self-determination/self-formation and self-control cannot be accounted for when taking the instrumental view of normativity that RCT implies.

As we have seen, when reducing norm-conformity to actions that are purely instrumental in character, the normativity involved in such actions is itself instrumental when the improvement of personal welfare is always the ultimate aim. This view contradicts the normative dimension of agency.[10] But this normative dimension seems to matter in the context of norm-conformity. That norms are often seen as prescribing how we ought to behave in certain situations seems to be left out by the economists view. This leads to the question II, which I want to address, namely what is the nature of this 'ought' or the type of normativity involved in norm-conformity?

II.

The underlying problem of explaining norm-conformity with tools from economics is that norm-conformity seems to involve an instrumental *and* a normative aspect. A quick look over the literature allows us to pick out overlapping notions such as 'expectations', 'obligatory', 'value-judgements', 'right', or 'wrong' and 'authoritative' when it comes to defining social norms (Anderson, 2000; Bicchieri, 2006; Coleman, 1987; Elster, 1989a; Hechter and Opp, 2001). These normative notions suggest a type of normativity attached to

---

[10]An exception is the position of ethical egoism, *i.e.* the normative ethical position that moral agents ought to only do what is in their own self-interest, can be considered as normative dimension of agency.

norm-conformity that is different from the consistency-requirement postulated by RC-models. Social norms are not regarded as mere regularities in people's behaviour which can be observed.[11] Neither are they considered to be completely contingent in the sense that people's conformity varies fundamentally depending on their expected utilities in each situation.[12] Instead, social norms can be seen as more or less stable "*prescriptions* that establish how one *ought* to behave" within a community (Woodward, 2009, p. 33, my italics). They are defined as "standards of behaviour that are based on widely shared beliefs how individual group members *ought* to behave in a given situation" within the community (Fehr and Fischbacher, 2004, p. 185, my italics). They are said to "reinforce certain patterns of behaviour ... by representing these patterns as desirable or *obligatory*" within a community (Pettit, 1990, p. 725, my italics). In this sense, a social norm is a rule prescribing or proscribing behaviour. — It presents a statement of the form of a conditional, hypothetical imperative "in situation X, you *ought* to do Y" (they are conditional only on the circumstances and not, as instrumental rationality assumes, exclusively on outcomes) or of an unconditional, categorical imperative 'you should do X' (Elster, 1989a, p. 98, my italics).

People accept the social norm as a standard of evaluation for the behaviour on a collective level. Thus, these standards of right and wrong, *i.e.* the social norms themselves, have to be at least 'passively' approved of, if not actively ratified, enforced and followed as well as implemented by the individual.[13] The important point here is that such a collective acceptance requires a shared understanding of what is regarded as socially appropriate/inappropriate among the

---

[11]Bicchieri contrast social norms with descriptive norms such as fashions and fads, whereas the latter are mere regularities which can be observed. According to Bicchieri and in line with my argument, what classifies collective behaviour as a descriptive or a social norm is a) the motives of the people involved and b) the expectations of the people within the community (Bicchieri, 2006).

[12]There is a lot of empirical evidence that people's conformity varies from situation to situation but generally, social norms are often observed to be stable over time.

[13]'Passively approved of' in this context means, that the existence of a norm is accepted even if people avoid situations where the conformity to a social norm would have been expected, like the example of the Ik people shows (Turnbull, 1972). Thus, even if people do not follow the social norm, its existence is common knowledge and thus not denied and can therefore be assumed to be passively accepted.

individuals within a community (Krupka and Weber, 2008). But is it not that such a shared understanding of the required behaviour needs to have a basis for agreement, something like a standard, a principle or rationale on which an evaluation can be based? Such a basis is normally provided by certain fundamental evaluative principles or values of which the individual approves. Because the individual approves of these values and principles, he uses them to evaluate and judge actions as the right/wrong thing to do and finally perform the action. I argue that these standards and principles shape the individual's motives; they become part of what motivates the individual in performing a certain action. And this is why motives matter when it comes to explanations of norm-conformity. But how exactly do they shape the agents motives or even provide him with a motive for action?

We have seen so far that social norms can be either defined as means to personal welfare-improvement; or they can be defined as standards for the evaluation of behaviour within a community. Both views require an evaluation of alternative courses of actions that is based on an evaluative standard for deciding which course to take. But in the two cases the individual is assumed to base his choices on different evaluative principles. In the first case, the evaluation is assumed to be based on the utility-principle. The individual only takes into account the satisfaction of his personal preferences as the main aim. In the second case, the individual bases his evaluation on a standard which is external to the agent, which is set or prescribed by the community. In this case, the individual *and* the social perspective matter for the agent's choice of a course of action. Besides taking his personal preferences into account, the individual has to approve or at least accept this evaluative standard or principle of the community. It requires a shared understanding of what is considered as right and wrong among the individuals within a community. One could argue that such a shared understanding is not necessary and that norm-enforcement is generally secured by informal sanctions (Bicchieri, 2006; Coleman, 1987; Mantzavinos, 2001). And it is true that, as long as we do not live in a dictatorship, punishment admittedly exists to align people's behaviour to what the community considers as right or wrong. But it cannot uphold permanent and constant norm-conformity without any approval at the individual level, or at least from a group of reasonable size within the society.

Rather, people approve/disapprove of something which they value and judge as right or good, something which is in line with what they are convinced of. Therefore, shared values matter to enforce social norms.

Let's come back to the question of how people's principles and standards of evaluation relate to their motives for norm-conformity. The gap between standard or evaluation for judging an action as right or wrong, the agents' motivation to perform an action he considers as the right thing to do, and the actual performance of an action can be closed by introducing the idea of commitment. Before doing so, it is important to note that especially social norms of co-operation are operative by supporting value-laden patterns such as promise-keeping and truth-telling over a period of time. Individuals and groups transform their principles and value-orientation into their individual and collective actions. Social norms such as fairness, reciprocity, and promise-keeping align people's behaviours with their values and principles. Norm-conformity is a means for an agent to implement the values and principles of that individual and/or at the same time of the group or society that he is part of. Social norms can therefore fill a gap between values and principles on the one hand and conduct on the other (Wallace et al., 2004). Values and principles are at the same time constitutive for the self-conception of groups, individuals, and their self-understanding as autonomous agents. Thus, they play a role in supporting certain virtues (Elster, 1989a).

This is important because it shows the link between people's judgements and their motivation. In the case of social norms, people's conformity to them can be one way of revealing what the individual considers as being the right thing to do in a situation. In this case, the individual approves of what is seen to be right in the society. But, beyond mere approval, the individual can identify with certain values and principles that are embodied in social norms. The identification with values and principles generates commitment to them for groups and their members. In this case, one could say that people feel committed to these values and principles. And people think they ought to conform to a social norm because these represent their values and principles with which they identify apart from the fact that they are also considered to be right by the rest of the society or the members of the group. And because the individual has committed

to these values and principles, he himself judges the conformity to a social norm as an action which is the right thing to do. Conforming to them makes it possible to reveal the agent's commitments. It represents his values and principles and thereby allows for living up to his standards. Thus, the social norm itself becomes a standard of evaluation that is based upon and is consistent with what we generally judge as being good or bad, right or wrong. Thus, people's norm-conformity is motivated by commitment to these values and principles in the first place. Commitment provides the agent with a reason for norm-obedience; this is how his commitment motivates him to conform to a norm. But this reason is not instrumental in nature, because the agent's judgement in these cases is not based on the principle of instrumental rationality. The agent's performance of the action, *i.e.* his conformity to the social norm is independent of his personal situation; it is independent from his self-interest and his preferences. The appropriate motivation which captures this idea is his commitment as one possible motive for conformity to social norms. And the type of 'normativity of commitment' involved in this kind of norm-conformity is not instrumental in character.

Taking up on this idea of a social norm being a standard of evaluation for approving behaviour, the understanding of why people conform to social norms differs significantly from the notions of agency in classical economics. And this is why the understanding of the normativity or the nature of the 'ought' involved differs too. Instead of granting the agent all types of preferences as explanation for his action to find out the motivation behind norm-conformity, I argue that it is instead rather useful to analyse the normativity involved in what motivates agents to conform to social norms. I claim that we have to distinguish between two forms of normativity involved in norm-conformity: one is what I call 'instrumental normativity', which is well captured by the economic framework and the notion of preferences. The second is a kind of 'normativity of commitment', which, so the argument goes, cannot be captured by economic theory. Depending on people's motivation, they consider conformity as being normative in either one or both of the two senses at the same time. This depends on the standard of evaluation they use. If they evaluate norm-conformity in terms of appropriateness or inappropriateness, they take the shared values as a basis for their evaluation. In this case, to follow a social norm is normative for a

person, *i.e.* the person recognizes and accepts the normative claim that that social norm makes on him/her, because the norm embodies the principles and values the individual or group is committed to. And the "acceptance of a certain normative consideration can mould the way [one] thinks" (Millar, 09.05.2007, p. 1) and acts. What motivates people is indirectly the recognition of the claim the norm makes on them and directly their commitment to the principles or values embodied in the norm. Against the economists' view, personal welfare-improvements are then not the reason for the choice of the agent; they are not what the agent aims at in the first place and thus are not what makes norm-conformity normative for them (Sen, 1977; Hausmann, 2005). In the case of norm-conformity as being an action motivated by commitment, the agent considers conformity as normative for him because of his commitment to what the social norm embodies, *i.e.* what is seen as right or wrong — and his commitment to values and principles which provides the basis for evaluation of an action to be right or wrong makes the performance of the action for him necessary, *i.e.* normative for him.

Why does this whole picture present a problem for the economist? To see this, we have to understand in detail what characterizes the motive of commitment and what it implies. The interesting issue about commitment is that it seems to matter to economists as well as to philosophers and thus perfectly represents the potential intersection of both disciplines. Whilst there is a lot of philosophical discussion going on about the nature of commitment, what it means for a person to be committed to something/someone and what commitment aims at, there are only very few economists who take the importance of commitment as a motive for agency seriously. One of these is Amartya Sen. In his article *Rational Fools: A Critique of the Behavioral Foundations of Economic Theory* published in 1977, Sen famously argues against the usefulness of the economist principle of self-interest as part of the assumption structure in economic models. He claims that not all motives for action can be reduced to ultimate self-interest. He remarks that the problem with economic theory is that "there is no choice-independent way of understanding someone's attitude towards alternatives" (Sen, 1977, p. 323). But, he goes on to say, there is one source of action which does not involve choice in the usual sense. That source of action is the motive of commitment.

What does it mean to say that commitment is choice-independent? This means that, when motivated by commitment, an agent takes an action even though the consequences of the action will not necessarily affect his personal welfare. So, there is no choice involved in the usual sense. The person would still perform the action, even if he believes that it will yield to a negative effect on his personal welfare; he would perform it even if, at the same time, he could perform an alternative action that would make him better off. Thus, taking commitment as a possible motive for action, the identity of personal welfare and individual preferences resulting in choice no longer holds (Sen, 1977).[14] In the case of commitment, personal welfare and personal choice are separated from one another. As we shall see, the concept of choice is no longer based on the idea that the chosen alternative always has to be better than all other available options, *i.e.* that the choice of a rational agent would result in the optimization (or Simon's idea of satisfaction) of the person's preferences, and therefore best improves the level of personal welfare.

To introduce his idea, Sen draws on a distinction between the motive of sympathy and a motive of commitment, both are different types of other-regarding or "altruistic motivation" (Pettit, 2007). Although a 'pro-social' motive, sympathy does not require a departure from personal welfare or individual utility maximization. As shown above with the example of altruism, an action motivated by sympathy is in the same way ultimately self-regarding, which means that "the concern for others directly affects one's own welfare" (Sen, 1977, p. 326). One example for the person acting from sympathy can be taken from Kant's *Groundwork*. Kant introduced the friend of mankind, who finds pleasure in spreading joy around him and who can take delight in the satisfaction of others. But Kant sees this satisfaction as an inclination like any other desire a person can have. He concludes that such an action, even if admirable, has no moral worth (Kant, 1997). In contrast, an action which is motivated by duty has moral worth; it is characterized as an action performed independently of personal desires. Kant illustrates this point by de-

---

[14]This identity mainly depends on the underlying understanding of the concept of a preference or the nature of reasons. As claimed before, personal welfare and individual utility are (technically) equalled in mainstream economics. This assumption provides the main basis for conclusions about what we understand as a preference and what the nature of reasons is (Sen, 1977, 1985).

scribing a philanthropist acting kindly towards others. He does not have any inclination which would positively or negatively affect his state of affairs. He simply acts because duty demands it. According to Kant, this gives the action a genuine moral worth (*ibid.*).

Abstracting from one's personal desires and inclinations implies that the action is not driven by them. Rather, an action motivated by commitment requires a "counterpreferential choice" from the part of the agent (Sen, 1977, p. 328). That does not mean, that the agent is not allowed to have any kind of additional inclination which makes the action attractive for him. This just means that under at least one counterfactual condition the personal welfare under the act chosen would be unaffected.[15] Hence, the agent would unconditionally perform the action, *i.e.* he would have no choice, because his commitment makes performing the action necessary for him. And even if these inclinations or desires would not be present, he would still perform the action in question. An action motivated by commitment can even imply a choice against one's personal welfare-improvement.[16] Personal welfare-improvements are not the *reason* for the action of the agent; they are not what the agent aims at in the first place (Sen, 1977; Hausmann, 2005). That means, an action motivated by commitment is completely non-egoistic.

It seems obvious that there is a close relationship between commitment and moral agency. According to Kant, acting from duty requires a sense of moral commitment. Being committed to moral principles is a necessary precondition for acting from duty. Taking Kant's motive of duty as one 'subgroup' of the general idea of commitment, Sen's understanding of commitment is that an action motivated by commitment is *per definitionem* independent from the satisfaction of any kind of inclination or immediate desires the agent is driven by in a given moment (Kant, 1997). In the case of commitment,

---

[15]It has to be noticed that it is difficult to imagine different states of the world and potential choices for action. However, I will not discuss this difficulty further here but just present the simple idea of commitment.

[16]I have to admit that an underlying premise of the argument is the empirical statement that commitment as characterized in this essay does exist. However, to ground this claim in empirical evidence seems to be a difficult task. People's motivations behind their actions are often seen as black boxes. There is a lot of empirical research still to be done on filtering out whether there is such a motivating force as commitment and, if yes, to disentangle this from other motives which are active at the same time the agent performs a certain action.

as is the case with actions done from duty, the agent considers the action as the *right* thing to do. Consequently, as when acting from duty, commitment requires the agent to be capable of undertaking judgment and evaluation, being self-reflected and conscious about his values and principles upon which he bases his evaluations.[17]

However, although morality implies commitment, morality should not be equalled or confused with it. Commitment results in the *right* action, although this does not necessarily mean that it results in what is seen as *morally right*. A person can also be committed to an ideology, a political party, or a religious view. A person can be committed to an organisation, to a social contract, to (moral) principles or to god. *Right* in this context thus just means that the agent considers the action itself (or the outcome of the action) as worth pursuing, independent of the personal benefit he can expect from the outcome of the action. And the agent comes to know that it is *the right thing to do* because he evaluates and judges the action on the basis of the values and principles he holds while comparing possible alternative actions instead of drawing on the evaluative principle of utility-maximization. In this sense, an action motivated by commitment is, as a moral action, unconditional and stable. But it does not necessarily aim at what is the morally right thing to do.

To sum up: the idea of commitment comprises an evaluation of possible alternatives that is not necessarily carried out by basing it on moral values; instead, it can be based on any set of values the individual in question holds. However, what commitment has in common with morality, at least understood in a Kantian way, is the existence of a non-welfarist goal that is altogether removed from individual preference-satisfaction and the fulfilment of immediate inclination. And the economist does not allow for this kind of behaviour.

Are there any ways out of this explanatory dilemma? Why not just assume an unconditional preference for either a certain value or commitment *per se*? The third question will be addressed to understand more clearly why even a very open economic model has prob-

---

[17]The character of commitment strongly depends on the definition of concepts. Frankfurt-minded philosophers would interpret commitment in a less evaluative and a less-cognitive way. According to Frankfurt people do not always know what they are committed to. Sometimes they believe they are committed to something they actually are not for example (*e.g.* Frankfurt, 1971). I just use Sen's understanding of commitment here, which seems very much in line with a Kantian view of moral agency.

lems integrating the motive of commitment. Where is this 'ought' or type of normativity grounded in? Does it differ between different people/cultures and why?

III.

One recent attempt at defining social norms while taking both types of normativity into account is by looking at people's expectations. People's preferences for social norms can be put in relation to normative expectations about actions that a group or the society has toward an individual or the individual has towards society (Coleman, 1987; Bicchieri, 2006). Cristina Bicchieri, one of the most prominent philosophers who work on social norms, stresses the importance of mutual expectations involved in norm-conformity. Her *Grammar of Society* published in 2006 is widely seen as "making a key contribution to our understanding of the motivations behind norm-compliance" (Küchle and Ríos, 2008, p. 118). She considers a social norm to be a kind of cluster of expectations among a sufficient number of people who believe that a social norm exists and who expect that many other people will follow it in certain situations. In this sense, people have a conditional preference for norm-conformity; conditional upon expectations about other people's belief in whether an action is seen as being right or wrong within a community (Bicchieri, 2006).

Bicchieri's account mainly consists of the following four (individually) necessary and (jointly) sufficient conditions for the existence of social norms: (1) People are collectively aware that a norm exists and that it applies to specific circumstances (contingency). (2) But the simple presence of the norm does not suffice to guarantee conformity. People also need to have a conditional preferences to conform to a social norm; conditional on (3) expectations of the following kind: A person believes that, based on past observations of other's behaviour or its consequences, the majority of the people will follow the norm as well (empirical expectations) and (4) a belief that others expect a person to conform (normative expectations) in the sense that they believe that everyone ought to conform. Here, expectations are taken to be reciprocal. Normative expectations can be accompanied by sanctions but do not have to be. The people who expect others to conform and who are willing to sanction trans-

gressions also have to *prefer* people to conform. What brings in the instrumental aspect of norm-conformity is that people have to have a (conditional) preference for conformity and the required beliefs in addition to the expectations. In line with the classical belief/desire-framework of rational choice theory, people are finally motivated by their preferences (Bicchieri, 2006). What brings in the normative aspect of conformity is the condition of mutually held 'normative expectations' about what people ought to do, based on what people consider as right or wrong. Let us look at this condition in greater detail.

The notion of mutually held 'normative expectations' introduces a type of normativity that differs from the mere instrumental normativity and from having a prudential reason for conformity. It implies that people not only think that others ought to conform, but that they also believe that they themselves are expected to conform in the sense that a reasonable number of people think that one *ought* to conform to the social norm. This can suffice to induce a preference for conformity when "individuals recognize the legitimacy of others' expectations and feel an obligation to fulfil them" (2006: 15); and, once induced, a preference for conformity guarantees that people follow the social norm. But it does not need to suffice. In some cases, Bicchieri claims, potential sanctions are needed to induce a preference for conforming to the social norm. And expecting sanction might motivate either in the sense that people fear sanctions and want to avoid them or because of a desire to please others and thus be rewarded.

What is of interest for the argument here is the nature of the 'ought' in Bicchieri's account. It can be implicitly contained in the normative or the empirical belief people hold about others' expectations. This is, people think they ought to conform because of their belief about others' expectations. In the case of an empirical belief, people infer from my (consistent) behaviour in the past that I will do the same in the future, and that this is what I believe (Bicchieri, 2006). If it is a normative belief things look a bit different. Although normative beliefs do not generally state an obligation, they do so in Bicchieri's account (*ibid.*). In this case, I believe that a sufficiently large number of people think that I have an obligation to conform to a social norm in the appropriate circumstances (*ibid.*). This is why a person will consider other people's expectations as legitimate.

It is not a kind of instrumental normativity, because the motivation for conformity understood in this sense can but does not have to incur material or immaterial sanctions to guarantee conformity. Also, conforming to the norm does not simply serve as a means to satisfy my ends Instead, people consider conformity to the social norm as their and other people's obligation or duty. But why does the person accept this obligation and thus considers other people's expectations as legitimate? Bicchieri claims that people think that they (and others) are obliged to conform to the norm because they believe it is the right thing to do. In these cases, social norms are seen as good or reasonable.[18] But how do people justify this obligation or, to put it differently, what is this obligation grounded upon? Bicchieri does not give a final answer. But she indicates that we have a preference to share a cake equally for example because it might be our duty to be fair. And we hold the belief that others expect us that we ought to share because it is the fair thing to do. But on what grounds do people believe that the norm of equal sharing is the right thing to do?

We could get around this problem by arguing that social norms are often not seen to be the right thing to do. But then the accepted obligation involved seems even more difficult to understand. To make her approach understandable, Bicchieri gives the following example of a Muslim woman wearing a veil: although it is a widely spread social norm, many people (in the Muslim community) might think that wearing a veil is an unpleasant thing to do and not many people would be prepared to sanction a transgression. This means, people would not consider it an obligation to conform to the norm. Each woman therefore does not necessarily believe that she ought to follow the social norm. But she might believe that she is expected to wear the veil in the sense that she believes that many people think that she ought to wear a veil and that they also prefer to wear a veil because it is an obligation or her religious duty to wear it. Because of social pressure she will finally end up wearing it; expected sanctions are necessary to guarantee conformity. Her reason to conform

---

[18]Of course, as Bicchieri rightly stresses, not all social norms are thought of being good and people do not think that others ought to follow them. Yet, conformity to them is widely observed. I will not consider this case here, but I am aware that this is an important issue which has to be stressed, when norm-conformity should be explained fully.

is thus a kind of 'instrumental reason' and the type of normativity involved has an instrumental nature. But let us look at this case in greater detail: what this case implies is that a religious community of extensive size would follow a norm which dramatically constrains the 'quality of life' of at least half of the members of the community, whereas 'quality of life' labels a lifestyle taken from the idea of freedom of secular western societies. And it assumes that people accept this constraint on their choices only because of other people's expectations, which, even worse, is based on false beliefs because, as Bicchieri describes, these expectations do not really exist if we would ask people for people's actual beliefs. What this example suggests, then, is that people live against what they really or truly value and consider as the right thing to do, simply because of a falsely believed existence of expectations.

This seems unreasonable to me. I guess that a vast majority of Muslim women would wear the veil even if they would not fear any kind of sanctions. And this is because they do it out of a religious conviction. They have committed themselves to the values and principles of their religious community. And they believe that conformity to the social norm is one way of expressing this conviction. This commitment is why they think they ought to follow the norm and why they also expect others to conform. Such a view immediately loses its validity when people have different values. Western women would have difficulty in believing in this kind of 'uncomfortable' commitment. Because of their commitment to other kind of values, their motivation and the normativity involved would be different immediately, *i.e.* rather instrumental in character. So, the conformity to the same norm can be motivated in at least two different ways, namely either fear of sanctions or conviction and commitment to values and principles that are embodied in the norm. Consequently, why people think they ought to follow a social norm also differs. As the one is instrumental in character and guarantees conformity only by expected sanctions, the conformity of a person who is motivated in this way is *contingent*. As the other one is 'committing' in character and guarantees conformity also without any kind of expected sanctions (the person conforms voluntarily and unconditionally), the conformity of a person who is motivated in this way is *necessary*.

I think this relates to an idea Hume raised (Hume, 2003). Sometimes norms and values or virtues appear to be related. Bicchieri

claims that once norms are invoked in a society, people start to follow them because they attribute some value to what the norm stands for or they have a deep conviction of the norm. Consequently, the norm is what shapes the values of the people within a community (Bicchieri, 2006). Accordingly, when we attribute an intrinsic value to the norm, this gives us a reason to recognize other people's expectations as legitimate and we feel an obligation to fulfil it. These expectations cause us to have a preference and having a preference causes us to conform to the norm we value or are deeply convinced of. So their 'value depends on widespread conformity' (Bicchieri, 2006). I think this seems somewhat contradictory. We consider an action worth pursuing because its properties embody certain values and principles we already have or hold and about which we are convinced. We are committed to these values and, by following a norm, we want to implement or realize them. Consequently, it is often our values and principles that shape social norms, not the other way round.[19] Because we are committed to the principles and values we hold, we take it as our obligation to conform to the norm. This obligation is independent, first, from our preferences and thus from what is in our self-interest and, second, from other people's expectations.

So, as I claimed before, I think that individuals draw on a basis of values and principles that people commonly hold. And because they commit themselves to these values and principles, on the one hand, they accept other people's evaluations and expectations about what kind of behaviour is considered as appropriate and find them legitimate. On the other hand, their commitment also justifies their own expectations towards other people's behaviour and their own norm-compliance. This, so the economist will argue, seems vague and does not explain properly the normativity involved. We should rather grant people to have preferences for fairness or for other social concerns which would replace this talk about values. And, if we

---

[19]Empirically it is probably true that in many societies certain social norms exist before new individuals build up values that support these norms. But what was there first and how (informal) institutions evolve are question which cannot be addressed in this paper. However, I think that, if what the social norms represent or imply would be contradictory to what we consider for ourselves as acceptable and would not be more or less inline with our values, conceptions, and ideas about the world, the way we live together in a society, and how we think we should treat each other, then the majority of the people would not conform to social norms.

cannot get around the idea of commitment, we could simply assume them to have a preference for commitment.

But what would a preference for commitment look like? Having characterized what the nature of commitment is, it seems plausible that the idea of a preference and what it implies — both technically and factually — seems to be misleading in this context. Commitment is inherently contradictory to a self-interested action and does not lead to personal-welfare improvement in the first place (Sen, 1985, 2005). One could object that Bicchieri allows for counter-preferential choices: in a situation where (material) rewards are present, I might prefer more to less but when people expect me to follow a norm of generosity 'I might prefer to behave generously'. Is this what we intuitively understand as counter-preferential choice? It rather sounds like having a dominant preference for conforming to the norm of generosity; we can change our preference or one preference dominates the other, but then we do not take a choice which contrasts our preferences. Thus, the action is still based on and not against our preferences.

This relates to the point that the dominant role of expectations and the way in which we form our preference for social norms seems quite odd. Bicchieri unifies preference-formation and the existence of expectations and does not focus on whether other's expectations make any sense for us. In the long run, it would be impossible for an agent to keep his promises in a world where nobody kept theirs. He would have to change his preference in accordance with the change of his beliefs about other's expectations, independent of whether he cares about taking the specific action 'keeping a promise', *i.e.* whether he judges the action of keeping promises as good or the right thing to do. It does not matter what kind of behaviour we believe to be expected, it could be keeping a promise or it could be jumping as high as we can when we see a green bird. But — according to the theory — whether we consider the action to follow a norm worth pursuing or the right thing to do (which is obviously not the same as having a preference) does not matter.[20] This conflicts with the idea of commitment. But why would people prefer what other people expect them to do, no matter how 'stupid' it seems to

---

[20]Note that on this account it is very likely that norm-guided behaviour will change with circumstances, because expectations may be different in distinct contexts, which changes the preference of the agent and thus his behaviour.

be? Why would all Muslim women accept to be strongly limited in their personal freedom, if they completely disagreed with what the norm says and even formed a preference to this extent?[21] Even if this point appeals only to our intuition and thus has yet to be proven, I doubt that any expectations we believe people to hold have this power to form people's preferences accordingly.

This leads us to the last point, namely that awareness, self-reflexion and conscious deliberation are not necessary conditions for following a norm in Bicchieri's account. She claims that we often unconsciously follow social norms, while she defines a mental state to be conscious when it is accompanied by roughly simultaneous, higher-order thought about that very mental state. This understanding of awareness is reminiscent of conscious deliberation modelled in rational choice theory. In a cost-benefit-analysis, we consciously weigh our options available. In this context, deliberation is based on the beliefs and desires of which we are aware in the abovementioned sense. I believe that Bicchieri understates the importance of this kind of awareness and claims that we often make use of heuristics to 'choose' our actions, especially when it comes to norm-conformity. And she might be right in many cases. However, in the context of norm-conformity motivated by commitment, awareness also requires making judgements and it implies self-consciousness achieved by a process of deliberating about how we act on the whole. We draw on our experience of how we acted in the past and why we acted as we did to make sure that our action was worth pursuing and to adapt/improve our action in the future, having in mind certain fundamental values and principles to which we commit ourselves. As said before, because of this commitment, people use their values and principles for assessment. These values and principles for assessment are different from the notion of value implied by Bicchieri's claim that some social norms may become part of our system of values, and that we may feel a strong obligation to obey them (Bicchieri, 2006). In-

---

[21]Of course, the economist would argue that Bicchieri's account is just a rational reconstruction and does not aim at showing how people's *real* preferences are (Bicchieri, 2006). But besides the fact, that a rational reconstruction of an account for norm-conformity seems problematic as people, depending on their motives, conform to social norms in a more or less stable manner, it seems problematic to talk about what shapes people's preferences in the first place, when we just assume that people have preferences. In this case the whole talk about expectations would loose importance.

stead, as mentioned before, the concept of commitment implies that some norm-governed actions embody people's values and principles. And because the norm embodies the values and principles we are committed to, it has a normative claim on us which is a general claim independent of the situation. So, once we judge an action as worth pursuing on the basis of our values and principles, we have the duty to pursue it; in case of social norms 'we obey ourselves' to conform. The awareness of the existence of these properties, of the fact that they provide us with a reason to act, and of the values and principles we hold is therefore crucial when we want to account for the motive of commitment.

Economists might argue that, if we do not have a preference for the norm, these values constrain us in following our desires, so why should we commit ourselves to them? Let me give a constructivist answer to that question: these assessment criteria are constructed by the agents themselves through their own will (Korsgaard, 1997). During the process of deliberation, where the agent engages in critical thinking, he enlarges his perspective and option-set and gets a deeper understanding of the situation and of himself (Wallace, 2003). He becomes conscious of values and principles on which he wants to base his actions. This process does change his intentions; it is a kind of self-determination. Instead of passively reacting to his desires, he deliberates and assesses his own reasons for his actions on the basis of these values and principles. And because they reflect his overall view of what is worth to pursue in life, he wills them to have a claim on him.

To sum up, Bicchieri's definition of social norms has the advantage of being operational. It provides us with a pragmatic framework. It has empirical, testable content which allows us to predict norm-governed behaviour, especially in experimental settings. At the same time it is highly flexible and sophisticated enough to include aspects like emotions and internalization of norms. However, while it is able to explain nearly all types of motives people can have for norm-conformity, it also shows that an action motivated by commitment cannot be explained. The only way out would be to induce a preference for commitment but this is not what we are looking for, as personal welfare-improvement and self-interest are not part of the motive of commitment. Hence, it is problematic to account for norm-conformity in a satisfactory way with an approach where choice is

based on the motive of (immediate) satisfaction of individual desires or preferences and thereby aiming at personal welfare-improvements.

IV.

It has to be analysed further whether the theory of human behaviour used in economics to explain norm-conformity fulfils its explanatory and predictive role when accounting for norm-conformity. It has to be asked how useful economic explanations of norm-conformity are. To enhance their usefulness and adequacy on the one hand, and be more realistic in what they predict on the other, we have to understand what really motivates people. And finding out the motivation, the normativity involved in the conformity of the people, plays a crucial role. Thus, analysing the normativity involved would help us to improve our explanations in economics and explain cases of cooperation which are just ad hoc explained by standard as-if-models based on the rational choice approach. This difference in motivation and thus in the normativity involved has two important consequences. First, it has important implications on a methodological level when giving an explanation of norm-conformity. The explanation based on an account where normativity is assumed to be purely instrumental seems to be misplaced, as it ignores the possibility for a normativity of a different type. Second, it has important implications on a conceptual level. For, what we define as a social norm in contrast to other rules depends on why people think they ought to follow the social norm, *i.e.* what motivates their conformity. And if some motives can be covered but others are left out we can have legitimate doubts about the usefulness of the explanation provided.

My guess is that there are two ways out of this 'explanatory dilemma'. Either, we fully dismiss economic models when it comes to explaining norm-conformity. Or, we conduct further theoretical and empirical research in this specific area to extend the economic model, improve our utility functions and provide better cognitive and behavioural assumptions of the agents.

IV   Conclusion

The aim of the paper was to show the difficulty economists have in accommodating the motive of commitment in their explanatory frameworks for norm-conformity. This is because in the case of norm-

conformity, the normativity involved can be instrumental and/or committing. Of course, people's behaviour is complex and a theoretical framework will probably never cover all of the potential motivations for following social norms. However, I conclude that if we aim at reasonable explanations of norm-conformity, all possible motives should be considered. The concept of preferences is often misleading when it comes to explaining norm-conformity, as it dismisses the normative dimension of agency, *i.e.* the ability of the agent to undertake judgements and make choices independently of one's personal desires or inclination, *i.e.* independent of what is in the person's self-interest. Instead of granting the agent all types of preferences, to find out the motivation behind norm-conformity, it is rather useful to analyse the normativity involved in what motivates agents to conform to social norms. The type of normativity or the 'ought' involved in social norms differs remarkably, depending on the motivation people have. I argued that we have to distinguish between two forms of normativity involved in norm-conformity: one is what I called 'instrumental normativity', which is well captured by the economic framework and the notion of preferences. The second is a kind of 'normativity of commitment', which, so the argument was, cannot be captured by economic theory. Depending on people's motivation, they consider conformity as normative either in one of the two senses or in both at the same time. The main difficulty of the economic theory to account for these aspects lies in its general understanding of what is meant when we say that somebody acts for a reason and what we consider as being rational within the framework. It does not account either for the kind of reason we are provided with when acting from commitment, nor for what we consider as a rational in this context. Thus, if we accept commitment as a motivational source, then we must challenge our notion of reasons, *i.e.* the nature of the desires and beliefs involved in reasons for conformity and our understanding of what it means to act rationality in this context.

References

Akerlof, G. A., 1982. Labor contracts as partial gift exchange. Quarterly Journal of Economics 97 (4), 543–569.

Anderson, E., 2000. Beyond homo economicus: New developments in theories of social norms. Philosophy and Public Affairs 29 (2),

99

170–200.

Anscombe, G. E. M., 1963. Intention, 2nd Edition. Basil Blackwell, Oxford.

Basu, K., Weibull, J. W., 2003. Punctuality: A cultural trait as equilibrium. In: Arnott, R. J., Greenwald, B. C. N., Kanbur, S. M. R., Nalebuff, B. J. (Eds.), Economics for an imperfect world: Essays in honor of Joseph E. Stiglitz. MIT Press, Cambridge, MA, pp. 163–182.

Bicchieri, C., 2006. The grammar of society: The nature and dynamics of social norms. Cambridge University Press, Cambridge.

Coleman, J. S., 1987. Norms as social capital. In: Radnitzky, G., Bernholz, P. (Eds.), Economic imperialism: The economic approach applied outside the field of economics. A Professors World Peace Academy Book. Paragon House, New York, NY.

Coleman, J. S., 1990. Foundations of social theory, 1st Edition. Belknap Press of Harvard University Press, Cambridge, MA [a.o.].

Conlin, M., Lynn, M., O'Donoghue, T., 2003. The norm of restaurant tipping. Journal of Economic Behavior & Organization 52 (3), 297–321.

Davidson, D., 1963. Actions, reasons, causes. Journal of Philosophy 60 (23), 685–700.

Davis, J. B., 2004. Identity and commitment: Sen's conception of the individual. Tech. Rep. TI 2004-055/2, Tinbergen Institute, Amsterdam.

Dawes, R., Thaler, R., 1988. Anomalies: Cooperation. Journal of Economic Perspectives 2 (3), 187–197.

Elster, J., 1989a. Social norms and economic theory. Journal of Economic Perspectives 3 (4), 99–117.

Elster, J., 1989b. Solomonic judgements: Studies in the limitations of rationality. Cambridge University Press, Cambridge [a.o.].

Elster, J., 1994. The nature and scope of rational-choice explanation. In: Martin, M., McIntyre, L. C. (Eds.), Readings in the philosophy of social science, 2nd Edition. A Bradford book. MIT Press, Cambridge, MA, pp. 311–322.

Fehr, E., Fischbacher, U., 2003. The nature of human altruism. Nature 425, 785–791.

Fehr, E., Fischbacher, U., 2004. Social norms and human cooperation. TRENDS in Cognitive Sciences 8 (4), 185–190.

Fehr, E., Kirchler, E., Weichbold, A., Gächter, S., 1998. When social norms overpower competition: gift exchange in experimental labor markets. Journal of Labor Economics 16 (2), 324–351.

Frankfurt, H. G., 1971. Freedom of the will and the concept of a person. The Journal of Philosopy 68 (1), 5–20.

Hausman, D. M., McPherson, M. S., 2006. Economic analysis, moral philosophy, and public policy, 2nd Edition. Cambridge University Press, Cambridge.

Hausmann, D. M., 2005. Sympathy, commitment, and preference. Economics and Philosophy 21, 33–50.

Heap Hargreaves, S., Hollis, M., Sugden, R., Weale, A., 1992. Theory of choice: A critical guide. Blackwell, Oxford.

Hechter, M., Opp, K.-D. (Eds.), 2001. Social norms. Russell Sage Foundation, New York, NY.

Henrich, J., 2000. Does culture matter in economic behaviour? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon. The American Economic Review 90 (4), 973–979.

Hume, D., 2003. A treatise of human nature. Dover Publications, Mineola N.Y.

Kahneman, D., Knetsch, J. L., Thaler, R., 1986. Fairness as a constraint on profit seeking: entitlements in the market. The American Economic Review 76 (4), 728–741.

Kant, I., 1997. Groundwork of the metaphysics of morals. Cambridge texts in the history of philosophy. Cambridge University Press, Cambridge.

Kerr, N. L., Garst, J., Lewandowski, D. A., Harris, S. E., 1997. That still, small voice: commitment to cooperate as an internalized versus a social norm. Personality and Social Psychology Bulletin 23, 1300–1311.

Kirchgässner, G., 2006. Das ökonomische Verhaltensmodell: Der Homo Oeconomicus. In: Nell, V. v., Kufeld, K. (Eds.), Homo oeconomicus: Ein neues Leitbild in der globalisierten Welt? Vol. 1 of Forum Philosophie & Wirtschaft. LIT Verlag, Berlin, Hamburg, Münster, pp. 81–107.

Korsgaard, C. M., 1997. The normativity of instrumental reason. In: Cullity, G., Gaut, B. (Eds.), Ethics and practical reason. Clarendon Press, Oxford, pp. 215–254.

Krupka, E. L., Weber, R. A., 2008. Identifying social norms using coordination games: Why does dictator game sharing vary? Tech. Rep. 3860, Institute for the Study of Labor (IZA), Bonn, available online from `http://ftp.iza.org/dp3860.pdf`, last accessed: 24.02.2009.

Küchle, G., Ríos, D., 2008. Book review of Cristina Bicchieri's Grammar of society. Economics and Philosophy 24 (1), 117–123.

Kurzban, R., Houser, D., 2005. Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. Proceedings of the National Academy of Sciences of the United States of America 102 (2), 1803–1807.

Lindbeck, A., Nyberg, S., Weibull, J. W., 1999. Social norms and economic incentives in the welfare state. The Quarterly Journal of Economics 114 (1), 1–35.

Mantzavinos, C. (Ed.), 2001. Individuals, institutions, and markets. Political economy of institutions and decisions. Cambridge University Press, Cambridge.

Millar, A., 09.05.2007. The explanatory role of normative considerations. Available online from `http://www.`

essex.ac.uk/philosophy/tpn/documents/Normativity/
MillarExplanatoryRole.pdf.

Okruch, S., 1999. Innovation und Diffusion von Normen: Grund-
lagen und Elemente einer evolutorischen Theorie des Institutio-
nenwandels. Vol. 491 of Volkswirtschaftliche Schriften. Duncker &
Humblot, Berlin.

Opp, K.-D., 1999. Contending conceptions of the theory of rational
action. Journal of Theoretical Politics 11 (2), 171–202.

Pauer-Studer, H., 2007. Instrumental rationality versus practical rea-
son: desires, ends, and commitment. In: Peter and Schmid (2007),
pp. 73–104.

Peter, F., Schmid, H. B. (Eds.), 2007. Rationality and commitment.
Oxford University Press, Oxford.

Pettit, P., 1990. The reality of rule-following. Mind 99 (393), 1–21.

Pettit, P., 2007. Construing Sen on commitment. In: Peter and
Schmid (2007), pp. 28–49.

Raz, J., 2005. The myth of instrumental rationality. Journal for
Ethics and Social Philosophy 1 (1), 1–28.

Rosenberg, A., 1995. Philosophy of social science, 2nd Edition. Di-
mensions of philosophy series. Westview Press, Boulder.

Samuelson, P. A., 1938. A note on the pure theory of consumer's
behavior. Economica 5 (17), 61–71.

Schmid, H. B., 2007. Beyond self-goal choice: Amartya Sen's analysis
of the structure of commitment and the role of shared desires. In:
Peter and Schmid (2007), pp. 211–226.

Searle, J. R., 2001. Rationality in action. Vol. 2000 of A Bradford
book. MIT Press, Cambridge, MA.

Sen, A. K., 1977. Rational fools: A critiique of the behavioral foun-
dations of economic theory. Philosophy and Public Affairs 6 (4),
314–344.

Sen, A. K., 1982. Choice, welfare and measurement. MIT Press, Cambridge, MA.

Sen, A. K., 1985. Goals, commitment, and identity. Journal of Law, Economics, and Organization 1 (2), 341–355.

Sen, A. K., 2005. Why exactly is commitment important for rationality? Economics and Philosophy 21, 5–13.

Stout, L. A., 2001. Other-regarding preferences and social norms. Tech. Rep. Georgetown Law and Economics Research Paper No. 265902; and Georgetown Public Law Research Paper No. 265902, Georgetown University Law Center, Washington, DC, available online from `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=265902`, last accessed: 22.02.2009.

Sugden, R., 2000. Team preferences. Economics and Philosophy 16 (175-204).

Turnbull, C. M., 1972. The mountain people. Simon & Schuster, New York.

Wallace, R. J., 2003. Practical reason. In: Zalta, E. N. (Ed.), The Stanford Encyclopedia of Philosophy (Winter 2003 Edition).

Wallace, R. J., Smith, M., Pettit, P. (Eds.), 2004. Reason and value: Themes from the moral philosophy of Joseph Raz, 1st Edition. Oxford University Press, Oxford.

Woodward, J., 2009. Why do people cooperate as much as they do. In: Mantzavinos, C. (Ed.), Philosophy of the social sciences: Philosophical theory and scientific practice. Cambridge University Press, Cambridge, pp. 295–357.

Naturalism and Normativity

Christos Kyriacou (University of Edinburgh)

C.Kyriacou@sms.ed.ac.uk

I   Introduction

In metanormativity debates, it is often considered that a naturalistic reductionist framework cannot accommodate our normative notions: moral, epistemic and semantic.[1] The question of this tension between naturalism and normativity, some of the main theoretical approaches to normativity in our disposal and the question of which one might be the most promising will be the topic of this talk. I will overview three of the main theoretical approaches to normativity on offer, examine some of their respective virtues and vices and for methodological reasons briefly explained express my sympathy for the expressivist approach.

First, some short clarificatory definitional preliminaries on the notions of 'naturalism' and 'normativity' are due. Philosophers have a notorious difficulty in analysing abstract theoretical concepts in terms of necessary and sufficient conditions and naturalism is one of these abstract theoretical concepts that is difficult to analyse in such conditions. The notion is often glossed in many interrelated ways (scientific, epistemological, ontological, and methodological) and this is not the suitable place to disentangle these various ways and opt for

---

[1]See Moore (1903); Ayer (1946); Gibbard (1990); McDowell (1994); Williamson (2000), etc.

the most elegant.[2] But for present purposes the following stipulation of naturalism is, I think, adequate: all states of affairs are natural states of affairs as these are disclosed to us by the causal interaction of our perceptual mechanisms and the natural world. Accordingly, I will take a natural property to be an empirical property and a nonnatural property to be a non-empirical property. That is, an empirical property is a property with independent causal powers and a non-empirical property is a property with no independent causal powers.[3]

The case for the notion of normativity is not different. It is difficult to pin down a definition of normativity and we need not set out here for such a definitional 'wild goose chase'. For present purposes we can stipulate that normativity is concerned with 'oughtness'. Philosophers often talk about what 'we ought to do', 'how we ought to live', 'what we ought to believe', 'how we ought to reason', 'what we ought to mean' and take associated concepts like goodness, rationality, justification, validity and meaning to be normative. As W. Sellars once put it, such normative concepts are concepts "fraught with ought."[4] Equally, they regard practical/moral, epistemic, logical and semantic discourses to be normative. These discourses are not about 'what is the case' but about 'what ought to be the case'.

II   Normativity and the Naturalistic Reductionism/Nonnaturalistic Anti-Redutionism Dilemma

Now that the technical notions of naturalism and normativity have been clarified a bit, we can go on with a brief overview of three of the main approaches to normativity. A question confronts anyone who wants to come to grips with the question of normativity: *How should we understand normative concepts*? In recent debates, attempted answers to this question have led to the formulation of a dilemma of theoretical approaches.[5] The first horn of the dilemma is

---

[2]For some discussion of various glosses on 'naturalism' see Papineau (2007) and for the natural/non-natural distinction Ridge (Fall 2008).

[3]The causal account of property has been propounded by Shoemaker (2002) and Lewis (2002).

[4]This phrase is attributed by Gibbard (2003, p. 21)) to W. Sellars.

[5]The dilemma is not meant to be exhaustive. There are many other approaches that elude the horns of the dilemma — e.g., Anti-reductive Cornell-style naturalism, Kantian constructivism, error theory etc. — and in this occasion we cannot go through them all. Also, the dilemma is a very old one going all the

Naturalistic Reductionism. It suggests that normative concepts can and should be non-circularly, reductively analysed in terms of non-normative, natural properties. Following the naturalistic reductionism project, it has been suggested that moral concepts like 'goodness' and 'justice' should be respectively analysed as 'pleasure' and 'giving back what isn't yours';[6] or epistemic concepts like 'justification', 'knowledge', and 'truth' should be respectively analysed in terms of 'reliable belief-forming processes', 'true justified belief', and 'the correspondence relation between truth-value bearers and truthmaking facts';[7] or semantic concepts like 'synonymy' should be analysed in terms of 'intersubstitutivity *salva veritate*'.[8]

Unfortunately, such attempts seem to meet Moorean "open question" semantic intuitions that allow counterexamples to emerge and defeat the suggested definitions. Moore (1903) famously argued that is always possible for a clear-headed agent to question, without any semantic confusion, whether a natural property N — e.g., pleasure — is also synonymous with a moral property M — e.g., goodness. In short, the reduction of a moral property to a natural property remains always an "open question." Moore assumed that if there was such an *a priori* property identity, a clear-headed agent wouldn't fail to grasp it and carried on — from the fact that the semantic 'openness' founders all such reductive efforts — to draw the conclusion that moral properties cannot be reduced to naturalistic terms. He contended that all such reductive attempts have committed — what he called — the "naturalistic fallacy" because there are no natural properties that could reduce moral properties.

Analogous Moorean "open question" semantic intuitions seem to apply to virtually all normative concepts. These "open question" semantic intuitions undermine attempts to reduce normative concepts in naturalistic terms and allow counterexamples to pop up and

---

way back to Plato. See my Kyriacou (Unpublished).

[6]The analysis of goodness as pleasure has been suggested in Plato's Protagoras, J.S. Mill's Utilitarianism and in various ways by modern Utilitarians. The analysis of justice as 'giving back what isn't yours' appears in Plato's Republic I and is subsequently refuted by the Socratic character.

[7]The reliabilist analysis of justification has been prominently proposed by Goldman (1992) and the analysis of truth as a correspondence relation, among others, by Russell (1912).

[8]The intersubstitutivity *salva veritate* analysis of synonymy is proposed and then refuted by Quine (1953).

refute the proposed reductions. Thus, intuitively, there are respective instances of bad pleasures — e.g., sadistic cat killings –, there are cases where the just thing to do is not to give back what isn't yours — e.g., when someone lends you a weapon and then she goes mad —, there are Gettier-cases of epistemic luck where true justified beliefs are not instances of knowledge — e.g., the classic cases with sheep-like dogs, façade barns etc. — cases where beliefs formed by reliable cognitive processes are not ultimately justified — e.g., Euclidean geometry in the light of Einsteinian physics —, cases where there are no natural truth-making facts of seemingly true beliefs — e.g., mathematical, logical facts etc. — and cases where synonymy is not 'intersubstitutivity *salva veritate*' — e.g., Quine's (1953) counterexample of 'renate' and 'cordate' meaning respectively 'creature with kidneys' and 'creature with a heart'.[9] Such normative concepts seem to resist successful reductive analysis and this decisively undermines the project of naturalistic reductionism. It appears then, that there are no natural properties that can reduce such normative notions.

It should be noted, though, that Moorean "open question" considerations are, as Frankena (1939) pointed out, question-begging but still such considerations can be taken to be enthymatic. As Frankena rightly objected, Moore's open question argument begs the question against the naturalistic reductionist because it assumes that there is no such reductive analysis of moral concepts to be found. This seems to trivialize the open question argument because it might be the case that we just didn't *yet* find the right analysis of moral (and other normative) notions and such reductive analyses will be found sometime in the future.

However, many anti-reductionist philosophers (nonnaturalists and expressivists alike) think that there is something valuable in the argument that should not be lost by the, indeed, correct Frankena objection. The argument indicates that intuitively our constant failure to find such successful reductive analyses inspires a legitimate pessimism about the prospects of such a discovery. Actually, the anti-reductionist pessimism is so pervasive that, even if we find a reductive analysis that seems to be initially immune to counterexam-

---

[9]The idea in Quine's counterexample is that 'renate' and 'cordate' are coextensive and therefore can be intersubstituted and preserve truth but yet they are not synonymous. Renate means 'creature with kidneys' and cordate means 'creature with a heart'.

ples, what we will think is that we haven't yet found the right coun-
terexample, not that the analysis is ultimately successful.[10] Thus,
despite the fact that the argument is question-begging and far from
being conclusive against naturalistic reductionism, it need not be
taken to be trivial. It can be considered to be an *inference to the
best explanation* of our constant failure to reach reductive naturalistic
analyses of normative notions.

The constant failure to provide successful analyses of normative
concepts has frustrated many philosophers who accepted the ver-
dict of Moorean "open question" considerations and turned to the
second — Platonic — horn of the dilemma: *Nonnaturalistic Anti-
reductionism*. Nonnaturalistic Anti-reductionism accepts that nor-
mative concepts are unanalysable, *sui generis* concepts that purport
to refer to the corresponding nonnatural properties (goodness, jus-
tification etc.).

Thus, Plato in his middle period dialogues (*Phaedo, Republic,
Theatetus*) took the properties of goodness, justice and knowledge
to be irreducible and nonnatural (or in Platonic jargon, "forms")
and Moore (1903) also famously proclaimed that the property of
goodness is "simple, indefinable, and *sui generis*" and any effort to
reduce it in naturalistic terms commits the "naturalistic fallacy."
Sellars (1956) explicitly drew a parallel between ethics and episte-
mology and argued that:

> . . . the idea that epistemic facts can be analysed without remainder
> — even 'in principle' — into non-epistemic facts . . . is I believe a
> radical mistake — a mistake of a piece with the so-called "natural-
> istic fallacy" in ethics.

For Sellars, epistemic concepts like knowledge and justification
are not naturalistically analysable, they belong to the irreducible
"space of reasons."[11] Finally, Williamson (2000) has also recently
argued that the concept of knowledge is basic and unanalysable.

---

[10]Compare: "Even if some sufficiently complex analysis never succumbed to
counterexamples, that would not entail that identity of the analysing concept
with the concept knows. Indeed, the equation of the concepts might well lead to
more puzzlement rather than less." Williamson (2000, pp. 30–31).

[11]As he has put it: "[t]he essential point is that in characterizing an episode
or a state as that of knowing, we are not giving an empirical description of that
episode or state; we are placing it in the logical space of reasons, of justifying and
being able to justify what one says." Note that it is controversial how to read
Sellars (1956). Some times is read as an expressivist and others as a nonnaturalist.

Further, philosophers like Putnam (1978, pp. 107–9) and Sosa (2001) have also directly appealed to Moorean "open question" considerations in order to claim that the property of truth is unanalysable while Davidson (2001) has appealed to analogous Moorean considerations in the Platonic corpus in order to reach the same anti-reductionist conclusion. They all conclude that the property of truth is irreducible and nonnatural. Additionally, McDowell (1994) takes "meaning facts" like synonymy to be unanalysable in naturalistic terms. For McDowell, meaning cannot be understood in terms of what he calls "bald naturalism." That is, a naturalism that identifies the natural realm with the causal realm (or the "realm of law" as he calls it) as we did basically with our stipulative accounts of naturalism and a natural property.[12]

Unfortunately again, the nonnaturalist horn meets some serious epistemological and ontological difficulties of its own that undercut its prospects of success.[13] First, there is the "epistemic access problem." Nonnaturalists often posit an antenna-like faculty of rational intuition which supposedly can track normative nonnatural properties through some sort of 'intellectual seeing'. But given that we have epistemic access to natural properties, objects, events etc. with ordinary causal-perceptual processes, it appears mysterious how we can have epistemic access to acausal, nonnatural normative properties. Thus, talk of 'intellectual seeing' appears epistemologically spooky, unless there is a way to cash out the ocular metaphor in naturalistic terms. Second, there is the "queerness problem." It is hardly clear what such acausal nonnatural properties could be. For surely they must be radically different from the mundane natural properties, objects etc. we perceive in our everyday life. If there were such properties, in Mackie (1977) often quoted words, they would seem to be ontologically 'queer'. They would be radically different from ordinary natural properties.

---

[12] McDowell would not sympathize with his classification as a nonnaturalist. He repeatedly says that he is critical to "rampant Platonic nonnaturalism" and that his position is a neoaristotelian "naturalized platonism," not a "bald naturalism." As he says, his naturalism is a "naturalism of second nature" not a naturalism that identifies nature with the realm of causal law. For reasons that cannot detain us here, though, I think that McDowell's position subsides to yet another form of nonnaturalism.

[13] Of course, there are many other difficulties for nonnaturalism like the moral motivation problem and the problem of satisfying the supervenience constraint but we cannot go through them all here.

III   The Anti-Reductionist Expressivist Alternative

The difficulties of the naturalistic reductionism and nonnaturalistic anti-reductionism horns of the dilemma have ushered some philosophers to search for a different approach to normative concepts. That is, a different approach that could avoid at least many of the difficulties that surround and undermine the two horns. One approach that was born from this endeavour (in Ayer's (1946) hands) was expressivism.[14]

Expressivism, like nonnaturalism, accepts the verdict of Moorean "open question" considerations and takes normative concepts to be irreducible to naturalistic terms. Crucially, however, it denies the factualist assumption that both horns of the dilemma share. Namely, it denies that there are normative properties and, since our perceptual mechanisms do not indicate the existence of such normative properties (natural or nonnatural), it comports nicely with a parsimonious naturalistic framework. For the expressivist approach, there are no normative properties (natural or nonnatural) that normative concepts purport to refer to. Actually, expressivist semantics are not referential at all. They are psychologistic (or ideational) semantics explaining the meaning of sentences in terms of the state of mind they express in the light of linguistic conventions. The semantics of normative discourse are taken to be expressivist and the state of mind expressed noncognitive — i.e., nondescriptive states like attitudes, desires, plans, sentiments, etc.

Nevertheless, in spite of the rejection of the factualist assumption, Quasi-Realist sophisticated expressivists like Gibbard (1990, 2003), Blackburn (1993, 1998) and Ridge (2006, 2007) in various different ways claim that we can still latch onto normativity and reconcile it with such a parsimonious naturalistic framework. They claim that expressivism can mimic the realist-soundings of normative discourse and adequately explain them without any loss of explanatory power. To that effect, they propose a minimalist 'ontologically light' reading of notions like 'property', 'truth', 'objectivity', etc., and claim that expressivism can legitimately earn the right to talk about minimalist normative properties, truth, etc., without conceding that traditional realism can do any better.

---

[14]Ayer's expressivism was meant to apply only to moral concepts but, as we shall see, there are contemporary expressivists who are expressivists about normativity in general. Another kind of approaches that were born from this endeavour are virtue theory approaches like, for instance, McDowell (1994).

By denying the factualist assumption of both naturalistic reductionism and nonnaturalistic anti-reductionism, expressivism draws out the rug from under the feet of both horns of the dilemma and opens some very interesting logical space between the horns of the dilemma. This new logical space promises to reconcile naturalism and normativity and elude the difficulties of naturalistic reductionism and nonnaturalistic anti-reductionism. On the one hand, expressivism comports nicely with a broadly naturalistic framework, as naturalistic reductionism would like it to be, while at the same time explains our "open question" semantic intuitions concerning our efforts to analyse normative concepts. For expressivists, "open question" semantic intuitions indicate that there no natural properties that can successfully reduce our normative concepts. On the other hand, sophisticated Quasi-realist expressivists like A. Gibbard, S. Blackburn and M. Ridge (in various ways) attempt to show that we can uphold the realist-soundings of normative discourse in an expressivist context, as the traditional realist nonnaturalist would like it to be, while at the same time remain committed to a broadly naturalistic framework that evades the "epistemic access" and "queerness" problems of nonnaturalism.

As often happens in philosophy, however, the expressivist approach meets its own difficulties. Expressivism faces its own thorny problems and some philosophers think that these problems are so deep they could even amount to a *reductio* against expressivism. The two perhaps thorniest problems for expresssivism are semantic problems: the notorious Frege-Geach problem (or the "embedding problem") and what we may call the "truth problem." The Frege-Geach problem is a semantic problem which imperils validity.[15] Very roughly, the kernel idea in the Frege-Geach problem is that expressivism seems to allow in unasserted truth-functional contexts (like a *modus ponens*) valid conclusions not to be drawn due to a "fallacy of equivocation." Normative moral and epistemic (and nonnormative) terms appear embedded in unasserted truth-functional contexts where valid inferences, if we are to respect validity, ought to follow from true premises. Validity in such unasserted truth-functional contexts can be preserved only if meaning synonymy is preserved from one premise to another.

---

[15]The problem was first formulater by Geach (1972a,b). Geach's formulation of the problem was inspired by Frege's (1977) discussion of negation.

However, if we take semantics to be expressivist, then in unasserted truth-functional contexts valid conclusions might not follow (as they ought to) because it seems possible, given that expressivism does not posit truth-conditions for sentence-meaning, that meaning synonymy from one premise to another might not be preserved. There might be a shift of meaning form one premise to the other and this would entail a "fallacy of equivocation." Sentences or parts of sentences with the same sentence-meaning may have different speaker-meaning and then the valid conclusion won't follow. If that is correct, then expressivism cannot handle validity and closely associated semantic properties like (in)consistency and entailment and for apparent reasons this is sufficient to shoot down *any* such theory of semantics.

Expressivism also seems to meet an equally thorny difficulty with the notion of truth. Expressivist semantics are not referential and some philosophers think that nonreferential semantics cannot adequately capture the notion of truth and the objectivity that surrounds it.[16] Nonreferential semantics, they claim, fail to accommodate the correspondence intuition that lies in the heart of the notion of truth. Sophisticated expressivists like Gibbard, Blackburn and Ridge have responded to such truth worries by allying expressivism to a popular sophisticated nonreferential theory which gives the notion of truth a minimalist reading: the deflationary theory of truth.[17]

Many people, though, are still unimpressed by the alliance of expressivism with deflationism and claim that appeal to a deflationary theory of truth cannot help expressivism. They argue that the deflationary conception of truth, despite its subtlety and sophistication, is still a nonreferential theory and, thus, it is too meagre to be able to hold on its shoulders the heavyweight notion of truth. They think that a deflationary theory of truth ultimately collapses to some form of relativism and therefore cannot capture our objectivity intuitions, as these swarm around the notion of truth.

IV   Methodology and Theory-Choice Theory

Given that each one of these three approaches meets serious problems, the question now is which project is the most promis-

---

[16]For this kind of worry see Shafer-Landau (2005) and Wedgwood (2007).

[17]For a classic statement of the deflationary theory of truth see Horwich (1990).

ing in overcoming its problems and accommodating normativity. I am most pessimistic about naturalistic reductionism. Since Plato we are looking for definitions of normative concepts and these do not seem to be forthcoming. Naturalistic reductionism is, of course, an open conceptual possibility but its constant failure instigates a thorough pessimism suggesting that the project is probably futile. Thus, for example, Williamson (2000, p. 31) says that "[T]he pursuit of analyses is a degenerating research program." Hence, I think that naturalistic reductionism should be set aside.

We are left with the anti-reductionist approaches of nonnaturalism and expressivism. Now, each approach has its pros and cons and they both have to struggle with their serious respective problems. My own intuitions, though, incline me to side with expressivism because to my eyes expressivism appears to be a more promising project than nonnaturalism. Here are three quick reasons for thinking that expressivism is a more promising project than nonnaturalism. First, intuitively, I find the epistemological and ontological difficulties of nonnaturalism much more difficult to be resolved than the semantic difficulties of expressivism. Second, the theoretical virtues of expressivism strike me as more explanatorily fruitful than the respective merits of nonnaturalism. Third, expressivism promises to reconcile naturalism and normativity and this is alluring since naturalism seems to evade the epistemological and ontological problems of nonnaturalism.

However, all these three reasons appear to be entirely *question-begging* against nonnaturalism. These three reasons assume without independent argument that what is essentially at stake between nonnaturalists and expressivists can be decided in favour of expressivism. This unsupported assumption rests on the implicit adoption of a naturalistic methodological framework. Expressivists, as philosophers with naturalistic orientation, have a different methodological starting point from their nonnaturalist colleagues. As naturalists, in the light of a scientific world picture, expressivists have no qualms in rejecting pre-theoretical normativity intuitions inherent in normative language and thought and opt for *ontological parsimony*. In contrast, nonnaturalists show much more *theoretical conservatism* and are unwilling to revise pre-theoretical normativity intuitions that are inherent in ordinary normative language and thought. They are

inclined to think that we should not by any means take our pre-theoretical normativity intuitions so lightly and attempt to revise them.

In the light of the different methodological starting points of expressivists and nonnaturalists, all three reasons cited for sympathizing with expressivism collapse. First, for the nonnaturalist we should undergo the epistemological and ontological difficulties and somehow assuage or even defuse them (if we can) because they are implied by our pre-theoretical normativity intuitions and these intuitions should be respected. Second, given that we must respect our pre-theoretical normativity intuitions, the explanatory virtue of ontological parsimony appears to be more like a vice than a virtue. Ordinary normative language and thought seems to imply the existence of normative properties and we should respect this 'brute fact'. Third, granted that a naturalistic framework cannot sustain our intuitions about the existence of normative properties and these intuitions must be respected, we have good reasons to reject naturalism rather than nonnaturalism.

Of course, there might be good independent reasons for choosing one instead of the other methodological frameworks and I suspect that the naturalistic methodological framework might be ultimately defensible but this discussion cannot be opened here. It would lead us to a swirl of metaphilosophical discussion too complicated to be even touched here. The debate between naturalistic and nonnaturalistic philosophical methodology runs so deep that often breaks down and the two camps remain entrenched in their positions and simply proclaim that they have conflicting intuitions. Even at this deeper level, the question often cannot be adjudicated and the two camps agree only about their disagreement concerning methodological intuitions.

A fourth methodologically neutral reason might help us to put some weight in favour of expressivism. In theory-choice theory, fertility is considered to be a true theoretical virtue. That is, the capacity of a theory to be flexible enough to incorporate new insights, develop in new directions and thus increase explanatory power is considered to be valuable.[18] Thus, for example, we have Kuhn (1977) saying that fertility is of "special importance." Now, expressivism seems

---

[18]See for example Nolan (1999). The T. Kuhn quote is found in Nolan's paper.

exactly to bear this sort of theoretical virtue. It shows a remarkable ability to incorporate and combine a variety of philosophical insights and develop in new directions. It shows potentiality to increase explanatory power, overcome objections and the necessary flexibility to adapt to new challenges and conditions.

Of course, this is not to insinuate that the nonnaturalist tradition is an inflexible, barren and monolithic tradition that cannot incorporate new insights and develop in new directions. It is an approach that often does exactly that in order to respond to objections, worries etc.[19] But in any case, expressivism is a relatively newborn tradition in comparison with the nonnaturalist tradition (that goes at least back to Plato) and yet in short time it has been formulated in a plethora of ways and continues to incorporate new insights and develop in new directions in a rapid way that the nonnaturalist tradition, I think, cannot match. Since Ayer's crude emotivism was propounded, the expressivist approach has grown to a whole philosophical tradition with numerous variations. Thus we have Ayer (1946) "emotivism," Stevenson (1937) expressivism of "suggestive force," Hare (1952) "universalizable prescriptivism," Gibbard (1990, 2003) "norm-expressivism," Blackburn (1998) "Quasi-Realism," Ridge (2006, 2007) "ecumenical expressivism" and others. Further, the expressivist tradition was initiated by Ayer as an approach to moral concepts like 'goodness' and 'wrongness' but recently it has been argued by Field (2000), Gibbard (2003), Ridge (2007), Chrisman (2007) and others that it can also expand and cover epistemic concepts like 'knowledge' and 'justification'.

This wide array of expressivisms indicates that expressivism is a particularly flexible and fertile approach to normativity with prospects and potentiality. One example of this theoretical flexibility to develop in new directions is also shown, I think, in my current work on a naturalistic account of epistemic justification in terms of norm-expressivism. In the next section, I will laconically introduce the norm-expressivist account of epistemic justification and very briefly apply it to the species of *a priori* epistemic justification.

---

[19]For example, Wedgwood (2007) nonnaturalist position can be seen as a sophisticated attempt that incorporates new insights in order to respond to objections, e.g. Conceptual role semantics for 'ought' against the moral motivation problem.

V An Example: Epistemic Norm-Expressivism and A Priori Justification

As we have seen, noncognitive expressivism arrived at the contemporary philosophical scene with Ayer (1946) *Language, Truth and Logic*.[20] Ayer's proto-expressivist "emotivism" for moral discourse was meant to open some new logical space between the problematic naturalistic reductionism and nonnaturalistic anti-reductionism approaches by denying an unquestioned assumption that both approaches share: the referential semantics assumption and the concomitant factualist assumption. Ayer (1946, pp. 110–1) suggested that moral sentences like 'Stealing is wrong' are not referential sentences. They don't purport to refer to some 'wrongness' property in virtue of which stealing is wrong, but rather they are expressing feelings and sentiments of moral disapproval for the act of stealing. Their semantic function is 'emotive' rather than referential. It is to express our moral sentiments and feelings (disgust, shame, guilt, anger etc.) of moral approval or disapproval for certain actions, states of affairs, etc.

Ayer's exposition of expressivist 'emotivism' was more a brief suggestion thrown on the table in germ form than a worked out theory of semantics for moral discourse.[21] Unavoidably then, his suggestion was crude and unrefined in many serious respects. Very briefly, here are some of these crude respects. First, Ayer (1946, p. 110) conceded that moral sentences are not truth-apt and this seemed to abandon our objectivity intuitions that surround the notion of truth in any normative domain. As we have seen in section 3, worries about truth still beset even sophisticated forms of expressivism.

Second, Ayer didn't consider the repercussions expressivist semantics might have for formal logic. It didn't occur to him that he might need to devise expressivist semantics that could handle moral (and other) sentences embedded in unasserted truth-functional contexts in conformity with such logical notions as negation, validity,

---

[20]D. Hume is often considered to be a precursor of expressivism, but whether he is truly an expressivist is contentious and we need not entangle ourselves in scholarly interpretation and controversy here.

[21]To be fair with Ayer, he didn't intend a full-scale exposition of a theory for moral discourse. His emotivism appeared as a part of his general Logical Positivistic "elimination of metaphysics" project. As he acknowledges in the Appendix Ayer (1946, p. 191), "the theory is presented in a very summary way."

consistency and entailment. As we have seen in section 3, these repercussions came to surface with the formulation of the Frege-Geach problem. Third, Ayer didn't present any elaborated theory of the role emotions and feelings (shame, guilt etc.) play in coordinating our moral lives, regulate social co-operation, etc.

In the light of these shortcomings, the expressivist approach needed some serious overhauling if it was to survive infancy and make its way to maturity. This task was initially undertaken by Gibbard (1990, 2003) and Blackburn (1993, 1998) who attempted to resolve the semantic problems of expressivism with logic and truth, on the one hand, by developing expressivist semantics for logic and, on the other hand, by appealing to a deflationary theory of truth which could sustain truth-aptness in such an expressivist context. They also offered detailed analyses of how moral sentiments coordinate our moral lives, regulate social cooperation and interaction, etc., and how the roles emotions play could be understood in terms of evolutionary psychology

In particular, Gibbard's (1990, pp. 4–5) sophisticated "norm-expressivism" sets out searching for an understanding of what 'rational' means and how an understanding of the meaning of 'rational' could subsequently elucidate the species of moral rationality. What makes our moral choices to 'make sense' or 'to be wise'. His approach to the meaning of 'rational' is patently expressivist. He rebuffs referentialist efforts to either reduce rationality to a naturalistic property or to assert that there is an irreducible nonnatural 'rationality' property. Instead, he claims Gibbard (1990, p. 7) that "to call something rational is to express one's acceptance of norms that permit it." The meaning of 'rational' should be understood in terms of the noncognitive state of mind it expresses, namely, the nonreferential attitude of acceptance (or endorsement) to norms that permit (or govern) the content of the sentence expressed.

Now, my intention is to borrow Gibbard's norm-expressivist framework for moral discourse and apply it to epistemic discourse. Specifically, I will use it as an approach to the meaning of propositional sentences predicating 'is epistemically justified'. Gibbard's (1990) discussion of epistemic norms is scanty because his focus is on moral rather than epistemic norms, but I think the transition from applying his theoretical framework from moral to epistemic norms can be legitimately made.

Christos Kyriacou (University of Edinburgh)

The project of applying the norm-expressivist theoretical framework from moral to epistemic norms is motivated by certain key intuitions. First, it is an intuitive thought, I believe, to expect normative discourse to be understood and explained in a unified way and if moral discourse can be understood in expressivist terms, then there should be no special reason why epistemic discourse cannot be understood along analogous expressivist terms. At least prima facie, such an effort does not seem to be incoherent and given that the expressivist approach to moral discourse has proven to be fertile (tendentious, though, to the core), it seems that it would be both interesting and challenging to explore an expressivist approach to epistemic discourse as well.

Second, this intuitive thought is sharpened by the parallel drawn between moral and epistemic discourses in contemporary debates. In a parallel way, debates on epistemic discourse seem to follow the path that moral discourse had already followed and culminated to the conception of expressivism about moral discourse (recall the Sellars' quote in section II). Expressivism about moral discourse is an approach that has more recently arrived as a bold approach that promises to steer between the horns of naturalistic and nonnaturalistic approaches, take the best of both and at the same time eschew the worst of both. In the same way, expressivism about epistemic discourse is an approach that has more recently arrived as a bold approach that promises to steer between the horns of naturalistic and nonnaturalistic approaches, take the best of both and at the same time eschew the worst of both. Thus, epistemic expressivism draws inspiration from the application of expressivism to moral discourse, the gains it reaps and the theoretical virtues it displays in order to explore the same gambit to epistemic discourse.

Very roughly, according to this norm-expressivist approach, the semantics of propositional sentences predicating 'is epistemically justified' are expressivist and the state of mind expressed noncognitive — i.e., attitudes of norm-endorsement. That is, attitudes of approval (or endorsement) of the norms that permit (or govern) the content of the sentence expressed. When we use propositional sentences predicating 'is epistemically justified', we don't purport to refer to a certain natural or nonnatural 'epistemic justification' property. Rather we express nonreferential mental states, namely, noncognitive attitudes of endorsement for the norms we take to govern the content of the sentence expressed.

119

I intend to apply this norm-expressivist account to various species of epistemic justification but here I cannot go through all these species of justification. But I will say a few things about how this norm-expressivist account of the genus of epistemic justification is supposed to work for the species of *a priori* epistemic justification. Let me very briefly set the scene where this norm-expressivist account of *a priori* justification is meant to take place. There has been a lot of discussion lately on whether we can find logical space for the *a priori* and, especially, whether we can find such logical space in a broadly naturalistic framework. The heart of the problem lies in the fact that the existence of *a priori* facts seems to be inconsistent with a broadly naturalistic framework.

On the one hand, apriority seems irreducible to a broadly empiricist framework like naturalism because mathematical, logical, facts, etc., do not seem to be natural facts (or reducible to such facts). Such facts do not seem to be the intentional objects of our perceptual experience. On the other hand, however, if we attempt to understand apriority in a nonnaturalistic way, then we meet again the epistemological and ontological difficulties that we have seen in section 2: the "epistemic access" and "queerness" problems. Thus, it seems that we meet another dilemma: either drop naturalism and embrace nonnaturalism in order to accommodate the *a priori* or drop the *a priori* and nonnaturalism and embrace naturalism.[22] The first horn is the Platonic route. Mathematical, logical, etc., facts are eternal, immutable abstract objects beyond the flux of the spatio-temporal natural world. The second horn is the Quinean (1953) route. There is no *a priori*. All beliefs are in principle revisable in the light of "recalcitrant experience."

In the light of this juncture and the epistemological and ontological difficulties that beset the Platonic horn, some naturalists followed Quine's radical empiricism and proclaimed that we can live without apriority. Other naturalists attempted to break some middle ground between the two horns, followed the moderate empiricism of logical positivists like Ayer and identified the *a priori* with the analytic (that Quine (1953) famously criticized). For various reasons that cannot detain us here, these naturalistic accounts of the *a priori*

---

[22]Again the dilemma is not meant to be exhaustive but I will stick to this simplistic dilemma because, in the light of space and time limitations, it provides a crisp way to introduce the debate.

have been criticized and found elliptical. These options, however, are not exhaustive for the philosopher who would like to naturalize the *a priori*. We can embrace Gibbard (1990) Norm-Expressivist conceptual framework about moral discourse and apply it to epistemic discourse and *a priori* justification in particular. Epistemic norm-expressivism can steer between the horns of the dilemma and open logical space for a naturalistic account of *a priori* justification.

Let us now apply this account to propositional sentences predicating 'is *a priori* epistemically justified' and explain how this is supposed to work. Propositional sentences predicating 'is *a priori* epistemically justified' express the speaker's attitude of endorsement for the norms of the relevant discourse (logic or mathematics) she takes to permit the content of the sentence expressed. First-person sentences like "I believe that 1+1=2 is *a priori* epistemically justified" will express my endorsement of the norms of arithmetic discourse that I take to permit the content of the sentence expressed. First-person sentences that are, in some sense, self-attributed like "My belief that Socrates is either alive or not alive is *a priori* epistemically justified" both attribute the sentence to myself and express an attitude of endorsement for the norms of logic discourse that I take to permit the content of the sentence expressed.

Second-person attribution of sentences like "Your belief that $1 + 1 = 2$ is *a priori* epistemically justified" will attribute that certain sentence to someone and express an attitude of endorsement for the norms of arithmetic discourse taken to govern the content of the sentence expressed. Third-person attributions of sentences like "S's belief that Socrates cannot be both alive and not alive is *a priori* epistemically justified" both attribute the sentence to S and express an attitude of endorsement for the norms of logic discourse taken to govern the content of the sentence expressed.

Generalizing from these examples, we can say that propositional sentences predicating 'is *a priori* epistemically justified': (a) attribute the sentence (if it is an attribution) and (b) express an attitude of endorsement for the norms of the relevant *a priori* discourse (logic or mathematics) the speaker takes to permit the content of the sentence expressed.

This is, of course, telegraphic but sufficient to give a first very rough idea of how a version of epistemic norm-expressivism could be used to explore an account of *a priori* justification (and other species

of epistemic justification). To revert to our discussion on theory-choice theory, this theoretical fertility of expressivism is one of the methodological reasons that incline me to sympathize with expressivism rather than nonnaturalism. Two points should be stressed here, though. First, this conclusion should not be taken to be a head-on defence of expressivism as a theoretical approach to normativity. Perhaps, it is not even a sideways-on defence of expressivism, since I do not really propose solutions to its problems or discuss in depth and defend its substantial commitments. It is only to make a very modest point: stress a methodological theoretical virtue that expressivism bears and nonnaturalism, I think, does not.

Yet even if this modest point is correct, it gives us only *a* good reason for pursuing expressivism and in the light of other countervailing reasons this reason might not be good enough. For example, if expressivism cannot surmount its semantic problems with validity and truth, then although it is an interesting theoretical project it *must* fail because these semantic problems amount to a *reductio*. Any theory that cannot accommodate such basic semantic notions is doomed to failure. Second, as I have already stressed, I don't mean to imply the unwarranted dogmatic view that nonnaturalism is not an approach to normativity worthy of pursuing. Far from it, nonnaturalism is a venerable approach with its own merits and might after all be the right approach to normativity.

One way or another, as philosophers we have to rely on our intuitions for a methodological starting point for any inquiry because these are the only tools we have at our disposal. Intuitions, even when they are proven deceptive, they are proven to be so in the light of other intuitions. Thus, expressivists and nonnaturalists alike are entitled to cautiously follow their intuitions having always in the back of their head the disquieting sceptical worry that these might be deceptive.

VI Conclusion

I have summarized some of the problems that plague three of the main theoretical approaches to normative concepts: naturalistic reductionism, nonnaturalistic anti-reductionism and expressivism. I expressed my pessimism for naturalistic reductionism and carried on to examine which approach of the remaining two appears to be most promising: expressivism or nonnaturalism. I expressed my initial

sympathy for expressivism, but the reasons for this initial sympathy were found entirely question-begging. These reasons rested on the contentious implicit endorsement of a naturalistic methodological framework that nonnaturalists do not adopt. Nonnaturalists endorse a different methodological framework from their naturalists' colleagues that insists on the logical priority of our pre-theoretical normative intuitions. I conveniently skipped the metaphilosophical question which methodological framework is the most promising and, instead, provided a fourth methodologically neutral reason that seems to support expressivism.

I contended that expressivism is a relatively newborn tradition in comparison with the nonnaturalist tradition (that goes back to Plato) and argued that, despite its relative youth, it seems to exhibit an important theoretical virtue that the rival nonnaturalist tradition does not match: theoretical fertility. That is, the remarkable flexibility to incorporate new insights, develop in new fertile directions and therefore increase explanatory power. An example of this ability to incorporate new insights and develop in new directions was given from my work on a norm-expressivist account of epistemic justification as this applies to *a priori* justification. I concluded with the suggestion that, as philosophers, we have to rely on our intuitions for the methodological starting point of an inquiry and, since our intuitions in the light of other intuitions often prove misleading, expressivists and nonnaturalists alike are entitled to cautiously pursue the projects they find more appealing.

References

Ayer, A. J., 1946. Language, Truth and Logic. Penguin Books.

Blackburn, S., 1993. Essays in Quasi-Realism. Oxford University Press, New York.

Blackburn, S., 1998. Ruling Passions. Clarendon Press, Oxford.

Chrisman, M., 2007. From Epistemic Contextualism to Epistemic Expressivism. Philosophical Studies 135 (2), 225–54.

Davidson, D., 2001. The Folly of Trying to Define Truth. In: Lynch, M. (Ed.), The Nature of Truth. MIT Press, Cambridge, Mass.

Field, H., 2000. Apriority as an Evaluative Notion. In: Peacocke, C., Boghossian, P. (Eds.), New Essays on the A Priori. Oxford University Press, Oxford.

Frankena, W., 1939. The Naturalistic Fallacy. Mind 48, 464–77.

Frege, G., 1977. Logical Investigations. Yale University Press.

Geach, P., 1972a. Ascriptivism. In: Logic Matters. Blackwell, Oxford.

Geach, P., 1972b. Assertion. In: Logic Matters. Blackwell, Oxford.

Gibbard, A., 1990. Wise Choices, Apt Feelings. Clarendon Press, Oxford.

Gibbard, A., 2003. Thinking How to Live. Harvard University Press, Cambridge, Mass.

Goldman, A., 1992. What is Justified Belief? In: Liaisons: Philosophy Meets the Cognitive and Social Sciences. MIT Press, Cambridge, Mass.

Hare, R. M., 1952. The Language of Morals. Oxford University Press, New York.

Horwich, P., 1990. Truth. Blackwell, Oxford and Cambridge, Mass.

Kuhn, T. S., 1977. The Essential Tension: Selected Studies in Scientific Tradition and Change. University of Chicago Press, Chicago and London.

Kyriacou, C., Unpublished. Plato, Moore and 'Ethical Nonnaturalism', unpublished Manuscript.

Lewis, D., 2002. New Work for a Theory of Universals. In: Kim, J., Sosa, E. (Eds.), Metaphysics. Blackwell.

Mackie, J. L., 1977. Ethics: Inventing Right and Wrong. Penguin.

McDowell, J., 1994. Mind and World. Harvard University Press, Cambridge.

Moore, G. E., 1903. Principia Ethica. Cambridge University Press, Cambridge.

Nolan, D., 1999. Is Fertility Virtuous in its Own Right? British Journal for the Philosophy of Science 50, 265–82.

Papineau, D., 2007. Naturalism. In: Zalta, E. N. (Ed.), The Stanford Encyclopedia of Philosophy. Available online from `http://plato.stanford.edu/archives/win2008/entries/naturalism/`.

Putnam, H., 1978. Meaning and Moral Sciences. Routledge, London.

Quine, W., 1953. Two Dogmas of Empiricism. Harvard University Press.

Ridge, M., 2006. Ecumenical Expressivism: The Best of Both Worlds? In: Oxford Studies in Metaethics, Vol.2. Oxford University Press.

Ridge, M., 2007. Epistemology for Ecumenical Expressivists. Proceedings of the Aristotelian Society LXXXI, 83–108.

Ridge, M., Fall 2008. Moral Non-Naturalism. In: Zalta, E. N. (Ed.), The Stanford Encyclopedia of Philosophy. Available online from `http://plato.stanford.edu/archives/fall2008/entries/moral-non-naturalism/`.

Russell, B., 1912. The Problems of Philosophy. Home University Library.

Sellars, W., 1956. Empiricism and the Philosophy of Mind. In: Feigl, H., Scriven, M. (Eds.), The Foundations of Science and the Concepts of Psychoanalysis, Minnesota Studies in the Philosophy of Science, Vol. I. University of Minnesota Press, Minneapolis, MN, pp. 127–196, reprinted in Sellars (1963).

Sellars, W., 1963. Science, Perception and Reality. Routledge & Kegan Paul Ltd, London.

Shafer-Landau, R., 2005. Moral Realism. Oxford University Press.

Shoemaker, S., 2002. Causality and Properties. In: Kim, J., Sosa, E. (Eds.), Metaphysics. Blackwell.

Sosa, E., 2001. Epistemology and Primitive Truth. In: Lynch, M. (Ed.), The Nature of Truth. MIT Press, Cambridge, Mass.

Stevenson, C., 1937. The Emotive Meaning of Ethical Terms. Mind 46 (181), 14–31.

Wedgwood, R., 2007. The Nature of Normativity. Oxford University Press, Oxford.

Williamson, T., 2000. Knowledge and its Limits. Oxford University Press, Oxford.

The Role of Explicitness in Assessing Practical Reasoning

Andrea Lechler (University of Reading)[*]

a.lechler@reading.ac.uk

I   Introduction

Many of our decisions are made on the basis of conscious practical reasoning, i.e. thinking about how to act. Usually such conscious practical reasoning is an integral part of our flow of activity and we act on its results without questioning it. Yet sometimes the correctness of such reasoning is cast into doubt, either before or after the result has been put into practice. This may be done, for instance, by a negative gut feeling about the resulting decision, by a negative evaluation of the consequences of acting on the reasoning, or by the disagreement of others. In such cases re-assessment of one's reasoning is called for.

Two kinds of assessment need to be distinguished here. First, one can merely assess the *outcome* of one's reasoning, for instance

through reasoning again to see whether one arrives at the same conclusion. Second, one can assess an entire *process* of reasoning, which is particularly important if one's aim is to reason better in future and to avoid repeating one's mistakes. The latter kind of assessment is the main focus of the present paper. Its approach to the issue is based on the idea that in order to adequately assess a process of practical reasoning one needs to first make it more explicit. The paper offers an account of how instances of practical reasoning can be made more explicit and how the resulting explicit versions can be assessed.

Section 2 presents a first step that can be taken in trying to make one's reasoning more explicit. This consists in expressing the conscious thoughts which go into one's reasoning in words. Section 3 argues against the view that such verbalized practical reasoning should be assessed as a classical logical argument. It suggests that one needs to make one's practical reasoning more explicit before one can go on to assess it. Thus section 4 discusses how unconscious attitudes or habits which play a role in one's practical reasoning can be made explicit. However, it argues that the result is not a sufficient basis for assessment either.

Since Elizabeth Anscombe claims to offer an alternative to the view that explicit practical reasoning should be assessed as a classical logical argument, section 5 turns to her work on practical reasoning. Anscombe agrees that making an instance of practical reasoning explicit involves pointing out what logical relations or truths were involved. Yet she holds that the same logical relations can play a role in different kinds of reasoning. Therefore making an instance of reasoning more explicit also needs to involve specifying its starting-point and the special order or form in which the logical truths are arranged. The latter can be understood as the actual reasoning *process* taking place. Starting-points, on Anscombe's account, are not considerations in this process of reasoning, but are rather what gives the process its goal and direction. This means that in addition to assessing whether the attitudes or logical relations involved in an instance of practical reasoning are justified or correct, one also needs to assess whether the reasoning process is appropriate given the starting-point.

Finally, section 6 uses David Marr's distinction between three levels of explanation in information-processing tasks to put Anscombe's

analysis of practical reasoning in touch with recent work in cognitive science and artificial intelligence. It argues that once one acknowledges that different logics can be appropriate for different kinds of reasoning, more possibilities for assessment are opened up. For instance, one can assess whether the chosen logic is most appropriate given the starting-point of the reasoning. Furthermore it can be argued that the actual reasoning process may be more 'formal' than Anscombe admits. This would mean that there are more ways of assessing this process than envisaged by her.

II   Putting it in words

Consider the following case, which will be used to illustrate various points made throughout the paper: It is 7pm. Rita has just had her dinner and realizes that she has not yet had her daily exercise. She decides to get some exercise now and starts thinking about what exactly to do. After considering various options she decides to go to the gym. Yet, when she tells her flatmate about her decision, he suggests that there must be something wrong with her reasoning because she always regrets going to the gym. Rita takes his advice and starts assessing her reasoning . . .

It seems that in order to bring this reasoning into a form in which it can be assessed systematically as a whole, the train of thought which constitutes the conscious part of the reasoning needs to be expressed in words.[1] This can be done inwardly, in speech, or in writing.[2] Having such a verbalized version of one's reasoning is especially important if the assessment is undertaken in cooperation with other persons, or if one wants to make one's assessment accessible to others. But it seems also necessary if one only wants to assess it for oneself. This step is easier if some of the reasoning is already conducted in inner speech, and thus expressed in words.[3]

What need to be made explicit in this way are both the contents of the thoughts one had and the aspects under which the contents were considered or entered one's mind. There is, for instance, a clear

---

[1] I will not discuss here the possibility of actions being the conclusion, and thus part of, a process of practical reasoning. I think that what I am going to say could easily be amended to allow for this possibility, but it would make things too complicated here.

[2] I will not consider here the question of whether other symbols could be used instead of words, and how their use relates to the use of words.

[3] Carruthers (2004) helped me to see this point more clearly.

difference between thinking of going to the gym as a possible option for what to do now and deciding to go to the gym, even though both thoughts have going to the gym as their content.

However, in trying to express one's thoughts in words one may face various problems. First, it can be very difficult to recall how one actually reasoned in a certain situation. The longer ago the reasoning took place, the harder it becomes to remember what was going on in one's mind. In particular, it becomes harder to distinguish one's actual reasoning from reasoning constructed in hindsight to explain or justify the conclusion. If one is seriously interested in assessing one's own reasoning, then such post-rationalizations are to be avoided.

Recall of one's reasoning is also difficult if the reasoning was part of a habitual chain of actions. Such reasoning is often very quick, does not have one's full attention, and is not conducted in inner speech.

A further problem, especially with reasoning that was not conducted in inner speech, is that one is confronted with a continuous train of thought. This makes it difficult to identify and describe individual thoughts or parts of the reasoning. One may even be unable to remember the exact order in which different thoughts entered one's mind. Moreover in many cases there are various possibilities for expressing thoughts in words, and it is not obvious which one to choose. It is even questionable whether thought contents which were not expressed in inner speech can always (or ever) be adequately translated into natural language terms. On the other hand this may not be a big problem in assessing one's reasoning for oneself because in that case words may just serve as pointers to one's thoughts, of whose contents one still has an independent grasp.

It is also difficult to express the various aspects under which a thought content can be considered. One way of doing this, which is often chosen in the literature, is to distinguish different attitudes, such as belief, desire, or intention.[4] These can, but need not, involve a feeling. They can be positive (i.e. pro-attitudes) or negative. Although there may be disagreement about the meaning of certain attitude terms, this seems to be a common and useful way of expressing one's thoughts. However, the category of attitudes does not

---

[4] If the content of one's thoughts can be expressed in propositional form, then it may be more precise to speak of 'propositional attitudes'.

seem to cover all kinds of thoughts which may play a role in practical reasoning. Often the aim of practical reasoning (or reasoning in general) is precisely to arrive at an attitude regarding a certain issue. Thus many thoughts in practical reasoning put forward a candidate for action to be evaluated through reasoning, without expressing any fixed attitude toward it.[5]

Let met illustrate the foregoing by looking at a set of verbalized thoughts which could be an explicit version of Rita's conscious practical reasoning.

> I believe that I haven't had any exercise today.
>
> I intend to get some exercise now.
>
> Shall I go for a run? [I believe that one way of getting exercise is to go for a run.]
>
> I believe that it is 7pm.
>
> I believe that it is too late to go for a run.
>
> [I won't go for a run.]
>
> I fancy going for a swim. [Shall I go for a swim?]
>
> I believe that I have just eaten.
>
> I won't go for a swim.
>
> Shall I go to the gym?
>
> I believe that I haven't been to the gym for a few days.
>
> I intend to go to the gym.

This example shows that it is often not clear how to describe under what aspect a thought entered one's mind. Thus what is here introduced as a thought which puts forward the option of going for a run could possibly also be described as the belief that one could go for a run to get exercise. I chose to express the aspect under which a thought enters one's mind through phrases like 'I believe that ...' or 'shall I ... ?', but there are other ways of how this can be done. The example also shows how difficult it can be to individuate thoughts. I put in brackets the dismissal of the idea of going for a run because this may not have been an individual thought temporally or phenomenologically distinct from the belief that it is too late to go for a run.

---

[5]Unless one regards wondering, questioning, feeling uncertainty, etc. as attitudes. The parallel case in theoretical reasoning would be a proposition put forward to be assessed for truth.

III   Should verbalized conscious reasoning be assessed as a classical logical argument?

How does such an explicit version of our conscious reasoning help us in assessing our reasoning? A question often discussed in the philosophical literature in this context is whether such a verbalized version of correct reasoning must be a correct classical logical argument. If this was the case, assessment of one's conscious reasoning would require putting it in words and determining whether the conclusion is entailed (in the sense of classical logic, which is the sense of entailment considered in this section) by the conscious considerations that went into the reasoning. If the argument contained in the reasoning could thus be shown to be valid, one could go on to assess whether it is also sound. This would mean examining whether the premises are true or justified.

Let us try to apply this method of assessment to the first two lines of our example. It seems obvious that the sentence 'I believe that I haven't had any exercise today' does not entail the sentence 'I intend to get some exercise now'. One could object that the reasoning does not involve sentences or statements about attitudes, but the attitudes themselves. However, classical logic does not have room for the possibility of one attitude entailing another, it is concerned with entailment relations between propositions. Thus a last option would be to look for entailment relations between the propositional contents of different attitudes. Yet, that I have not had any exercise today does clearly not entail that I will get some exercise now. Similar results would be obtained if one tried to apply this method of assessment to the remainder of the example.

More extended arguments have been presented in the literature to make a similar point. However, it seems that the considerations presented here are sufficient to show that the example considered here cannot be assessed by treating it as a classical logical argument. There seem to be at least three possibilities for how to proceed from here.

First, it could be claimed that I just looked at the wrong kind of case and that in other cases verbalized versions of conscious practical reasoning *are* classical logical arguments. However, even if this was the case, it is not clear how it could help us with cases like the one just considered. As this appears to be a typical example of conscious practical reasoning, it seems legitimate to ask how it can be

assessed. Moreover, it may be that a correct account of assessing such reasoning will also provide an account for assessing other instances of practical reasoning. Such a unitary treatment of various kinds of practical reasoning seems preferable to an account which has to distinguish different kinds of practical reasoning at this stage of assessment.

A second option would be to search for other ways of making sense of verbalized conscious reasoning (i.e. other than seeing them as classical logical arguments), and for taking this as a sole basis of assessment. Yet I cannot think of any plausible such option, so I will not further pursue this path here. Instead I will explore a third path, namely going beyond the conscious parts of our practical reasoning before assessing it.

IV   Beyond the conscious

The previous sections only considered what is before our conscious mind when engaging in practical reasoning. However, much of our reasoning is a matter of unconscious mental processes or reasoning habits. In assessing our practical reasoning it is essential that we also try to make these aspects more explicit. Again a common way to do so is by pointing to attitudes – yet in this case not attitudes expressed by conscious thoughts, but attitudes which are not present in consciousness.[6]

What can be said about the nature of such attitudes? It seems that they can be either merely occurrent mental states, or long-standing mental states or habits. Such mental habits may precisely consist in engaging in reasoning of a certain kind.[7] For instance, the transition between raising the option of going to the gym, the belief that one has not been to the gym, and the intention to go to the gym may be mediated by a desire to see a particular fitness instructor. It may be difficult to determine whether this desire is an occurrent mental state or a long-standing attitude in the form of a habit of preferring options which contain a certain likelihood of seeing him.[8] This is because making implicit attitudes more explicit is complicated by various factors.

---

[6]The importance of this step is particularly emphasized by Anscombe (1957, 1979) and Brandom (1994).

[7]For an account of this kind see Brandom (1994).

[8]Note that attitudes which figure in one's conscious reasoning may also be an expression or instance of a long-standing mental state or habit.

A first problem is that in order to recognize the influence of a long-standing attitude one has to go beyond what is immediately given in the situation where the reasoning takes place and look at a larger period of one's life. Yet even having such wider knowledge does not guarantee that one can clearly identify the relevant long-standing attitude because it may be that several of one's long-standing attitudes could be seen as underlying one's reasoning, and it is also generally difficult to know what long-standing mental states or habits one has.

A further problem is the large involvement of unconscious mental activity in one's reasoning, and the difficulty of acquiring knowledge of one's unconscious mental life. The difficulty of gaining knowledge about unconscious mental processes sets a limit to how far one can go in making one's practical reasoning more explicit. A further limit to the process of making reasoning explicit is provided by pragmatic considerations about time and resources available in assessing one's reasoning.

A problem that comes up again in this context is the danger of post-rationalization. Given the difficulties just described, it may often be hard to avoid making an implicit attitude explicit in a way that one now thinks makes sense, rather than making explicit what was actually going on when one reasoned.

A critical question is also how exactly to describe the relevant attitudes. Moreover, many habits underlying our reasoning may not be describable as attitudes. Thus in some cases the best one can do may be to state between what kinds of thoughts or occurrent mental states transitions are habitually made in one's reasoning.

Of course, despite going beyond what is present in consciousness, speaking of such attitudes or habits still only describes the mental processes underlying one's reasoning at a high level of explanation. What mental mechanisms exactly are at work at a lower level is a further difficult issue, which cannot be considered here and which besides I think can be largely neglected in assessing one's practical reasoning from the reflective point of view (which is what this paper is concerned with).

A more extended treatment of the issue may be able to solve some of the problems mentioned so far, but for many of them it may not be possible to give a theoretical solution. Therefore it is important

to bear in mind that making one's reasoning more explicit is here not considered as an end in itself. Rather, it is a means to assessing one's reasoning, which in turn is a means to improving one's reasoning. In many situations where assessment of one's reasoning is crucial it may be better to have an inaccurate, incomplete description of one's reasoning than to have nothing at all to assess. It may even be that there are often different ways of making one's practical reasoning more explicit, and that in some cases it does not matter very much for the purpose of assessment which is chosen.

How does this further step in making a process of practical reasoning more explicit help us in assessing it? One possibility discussed in the philosophical literature is that the attitudes thus made explicit constitute the missing premises of a classical logical argument. However, this view has been argued against.[9] Rather than analyzing the arguments provided in the literature it will again suffice to apply this suggestion to our example. Rita could add to the belief that she hasn't had any exercise today a desire or goal to get physical exercise every day, or a more fundamental desire or goal to stay healthy (which can explain this more specific desire or goal). Yet the propositions stating that one has these two attitudes would still not entail the proposition that one intends to get some exercise now. As above the suggestion cannot be saved by looking at the attitudes themselves or their propositional contents.

It could again be objected that this is just the wrong kind of example. For instance, someone might suggest that if Rita's *only* way of getting exercise today was to exercise now, then the explicit version of the reasoning could be assessed as a classical logical argument. Although there seem to be good arguments for thinking that this claim is wrong,[10] these will not be examined here. This is because even if the foregoing was right, it would not help us in the current case; and our aim is again to find a method of assessment that covers more types of practical reasoning. The following section will look at Elizabeth Anscombe's treatment of practical reasoning, which offers a useful starting point for identifying such a method.

---

[9]E.g. Anscombe (1979); Brandom (1994, pp. 246ff).

[10]For instance, Anscombe (1979), Searle (2001) and Broome (2002) present such arguments.

V   Anscombe on assessing practical reasoning

Anscombe's account of practical reasoning has been, and remains, very influential in the literature.[11] And even though I will later criticize it, we will see that it contains some important insights. She stresses the point made in the last section that explicit practical reasoning of the type just discussed should not be assessed as a classical logical argument (no matter whether one looks at the attitudes involved, the assertion that a person has these attitudes, or the propositional content of these attitudes). Nonetheless she thinks that logic plays an important role in practical reasoning. According to her practical reasoning is grounded on certain logical connections between the propositions involved, which can be stated in the form of conditionals or 'hypotheticals'.[12] These 'logical truths' or 'falsehoods' are what she calls the premises or considerations which go into the reasoning. However, Anscombe holds that practical reasoning cannot be wholly accounted for by means of formal logic. The same logical truths can be put to different services in different kinds of reasoning, so formal logic is not sufficient to distinguish different kinds of reasoning. To illustrate this point Anscombe presents the following reasoning patterns. The first is meant to show the form of practical reasoning, the second stands for theoretical reasoning as usually thought of, and the third represents the search for an explanation.

> Wanted: that p. (Or: Let it be that p.)
>
> If q, then p.
>
> If r, then q.
>
> Decision: r!

---

[11] My exposition of Anscombe's views is based on her discussions of practical reasoning in *Intention* (Anscombe, 1957) and *Von Wright on Practical Inference* (Anscombe, 1979). The differences between the two treatments of the issue do not seem to matter in the context of this paper, so I will consider them as presenting one position and quote from both sources equally (although more attention will be paid to the more extended discussion in the later work).

[12] Anscombe (1979, pp. 124ff).

> r. (Or: Suppose r.)
>
> If r then q.
>
> If q then p.
>
> p.

> Given: p.
>
> If q, then p.
>
> If r, then q.
>
> To investigate: r. (Is 'r' true?)[13]

We can see that the same conditionals are involved in all three cases, but they are used in very different ways. First, they are used in instances of reasoning with different kinds of attitudes as 'starting-points' and conclusions. Anscombe's notion of a starting-point in reasoning, which she gets from Aristotle, is worth looking at in more detail. In the patterns of reasoning just quoted the starting-points are specified in the first line of each instance of reasoning. However, Anscombe makes clear that when she speaks of starting-points, she does not mean "a historical order of actual consideration."[14] Thus a starting-point is not the conscious consideration which the reasoner happens to start with. To understand the notion of a starting-point in reasoning better it will be useful to present Anscombe's view of 'wants', which she sees as starting-points in practical reasoning.

Anscombe's general view is that in order to act we always need a want that moves us to action.[15] Therefore practical reasoning cannot lead us to action unless there is a want which functions as a 'motor force for acting on the premises'.[16] The following quote illustrates this point:

> 'Dry food … suits anyone etc., so I'll have some of this' is a piece of reasoning which will go on only in someone who wants to eat suitable food. This is to say, it will at any rate terminate

---

[13] Anscombe (1979, p. 133).

[14] Anscombe (1979, p. 142).

[15] It is not entirely clear whether by 'wants' she means occurrent mental states or long-standing mental states or habits, or both. She does not seem to consider the distinction important for her purposes.

[16] Anscombe (1979, p. 115).

in the conclusion only for someone who wants to eat suitable food. Someone free of any such wish might indeed calculate or reason up to the conclusion, but leave that out, or change it to – 'So eating this would be a good idea (if I wanted to eat suitable food).[17]

Thus the function of wants in practical reasoning is very different from that of factual premises considered in instances of practical reasoning. Anscombe claims that "the wanting, the drive towards the end, does not properly go into the reasoning at all."[18] The same holds for statements or beliefs that one has the relevant wants. Only the content of a want may figure in the premises. For instance, the belief that I have not had any exercise today mentions exercise and thus points to my end of getting exercise.

A want is thus not one of many considerations in trying to solve a problem or answer a question; it rather sets up the problem or question to be solved or answered through reasoning. The task which practical reasoning is meant to perform is finding a means of satisfying a certain want, or reaching a certain goal. So what Anscombe seems to mean by a starting-point is what introduces the task to be tackled through reasoning and gives the reasoning its direction and force. If one conceives of reasoning as calculation (as Anscombe does), then the starting-point can be said to determine what is to be calculated. This means that the starting-point also determines what counts as or is accepted as a conclusion. Thus Anscombe says that specifying the end helps us see "where the reasoning should stop and a decision be made."[19]

Note that formulating the goal of practical reasoning as finding a means to one's ends accounts for the kinds of cases which Anscombe regards as typical of practical reasoning, namely cases where the conclusion is not a necessary means to one's end. Her conception of a starting-point may be slightly narrower than the one presented here. Nonetheless the account presented here seems to be in accordance with what she says and provides a unified picture of her account.

The notion of a starting-point also applies to types of reasoning other than practical reasoning. For instance, in the second reasoning

---

[17]Anscombe (1957, p. 66).
[18]Anscombe (1979, p. 144).
[19]Anscombe (1979, p. 116).

pattern above the starting point is the belief or supposition that r and the question of what follows from r in certain respects. The third reasoning pattern can be understood as starting with the fact that p and the question of how to best explain p.

Speaking of starting-points allows Anscombe to assign wants and other attitudes an important role in reasoning without having to accept the problematic position that the logical relations involved hold between these attitudes. Some such attitudes give instances of reasoning their direction, and others (or even actions) arise as a consequence of the reasoning and the attitude directing it.

A further sense in which conditionals or logical truths can be used differently in different kinds of reasoning concerns their order. Thus the conditionals 'If r then q' and 'If q then p' occur in different orders in the reasoning patterns above. Again Anscombe can be interpreted as not speaking of the temporal order in which the conditionals enter a person's mind in actual reasoning. At certain points of her discussion it seems as if she was mainly interested in practical reasoning as post-rationalization or justification of action in hindsight, rather than actual reasoning. On the other hand she speaks of practical reasoning as "leading to action."[20] And it seems to be actual reasoning resulting in actions which she is concerned with in discussing the case of a man who receives the command to bring one of several committees before an administrator.[21] Anscombe describes in detail the reasoning process he goes through in trying to obey the command and shows how various conditionals figure at different points in his reasoning. Thus her discussion highlights the fact that practical reasoning is a process taking place in time where the order of the logical truths involved plays an important role.

Through keeping strictly separate the questions of what logical relations are involved and what reasoning process takes place, Anscombe can accommodate for the fact that reasoning takes place in time while still allowing an important role for logic. This is not possible on accounts of reasoning where the reasoning process is equated with a logical inference or argument.[22] So Anscombe can hold that different kinds of reasoning display different forms or pat-

---

[20]Anscombe (1957, pp. 60ff).
[21]Anscombe (1979, pp. 123ff).
[22]von Wright (1972) discussion, which Anscombe (1979) criticizes, is one such example.

terns, without having to resort to the idea of different kinds of logical inferences. This is because reasoning is seen as a process taking place in time and the nature or form of this process can be different for different kinds of reasoning.

What role does the idea of making reasoning more explicit play on Anscombe's account? What the previous sections left us with was a set of conscious and unconscious attitudes involved in an instance of practical reasoning. Such an explicit version of one's reasoning can help one identify the starting-point of one's reasoning, the logical relations involved and of course important parts of the actual reasoning process that took place. Identifying these different aspects can itself be seen as a further step in making the reasoning more explicit.

Where does this leave us with regard to the question of how to assess an instance of practical reasoning? How can making one's practical reasoning more explicit in the way just described help one in assessing it? It seems that having such an explicit version at hand enables one to assess various aspects of an instance of practical reasoning.

One can first ask whether the conditionals or logical relations going into the reasoning are true. In cases where no conditional is mentioned explicitly one needs to assess whether what is regarded as a means really entails the end.[23] Second one can assess the various attitudes initiating the reasoning or contributing content to the reasoning. In the case of practical reasoning an important question is whether the thing wanted is in fact desirable and in accordance with one's other ends.[24] Third one can examine whether the reasoning process makes sense given the starting-point, for instance whether appropriate logical relations are taken into consideration.[25]

---

[23]Of course, this relation can be made explicit through a conditional before assessing it. This would probably be Brandom's approach here if he were to acknowledge that the same inferential relations can play a role in both theoretical and practical reasoning. See e.g. Brandom (1994).

[24]Anscombe (1957, pp. 70ff); Anscombe (1979, pp. 144ff).

[25]A difficulty in this context is posed by reasoning habits which cannot be understood as attitudes. A transition according to such a reasoning habit may not display any obvious logical relations. It may have been formed on the basis of certain logical relations, but such cases cannot be captured by the present account of assessment. This issue would need much more elaborate treatment, for which there is unfortunately not time here. I will therefore focus on the kinds of cases discussed by Anscombe.

This shows that Anscombe's distinction between the starting-point of one's reasoning, the logical relations involved in one's reasoning, and the process constituting one's reasoning allows a further step in making one's reasoning more explicit and opens up various possibilities for assessing one's reasoning. However, as the following section will show, Anscombe does not fully use the potential of this distinction for the analysis and assessment of practical reasoning reasoning. She has a limited conception of what sorts of starting-points and logical relations can play a role in practical reasoning (or reasoning in general). And she does not pay sufficient attention to the actual process of reasoning because she thinks that it cannot be formalized. Both points are connected to her restricted view of formal logic.

VI   Beyond classical logic

In order to show how Anscombe's view can be built upon, it will be useful to point out the similarities between her analysis of practical reasoning and David Marr's discussion of information-processing tasks. Marr thinks that in order to understand the functioning of devices which carry out such tasks one needs to distinguish three levels of explanation. The first level describes the nature of the problem being solved. It states what is to be computed and why this is to be computed. Furthermore it presents constraints, or a logic, imposed on the process by this task setup. The second level of explanation specifies a way of representing the input and output of the process and an algorithm for the computations. The third level is concerned with the physical realization of the computational process.

Marr proposes this distinction in the context of his account of vision, but he takes it to apply to all information-processing tasks and it has in fact been very influential in different areas of neuroscience, cognitive science and artificial intelligence. Interestingly it seems that Anscombe's distinction between the starting-point of one's reasoning, the logical relations involved in it, and the process constituting the reasoning can be restated in terms of Marr's distinction. What she calls a starting-point can be seen as corresponding to the reason or goal for the computation taking place (i.e. Marr's 'why'). This starting-point leads to a certain description of what needs to be computed. The constraints which the computation needs to satisfy are on Anscombe's view provided by classical logic,

or whatever else Anscombe means by 'logic'.[26] What Anscombe describes as the 'form', 'pattern' or 'order' of the reasoning can be seen as corresponding to Marr's second level.

If my above analysis of Anscombe is right, then the way I assigned her starting-point and logic a place in Marr's model appears uncontroversial. However, suggesting that Anscombe has in mind an algorithm when she speaks of the special form or pattern of practical reasoning does not seem to be in accordance with her account. She explicitly states that the actual reasoning process cannot be formalized. This belief is probably the reason why the notion of a reasoning process, though clearly present in her work, does not get as much attention as the notions of starting-point and logic. Nonetheless Anscombe speaks of practical reasoning as 'calculation'.[27] Thus, if we focus on Marr's claim that level 2 is concerned with *how* the computation is done, then it seems legitimate to express Anscombe's distinction in terms of Marr's model.

On the other hand, at this point we can also start questioning Anscombe's account. Recent work in cognitive science and artificial intelligence suggests that practical reasoning of the type discussed by Anscombe *can* be formalized.[28] I will in the following use Keith Stenning and Michiel van Lambalgen's work in the psychology of reasoning as an example of a contemporary account of reasoning that goes beyond Anscombe's work.[29] Stenning and van Lambalgen show how human reasoning can be logically modelled once one adopts a more sophisticated view of formal logic. They argue that many types of human reasoning should be understood as closed-world reasoning, and thus as following a nonmonotonic logic. They suggest logic-programming as a kind of closed-world reasoning which provides a logical formalism and at the same time an algorithm to model certain kinds of human reasoning. Practical reasoning is not considered by them in much depth, but they indicate that the kind of practical reasoning discussed by Anscombe could be modelled in a similar way. This means that modelling practical reasoning in formal logic may require going beyond classical logic.

---

[26] Her conception of logic is not made entirely clear in the paper.
[27] E.g. Anscombe (1957, p. 79).
[28] For an overview see e.g. Thomason (2008).
[29] Stenning and van Lambalgen (2008).

Assessing the proposals of Stenning and van Lambalgen in detail would go beyond the scope of this paper. But I want to discuss briefly how they (or others) could reply to Anscombe's objections against the idea of formalizing practical reasoning. Anscombe's worries are summarized in the following passage:

> though there is a 'validity' of practical inference, it is not of a purely formal character. By that, I mean one that can be displayed by the use of schematic letters, such that any substitution instance of the forms so given will be valid. The restriction of subject-matter to future contingents may be formally characterizable. The restriction, which we have mentioned, on inferences to *bringing it about that p and q*, apparently cannot be.[30]

The last point needs some explanation. Anscombe is here speaking of cases where both 'if p, then r' and 'if p and q, then r' hold. If one's end is to bring about r, it could be said that all one needs to do is to bring about p. On the other hand, there may be situations where one nonetheless rightly chooses to bring about p and q. Anscombe's claim is that one cannot present formal logical conditions for when the second conditional is relevant and what choice of action is appropriate in such a situation.[31]

A first route that could be taken in trying to rebut Anscombe's claim is pointing to the possibility of nonmonotonic interpretations of conditionals. The problem she describes may be less likely to occur there because one cannot just add any condition to the antecedent to get a new true conditional. On the other hand, Anscombe's interpretation of conditionals does not seem strictly classical, so pointing to the possibility of nonmonotonic interpretation of conditionals cannot fully solve her problem. And even in nonmonotonic logic one may have deal with cases of the type discussed by Anscombe.

Therefore we need to look in more detail at the question of whether logical conditions for the relevance of considerations can be formulated. There seem to be good candidates for such conditions. For instance, conditions for the choice of a means to one's end may be that it fits in with one's other ends, or that it requires minimal effort. Good reasoning may choose the means that best

---

[30] Anscombe (1979, p. 134).
[31] Anscombe (1979, pp. 131f).

meets a range of such conditions. They may be involved in both helping one to choose one of several considered means, and in determining what means are considered at all. The latter question could be seen as the aspect of a reasoning process that is least amenable to formalization. Relevant thoughts often appear to just pop up in our minds. Yet certain conditions for choosing means together with an appropriately structured long-term memory may be able to account for this issue. Stenning and van Lambalgen suggest such a prominent role for the organization of long-term memory, though a more detailed account is not part of their project.[32] It would also go beyond the scope of the present paper to examine this issue in more detail. Nonetheless the foregoing considerations indicate that the possibility of formalization should not be dismissed to quickly.

When discussing these issues it is important to bear in mind that what is to be given a formal description is not just what is present in consciousness while reasoning. One of Stenning and van Lambalgen's central claims is that a lot of our reasoning is unconscious, so a lot of what needs to be modelled by formal logic are unconscious mental processes.

Some of Anscombe's qualms about the possibility of giving such a formal account of practical reasoning processes may be due to her definition of 'formal'. Recent work in, for instance, linguistics or cognitive science indicates that what has traditionally been conceived of as content as opposed to form may be given a formal characterization as well. In fact, this is one of the key points of Stenning and van Lambalgen.[33]

Obviously it is difficult to fully model an actual decision making process because so many factors go into it and one would need to model all the knowledge a human being possesses. But it may still be possible to show that a formal model of the kind of processing that is going on in practical reasoning can be given and that certain reasoning processes can in theory be given a formal description.

Note further that even if Stenning and van Lambalgen are right and certain kinds of reasoning can be modelled using formal logic in the way they describe, this does not show that people really follow the relevant logic when reasoning in this way. Nonetheless there seems to be a strong presumption in favour of structured activity,

---

[32]Stenning and van Lambalgen (2008, pp. 238f).

[33]See e.g. Stenning and van Lambalgen (2008, p. 18).

not the least because learning can be more easily accounted for if one takes an activity to be structured. Therefore a formal account of practical reasoning processes seems preferable to one which only allows formal features in the form of logical constraints on reasoning processes. It could be objected that there is also the possibility of structured processes which cannot be described in a formal way. However, whether this is a serious possibility and what the nature of such processes would be cannot be examined here.

I said above that nonmonotonic logic may be more appropriate for modelling certain kinds of reasoning. Why is it that a lot of our reasoning seems to employ a nonmonotonic notion of consequence? On Marr's account what logic constrains a computation depends on the task to be solved. In Anscombe's terms this would mean that what logic is appropriate for a reasoning process is determined by the starting-point. This point is not obvious on Anscombe's account because she does not acknowledge that there is a range of logics which may constrain our reasoning, and that different starting-points may lead to different logics. Yet Stenning and van Lambalgen's distinction between a credulous and a sceptical stance in interpreting discourse (which they regard as a reasoning process) gives an example of different starting-points requiring different logics. They describe the distinction in the following way:

> Credulous reasoning is aimed at finding ideally a single interpretation which makes the speaker's utterances true, generally at the expense of importing all sorts of stuff from our assumed mutual general knowledge. Skeptical reasoning is aimed at finding only conclusions which are true in *all* interpretations of the explicit premises.[34]

Stenning and van Lambalgen argue that whereas classical logic may be the right sort of logic for sceptical reasoning, nonmonotonic logics are more appropriate for credulous reasoning. Although they mainly apply this distinction to discourse interpretation,[35] it is easy to see that a similar distinction can be made in the case of practical reasoning. Thus in planning what to do to reach a certain goal one can either assume that nothing abnormal will happen, or one

---

[34]Stenning and van Lambalgen (2008, p. 22).

[35]The cases they discuss are mainly people's interpretations of reasoning tasks presented to them in experiments.

can try to provide for every possible circumstance. In the first case nonmonotonic logic will be adequate, in the second case classical logic may be appropriate – though it is not clear whether one would ever come up with a plan in the latter case.

If the foregoing is right, then there are more aspects to be considered in assessing an instance of practical reasoning than those suggested by Anscombe. One can also assess whether both one's understanding of what is to be achieved through reasoning and the logic constraining the reasoning are in accordance with the starting-point. Furthermore one can assess the adequacy of the actual reasoning procedure. If one accepts that successful reasoning might follow an algorithm, then one could ask whether the right algorithm was chosen and whether it was followed correctly.

Making these aspects of one's reasoning more explicit and subsequently assessing them is again difficult. It is often hard to know what logic or algorithm one used, or even what exactly one's goal in reasoning was. Another apparent problem with this account is the question of how one can assess the choice of a particular logic. Assessing the appropriateness of a logic is also a type of reasoning, which follows its own logic. This raises the question of what justifies using one logic to assess another. Is there a hierarchy of logics, and in particular, is there one highest logic, with which all others can be assessed? Although I think there are ways of answering these questions which can make sense of the idea of assessing a choice of logic, the issue is far from trivial and a more thorough discussion of it will have to wait for another occasion.

VII Conclusion

The aim of this paper was to show in what ways an instance of practical reasoning can be made more explicit, and how having a more explicit version of the reasoning can help in assessing it. A first step in making one's reasoning more explicit was putting one's conscious thoughts in words. A second step consisted in expressing unconscious attitudes or reasoning habits which played a role in the reasoning in words. After each of these steps I asked whether it would be appropriate to assess such a more explicit version of a reasoning process as a classical logical argument. In both cases the answer given was negative.

I next showed on the basis of Elizabeth Anscombe's discussion of practical reasoning that logic can still be shown to play an essential role in reasoning. Yet this role can only be appreciated if one distinguishes the logical relations going into the reasoning from the starting-point of the reasoning and the actual reasoning process. The starting-point of an instance of reasoning can be seen as defining the goal of the reasoning and thus what is to be reasoned about. A reasoning process, on the other hand, is the performance of the actual reasoning. Thus making reasoning more explicit requires specifying a starting-point, logical relations, and the structure of the reasoning process. This means that one can assess the correctness of the logical relations considered in the reasoning, the attitudes involved in it (e.g. the starting-point), and the appropriateness of the reasoning process given the starting-point.

The last section showed how this account of making practical reasoning more explicit and subsequently assessing it can be extended if one adopts a wider conception of formal logic. In this case different starting-points can be seen as favouring different logics, and different kinds of processes are suitable for reasoning problems constrained by different logics. Such processes may even be given a formal description. This makes further kinds of assessment possible. Thus one can assess whether one's understanding of the problem and the choice of logical constraints are appropriate given the starting-point, and whether the reasoning process tackles the problem correctly. I could not show here that formal logic can really play the role it is assigned by researchers, such as Keith Stenning and Michiel van Lambalgen. However, their account seems prima facie plausible, so the different possibilities of assessment entailed by it should be taken seriously in discussions of practical reasoning.

References

Anscombe, G. E. M., 1957. Intention. Basil Blackwell, Oxford.

Anscombe, G. E. M., 1979. Von Wright on Practical Inference. In: Schlipp, P., Hahn, L. (Eds.), The Philosophy of Georg Henrik Von Wright. Open Court, La Salle, IL.

Brandom, R., 1994. Making it Explicit: Reasoning, Representing, and Discursive Commitment. Harvard University Press, Cambridge, Mass.

Broome, J., 2002. Practical Reasoning. In: Bermúdez, J., Millar, A. (Eds.), Reason and Nature: Essays in the Theory of Rationality. Oxford University Press, Oxford, pp. 85–111.

Carruthers, P., 2004. Practical Reasoning in Modular Mind. Mind and Language 19 (3), 259–78.

Searle, J., 2001. Rationality in Action. MIT Press, Cambridge, MA.

Stenning, K., van Lambalgen, M., 2008. Human Reasoning and Cognitive Science. MIT Press, Cambridge, MA.

Thomason, R., 2008. Logic and Artificial Intelligence. In: Zalta, E. N. (Ed.), The Stanford Encyclopedia of Philosophy. Available online from `http://plato.stanford.edu/archives/fall2008/entries/logic-ai/`.

von Wright, G. H., 1972. On So-Called Practical Inference. Acta Sociologica 15 (1), 39–53.

Perception, Causation and Justification

Graham Peebles (University of Glasgow)

`peeblesgraham@googlemail.com`

This paper is a brief discussion of McDowell's (1994) analysis of what he calls "brute impact" in perceptual experience, where "brute impact" is understood as a causal impingement of the world on the mind through perceptual experience. The threat that this apparently carries is that the causation renders our perceptual experience importantly non-rational. The competing picture, which can be characterized as Davidsonian, is that, in perceptual experience, there is a causal element. This can be read in two ways: either perceptions cause beliefs or the functional role of perceptions is played by beliefs which are themselves caused, but not by non-doxastic perceptions. Theories of this type are familiar from the literature.[1] I discuss and reject McDowell's motivations and arguments against a theory of this type.

I

McDowell (1994) presents two essential constraints on our beliefs, characterized as "internal" and "external constraints." The internal

---

[1] See in particular Roxbee-Cox (1971) and also Craig (1976) and Glüer (Forthcoming). Also see Armstrong (1968) and Pitcher (1971) for similar but importantly different theories. The characterization of these theories as "Davidsonian" is borrowed from McDowell (1994) and is only meant to refer to this aspect of a theory. Exactly what Davidson's theory of perception is, is rather unclear and I do not mean in any way to offer an exegetical account of it.

constraints are those constraints that hold on beliefs hold in respect of their rational relations to other beliefs. Beliefs, as exercises of our conceptual faculties are exercises of, as he puts it, "freedom," which can most simply be understood as requiring justification.[2] The external constraint is that which constrains belief "from the outside," a non-doxastic constraint. Were our beliefs to lack the first constraint, they would lack the necessary rational qualities. Were they to lack the external constraint, the constraint provided by perceptual experience, beliefs would lack empirical content. The concern here is with the former argument even though it is intermingled with the latter. This intermingling arises from the fact that perception plays a familiarly empiricist dual role: as well as providing the external constraint on our beliefs it also provides the empirical content.[3]

The demand for an external, a non-doxastic or empirical, constraint, is classically satisfied by the Given:

> when we have exhausted all the available moves from one conceptually organized item to another, there is still one more step we can take: namely, pointing to something that is simply received in experience. It can only be pointing, because *ex hypothesi* this last move in a justification comes after we have exhausted the possibilities of tracing grounds from one conceptually organised, and so articulable, item to another.[4]

However, this pointing towards the Given cannot suffice. Although an external constraint, the Given can only be identified as something bearing nonconceptual content (the invocation of the Given is not a conceptual move), rendering it inadmissible to satisfy the required role. The reason for this is that, as it is nonconceptual, it cannot constrain our thought in the required way: in being nonconceptual it simply cannot figure in any thought whatsoever; it would "be nothing to me." What this "being nothing to me"

---

[2]We could perhaps read "requiring" here as implying that it is a necessary condition on a belief that it be at least capable of being offered a justification, not that it be required to be conclusively justified.

[3]The notion of an "external" constraint is somewhat ambiguous in and of itself. It could either be a constraint that is mentally external or a constraint that is doxastically external. Sense-data, for example, could serve as the latter but not the former. This leads to the two different readings of "empirical" in the logical and metaphysical sense that McDowell requires to be satisfied.

[4]McDowell (1994, p. 6).

means of course, is that a nonconceptual content *could* "constrain" thought, but not in the right way. Such a nonconceptual would constrain on a sub-personal level, it would be constraint of "brute impact." Therefore, according to McDowell, "the idea of the Given offers exculpations where we wanted justifications."[5]

Thus the nature of the external constraint required is not non-conceptual, it is merely *nondoxastic*. The empiricist, foundation-alist (and internalist) quality of McDowell's view is clear, despite the reconceptualization of whatever now plays the functional role of the Given. Perceptions must be justificatory, and therefore must be within the space of reasons (and hence within the space of concepts), but they cannot, however, offer an *inferential* justification. Rather, the constraint they provide must be a constraint of non-inferential justification. Perceptions, therefore, play the classical *role* of sense-data or other empiricist constructs in this respect.

The alternative to this conceptualization of the Given (although that, strictly speaking, is not McDowell's proposal) would be "brute impact." Such "brute impact" would be an "exculpation," and, as such, could not play a justificatory role. What is clear from this preliminary is that the model of experiences as "brute impacts" is simply unacceptable. The reason, it appears, is that "brute impacts" are in some sense necessarily nonconceptual. The rational and the conceptual are, for McDowell, co-extensive and, if brute impacts are nonconceptual, then they are non-rational.

II

There are two connected worries in play here. One, if perception is non-rational, then it can only exert a non-rational constraint on belief. Two, if perception does not have "empirical content," neither can our beliefs. According to McDowell's disjunctivism we should deny the "highest common factor" object of perception in veridical and non-veridical perceptual experience (this should not be denied in the subjective epistemological sense).[6] What is denied is

---

[5]McDowell (1994, p. 8).

[6]This is, after Austin (1962), sometimes denied. However, it plainly should not be. Whatever the objects of hallucination are, they have at least one monadic property, that of being subjectively indistinguishable from at least one possible object of veridical experience. One may read the disjunctivist as proposing that there are no objects of perception in the non-veridical case, however this reading is unduly strong and, given the above concession, seems at the very least highly

the metaphysical equivalence of the two objects of experience, their shared "highest common-factor." If there is a metaphysically thick highest-common-factor-bearing object to veridical and non-veridical perceptual experiences, the question arises: what shared properties constitute the highest common factor? The answer usually offered is their intrinsic properties (think sense-datum). As the locution "highest common factor" suggests, these objects of indistinguishable veridical and non-veridical perceptual experiences would be interchangeable.[7] If perceptual experiences are of this type, then we are essentially unable, so the argument goes, to acquire genuinely empirical content (or, more clearly, unable to access the world). Whether or not the causal theory of perception *per se* does imply something like sense-data is a deep subject in its own right and somewhat orthogonal to the issue at hand.[8] Nevertheless, veridical perception must satisfy three criteria: it must be conceptual; it must be non-doxastic; it must not have objects of a type with non-veridical perceptual experience. The third desiderata comes apart from the other two for our concerns: conceptualized sense-data would satisfy the first two but not the third. The third argument pertains to the metaphysics of perception, the first two to the logic of perception.

McDowell's solution is that perceptions are takings in of the world, but takings in where our conceptual capacities are exercised passively, this is opposed to belief, where there is the taking of an attitude to the content. Thus experience satisfies the three criteria: perception provides direct metaphysical access to the world, is conceptual, and is doxastically external. In perception we are confronted with a *fact* in the world, a fact that perceptual experience makes available for belief, and, due to the inherently conceptual nature of the factually constructed world, our perception, although a conceptual activity, is an activity conducted *without recourse to an intermediary*. In taking up this perceptual "entitlement" (to be en-

---

implausible.

[7]This, of course, is a mistake in the metaphysical sense. It may be that they are interchangeable, but this does not follow from the possession of the monadic property alluded to in footnote 6.

[8]The causal aspects of perception, sense-data and disjunctivism are discussed originally in Hinton (1967, 1973) and illuminatingly in Robinson (1994) and Child (1994). Even if it could be shown that sense-data, or their correlates, were not implied by the causal theory, it seems that McDowell would still have the worries discussed below.

titled is not to act) to belief, the subject endorses the content of the perceptual experience, thereby performing a rational act. Were the perceptual nonconceptual, the subject could not perform the essentially rational act of endorsement, hence his belief could not be justified.

III

In perception we are supposed to be related directly, and without intermediary to the world. This, however is to be a passive exercise of our conceptual capacities. When McDowell says the following, he really means it: "*That things are thus and so* is the conceptual content of an experience, but if the subject of the experience is not misled, that very same thing, *that things are thus and so*, is also a perceptible fact, an aspect of the perceptible world."[9] On this picture, as I read it, the important aspect (aside from the direct realism) is that there is no causation involved in perception. The subject is merely related (non-causally) to the object via the relation of entitlement (entitlement is perception).[10] "Entitlement," as I understand it, is supposed to replace a more traditional intentional or representational theory of perception, although it is unclear exactly why. It is unclear, for example, why entitlement should be any metaphysically thinner than an intentional theory, or why the intentional theory requires a casual element.[11] As already pointed out it is also certainly not clear, to say the least, that a causal element of perception necessarily implies a perceptual "intermediary."[12]

---

[9]McDowell (1994, p. 26).

[10]Putnam (2002) points out, the relation between the fact *qua* the fact and the fact *qua* entitlement, that it would be unwise for McDowell to reify facts in the Tractarian sense, suggesting merely that McDowell holds that "the same proposition is fit to describe how it is in the world and the content of perception" (p. 183). This, however, seems to be too charitable. And, if McDowell is guilty of neutral monism – i.e., holding that the object of perception is the content of a mental state and there is no mental content/object distinction –, then the tilt into idealism threatens: see the first option in footnote 2.

[11]The intentional theory is compatible both with the rejection of the causal theory and with direct realism, as well as the opposite, or any combination of the two. Certain non-casual readings of the intentional theory can come to look rather like a version of disjunctivism. The modern representational theory does look committed to a causal element.

[12]See footnote 6.

These concerns can be set aside, however, for there is an independently successful argument against this position. Perceptual entitlements are supposed not to be judgings, but rather conceptual openness. In judging, one endorses the proposition featured in experience, where endorsing is understood as a conceptual move, and, as such, is apt for consideration as appropriate or not appropriate. What individuates perceptual entitlement from belief is that perceptions lack the quality of endorsement.

Recall Davidson's assertion that "nothing can count as a reason for holding a belief except another belief." McDowell sympathizes with this, but only to an extent: only a propositional state can be a reason; but it does not have to be a belief. Stroud (2002) clearly states the problem with this (the subject is justifying her belief that there is a square in front of her):

> Perhaps in saying 'Because it looks square' the person is referring to such an experience (a non-doxastic conceptual state). Its looking that way to her is certainly an experience she might have. But if she has such an experience, it must be true that that object looks square to her, and if that is the way it looks to her, she must be aware that that object looks square to her. Experience involves awareness. But for her to be aware that that object looks square to her is for her to accept or endorse as true the proposition 'That object looks square (to me).' She is aware that that proposition is true. So it cannot be that in receiving an impression that is an appearance there is no acceptance or judgement at all ... McDowell says that in sense-experience 'one finds oneself saddled with content ... (which is) available to one, before one has any choice in the matter'. But to be 'saddled' with a certain content in perception is not simply for that content to be 'available' to be entertained or contemplated, as it is in the unasserted antecedent of a conditional proposition for instance. To take in some content in perception is to have accepted or endorsed that content, or to find oneself accepting or endorsing it.[13]

This point is quite devastating for the distinction between passive exercise of conceptual capacities and non-passive exercises along the lines of only the latter constituting the taking of an attitude. If per-

---

[13]Stroud (2002, pp. 86–87).

154

ception must be an attitude of some kind, the notion of entitlement as a non-attitudinal state must be given up.

We can conclude that there are no compelling arguments to be adduced in favour of the view that perception is a *non-doxastic* conceptual state. Indeed, once the move has been made to conceptualize perception, the presentational nature of perception suggests that this second step follows quickly.[14]

IV

Perhaps, though, something has been missed. Here, it appears, is the nub of the concern:

> Quine speaks of facing the tribunal of experience, which seems to imply a vulnerability to rational criticism grounded in experience. But he conceives of experience as 'the stimulation of ... sensory receptors'. And such a conception of experience makes no room for experience to stand in rational relations to beliefs or world-views. The cash value of the talk of facing the tribunal of experience can only be that different irritations of sensory nerve endings are disposed to have different impacts on the system of statements a subject accepts, not that different courses of experience have different rational implications about what system of statements a subject ought to accept. In spite of the juridical rhetoric, Quine conceives experience so that it could not figure in the order of justification, as opposed to the order of lawgoverned happenings.[15]

And, later, in a footnote:

> A characteristic formulation (of Quine "trying to have it both ways") is this: 'The stimulation of his sensory receptors is all the evidence anybody has to go on, ultimately, in arriving at his picture of the word'. ... What goes on in arriving at one's picture of the world is not the stimulation of one's sensory receptors, experience as Quine officially conceives it, but how things appear to one, which belongs in a quite different conception of experience.[16]

---

[14]This assumes that there are independent arguments for conceptualizing perception. These are not offered or discussed here but assumed.

[15]McDowell (1994, pp. 132–133).

[16]McDowell (1994, p. 134–135).

The stimulation here is causal stimulation. My sensory receptors are stimulated because they lie at the end of a casual chain. At this point, we should sharpen our usage of "causal." We can then characterize something like the following law: *if conditions C obtain, then event E will occur.*[17] A full fixing of a particular specification of C would inevitably result in the occurrence of E. In this sphere of the world there is no room for rationality, as constitutively defined by freedom. This should be distinguished from another type of, mental-to-mental, causation. For example a subject's believing that p and also q can be the cause of his coming to believe r, but it is not the case that this is a causal explanation of the first type. It is non-nomological.

The stimulation of our sensory receptors are an explanation of the first type. Our sense organs, nervous system, and brain are physical objects subject to nomologically governed causes and effects. McDowell's underlying worry may be the following. If perceptions are to be identified with the results of causal chains such as these, then they are not apt to feature in thought, they are not rational. Recall that McDowell's disjunctivist theory of perception has no causation involved at all and note that, even though there will be some causation of this type involving light entering eyes etc, this will not *causally constitute* an experience according to his disjunctivism. There is no causation of this type at all in McDowell's account of the content of experience.

According to Davidson "the relation between a sensation and a belief cannot be logical since sensations are not beliefs or other propositional attitudes. What then is the relation? The answer, I think, is obvious: the relation is causal."[18] The argument from McDowell perhaps, then, amounts to this: as sensations are at the end of a causal chain, they are not apt to feature in rational explanations. And, if Davidson is to be taken at his word, as sensations cause perceptual beliefs, these perceptual beliefs are not apt to feature in rational explanation either. If the perception (sensation) is the result of a nomological cause, it cannot be conceptually articulated, and, if so, cannot serve to justify our beliefs. If it does offer a

---

[17]This is just very loosely formulated to capture the nomological necessity: C may include whatever one likes.

[18]Davidson (1986, p. 311).

constraint it offers a nomological constraint that would render perceptual belief nonconceptual. If sensations are epiphenomenal on certain casued beliefs that, in some way, play the functional role of traditional perceptions, then this argument applies equally well to them. The crucial premise here is that if a state or content is the result of a first-type cause it cannot be conceptual.

Whilst it may be true that it should be held that causal explanations of the first kind do not characterize rational transitions, perception is not a rational *transition*. In this sense it is not an act. The features pointed out by Stroud, and the antecedently necessary conceptuality of perception, are perfectly compatible with the perception being caused. To make this more clear, a state's being the result of a *transition* of the first causal type – i.e., nomologically governed, and, therefore in some sense, necessary – perhaps threatens the rational status of the state; however, the state being the result of a *non-transitional* causal process does not threaten its rational status. In this Davidson is correct. The fact that perceptions are at the end of a first-type casual nexus does not imply that they are not conceptual – i.e., within the spaces of reasons and concepts.

That the state or content stands at the end of this first-type casual nexus does not imply that it cannot stand as the first content apt for being taken up into a rational nexus. The subject can have a conceptually articulated perception that is (first-type) caused, or he can have a perceptual belief that is (first-type) caused by the non-conceptual perception.[19] If the underlying worry is that featuring at all in a causal nexus of the first-type compromises rationality and conceptuality, then this can be seen, perhaps, to be partly true in that the state could not be justified in a particular, direct, manner. That is to say, it could not be justified by appeal to a conceptually articulated mental content that featured in my psychology in a logically prior manner to the state requiring being justified. This, however, is unimportant: it may be sufficient for a state to be rational that it can be offered a justification, but it is not necessary.

---

[19]The former is preferable, but it is not the place to argue for this claim, only to argue against an objection that rules out both this and the latter claim. If one takes the points made by Stroud seriously, and one should, then this supports the theory that, in perception, beliefs are caused. Especially when considered in conjunction with McDowell's point regarding the questionable intelligibility of a non-conceptual mental state. Again, this is not the place to consider these arguments in detail.

On the McDowellian picture the subject's perceptual belief is held for a reason, namely his perception; he takes that perception as a reason, endorses it, and therefore cites that as his reason for his belief. On both the Davidsonian pictures outlined above, this is not true of the subject's perceptual belief. But this, however, should not be troubling as, on the McDowellian picture, the subject has no such justificatory reason, as outlined immediately above, for the undergoing of the perceptual state either. Indeed this is true of all at all plausible theories of perception: it may be that some higher-order reflection on perception yields some change in one's attitudes to the content, but it does not furnish the subject with a *reason* for or against the very *undergoing* of that state. If this is a worry, then the worry is clearly misplaced. Furthermore, the fact that there is no causation of the first-type in McDowell's picture does not eliminate this feature of perception: there is no reason of the type just adduced – i.e., a reason for the subject's perception – to be found in the subject merely standing in a particular relation to a fact. This appears to be equally as "brute" as is the standing at the end of a first-type casual nexus. A related worry may be the following. If perceptions lie at all in a first-type casual nexus, then the subsequent state must also lie in a causal nexus of this type. In other words, if the perception is so-caused, so must the perceptual belief, and so must any other state that stands in what is no longer a relation of rational transition. But this requires independent argument. It may, in fact, be true, given some premises regarding the metaphysics of the mental, but it does not follow from the fact that the perception is caused to the fact that any subsequent states the subject finds himself in are also so caused.[20]

V

In summation, I have objected to McDowell's arguments against the feasability of holding that a perceptual state, particularly when that is construed as a belief, could be the result of nomological cause. When McDowell says that pointing to the Given, if possible, could only offer an "exculpation" rather than a justification, we can see now that this is correct and is not a point against a theory. Not least because it is inherent in the foundational model that there be a

---

[20]See Kim (1993).

final content offered as ultimate justification and, therefore, a form of exculpation is implied. If perceptions are held to be a type of belief that is caused (perhaps perception can be partly individuated in terms of this), then it should not come as any more of a surprise that the subject should be exculpated for this belief than it does that a subject should be exculpated for his perceptual entitlement.[21]

References

Armstrong, D. M., 1968. A Materialist Theory of the Mind. Routledge, London.

Austin, J. L., 1962. Sense and Sensibilia. Oxford Univerity Press, Oxford.

Child, W., 1994. Causality, Interpretation, and the Mind. Oxford University Press, Oxford.

Craig, E., 1976. Sensory Experience and the Foundations of Knowledge. Synthese 33 (1), 1–24.

Davidson, D., 1986. A Coherence Theory of Truth and Knowledge. In: LePore, E. (Ed.), Truth and Interpretation. Perspectives on the Philosophy of Donald Davidson. Oxford: Blackwell, pp. 137–57.

Glüer, K., Forthcoming. Towards a Doxastic Theory of Perception. Mind and Language.

Hinton, J. M., 1967. Visual Experience. Mind 76, 217–27.

Hinton, J. M., 1973. Experiences. Clarendon, Oxford.

Kim, J., 1993. Supervenience and Mind. Cambridge University Press, Cambridge.

McDowell, J., 1994. Mind and World. Harvard University Press, Cambridge.

Pitcher, G., 1971. A Theory of Perception. Princeton University Press, Princeton.

---

[21]This would amount to a rejection of the criterion in footnote 2, as discussed in the penultimate section.

Putnam, H., 2002. Mcdowell's Mind and Mcdowell's World. In: Smith, N. S. (Ed.), Reading McDowell on Mind and World. routledge, London.

Robinson, H., 1994. Perception. Routledge, London.

Roxbee-Cox, J. W., 1971. An Analysis of Perceiving in Terms of the Causation of Beliefs. In: Sibley, F. (Ed.), Perception: A Philosophical Symposium. Methuen, London.

Stroud, B., 2002. Sense-Experience and the Grounding of Thought. In: Smith, N. (Ed.), Reading McDowell on Mind and World. Routledge, London.

Moral and Rational "Oughts": The Distinction between 'Demanding' and 'Recommending' Normativity

Peter Schulte (University of Bielefeld)

petedayton@gmx.de

Most theorists agree that morality and rationality are both normative: the moral claim "you ought to help others" is a genuine normative judgment, as well as the prudential maxim "you ought to save money for the future." But it seems that there is a crucial difference between these two kinds of judgments. As Kant puts it: "The maxim of self-love (prudence) just *advises*; the moral law *demands*."[1] Although this difference is quite obvious and fairly clear-cut, there is no widely accepted explanation of it.[2] In my paper, I will argue that an expressivist account of normativity (roughly along the lines of Gibbard (1990, 2003) has the resources to provide such an explanation.

---

[1] Kant (1788/1968, p. 36) (My translation.) See also Pink (2004) for a more elaborate description of the distinction between the demanding and the recommending force of normative judgments.

[2] Kant, of course, proposed such an explanation: He tried to account for the difference between demanding and recommending force by appealing to the difference between categorical and hypothetical imperatives. I do not think that his proposal works, but unfortunately, I cannot go into this discussion here.

I 'Demanding' and 'recommending' normativity

But first, let me characterize the crucial difference in an intuitive, 'pre-theoretical' way. Taking my cue from Kant, I will do this (i) by focusing on natural language classifications, and (ii) by examining our reactions to 'people who fail to do what they ought to do' (or 'trespassers', as I will call them from now on).

Consider (again) the moral statement "you ought to help others". It is natural to say of someone who utters this sentence that he is making a demand, i.e. that he is demanding that we help other people. (Imagine, e.g., a journalist writing an article on the latest speech of the Pope: He can report the Pope's statement "you ought to help other people" by writing "the Pope demanded that we help other people.") By contrast, it is not adequate to characterize judgments of rationality (like "you ought to save money for the future") as demands; instead, these judgments are naturally described as 'recommendations' or 'pieces of advice'. This is one clear difference between morality and rationality, and it seems to be directly related to the difference between 'ordinary' demands (like "Give me my money back!" or "Please, pass me the salt") and 'ordinary' recommendations (like "Don't order this meal, it's horrible," "Take the highway, then you'll be there in time!").

This also holds for the second intuitive difference, the difference that is manifest in our reactions to trespassers. When someone violates a moral principle that we accept, we get angry at him or resent him (emotional reaction), we denounce or condemn his action (verbal reaction), and we may even punish him for what he did (practical reaction). But when someone acts against one of our accepted principles of rationality, reactions like these seem totally inappropriate. We can (and do) criticize such actions as "irrational" or "foolish," but we do not (and should not) resent, condemn or punish someone for his irrationality. Again, there is a striking similarity here to the difference between ordinary demands and recommendations: We get angry at people who refuse to comply with our demands, we reproach them and may even punish them (not by sentencing them to prison, of course, but by refusing to speak to them again, for example). People who refuse to follow our recommendations are usually not treated that way; we merely call them "irrational," "foolish" or just "stupid."

162

These observations suggest two things. First, they suggest that that there is a clear intuitive difference between moral judgments and judgments of rationality, and that this difference is aptly characterized as "the difference between 'demanding' and 'recommending' normativity." But secondly, these observations also suggest (at least to my mind) that it might be a good idea to start our investigation of the normative difference between morality and rationality by analyzing the difference between ordinary demands and recommendations.

II   Ordinary demands and recommendations

An imperative like "Put your straw hat on!" can be used to make a demand or to make a recommendation. But what is the difference between these two speech acts? Standard speech act theory suggests that the difference lies (at least in part) in the *sincerity conditions* of these acts: A sincere demand requires a different attitude on the side of the speaker than a sincere recommendation.

If the utterance of "Put your straw hat on!" is a demand, standard speech act theory says that it is sincere if and only if the speaker *wants* the addressee to put her straw hat on. If the speaker does not have this want or desire, while the context makes it clear that her utterance is to be taken as a demand (request, order, command, etc.), then she is insincere. She is pretending to have an attitude that she really does not possess. This, I think, is quite plausible; so we can agree with standard speech act theory that the sincerity condition of a demand that p is the speaker's desire that p.

With recommendations, matters are different. It is surely not *sufficient* for the sincerity of a recommendation that I want the addressee to do what I have recommended; otherwise, all suggestions made for selfish purposes (like, e.g., the recommendation of a car salesman to buy a car from his company) would automatically count as sincere. Furthermore, it seems to me that it is not even *necessary* for the sincerity of a recommendation that the speaker has the desire in question: I may sincerely recommend some course of action to you while at same time not caring at all whether you follow my recommendation. (Usually, of course, I do care, at least to some minimal degree; it is only in special circumstances that a speaker is motivated to make a recommendation without being interested at all in what the addressee does. Nevertheless, I think that these special circumstances may sometimes obtain.)

So, what are the sincerity conditions for recommendations? According to standard speech act theory, the sincerity conditions consist in the speaker's belief that the recommended action is "good for the addressee," "in the addressee's best interest" or something similar.[3] If I recommend that you put your straw hat on, while not believing that this will benefit you in any way, I am insincere.

This sounds plausible. Nevertheless, there is a problem with this answer: It is not very illuminating. Remember that we turned to ordinary demands and recommendations in order to explain the difference between demanding and recommending normativity. But standard speech act theory analyzes ordinary recommendations as speech acts that have as sincerity conditions beliefs about the addressee's best interest – in other words, *beliefs about what it is rational for the addressee to do.* Ordinary recommendations are explained by beliefs that have 'recommendingly normative' character. On the standard account, therefore, we cannot explain recommending normativity by looking at ordinary recommendations: Such an explanation would be viciously circular.

To get out of this circle, I suggest an alternative analysis of recommendations. This analysis is not necessarily incompatible with the standard account, but, in my view, it reaches a deeper level of explanation. My proposal is this: The sincerity conditions of recommendations are not 'simple' wants or desires, but rather *conditional wants* – or, as we can also say, *conditional intentions.*

Take the case where I recommend you to put your straw hat on (maybe because the sun is shining and you would get a sunburn if you don't cover your head). It seems quite natural to say that my recommendation is sincere only if (roughly) *I would do the same thing if I were in exactly the same situation that you are in now.* To see this more clearly, imagine that I also carry a straw hat with me but do not put it on myself. In that case, you can reasonably ask why I refrain from putting my hat on, and I can respond by pointing out differences between our situations: I can say "My skin is not as sensitive as yours" or even "I like to get sunburns, but I know that you don't." But I cannot say: "Yes, for all that matters, I am in exactly the same situation that you are in, and I don't wear my hat – but what has that got to do with my recommendation?"

---

[3]Cf. Searle and Vanderveken (1985, p. 203); Bach and Harnish (1979, p. 48).

I cannot say this because the connection between my own behavior and my recommendation is intuitively clear to us: In some sense, my behavior in the imagined case 'contradicts' my recommendation, I'm guilty of "not practicing what I preach." In other words, my behavior suggests that my recommendation was insincere, because it shows that I do not have the right conditional intention (i.e. the intention to put a hat on if I am in exactly the same situation that you are in).

I conclude from these considerations that conditional intentions (or conditional wants) constitute the sincerity conditions of recommendations, while unconditional, 'simple' wants constitute the sincerity conditions of demands.

III   Two kinds of normative judgments

What do we learn from the considerations in the last section about the distinction between demanding and recommending normativity? In order to answer this question, it is helpful to reformulate our conclusions about ordinary demands and recommendations. In the last section, I have talked about the *sincerity conditions* of these speech acts. But the same theses can be formulated in terms of what mental states these speech acts *express*. According to the standard use of the term, a speech act A expresses a mental state M if and only if M constitutes the sincerity condition of A. Therefore, we can restate the conclusion of the last section in the following way: Ordinary demands express unconditional wants, while ordinary recommendations express conditional wants (conditional intentions).

Now, my thesis is that expressivists can use this insight about ordinary demands and recommendations to provide an analysis of the distinction between demanding and recommending normativity. The central point is this. According to expressivism, normative judgments are in an important respect like ordinary demands and recommendations: *they also express pro-attitudes or 'wants' (in the widest sense of the term).* Expressivists hold that by uttering a normative sentence like "one ought to help other people," we are not expressing a cognitive, belief-like state, but rather a pro-attitude towards some form of conduct (in this case, towards helping other people).

What the exact nature of these normative pro-attitudes is, and how they differ from ordinary wants, desires and intentions, is a matter of intense debate. Gibbard (1990) has characterized normatives

attitudes as "states of norm acceptance," i.e. states with specific tendencies towards action and avowal;[4] Blackburn (1998) has described them as complexes of first- and higher-order preferences, constituting a "spiral of practical and emotional ascent;"[5] and there are other proposals. I do not have time to consider these accounts in detail here – but fortunately, I do not need to.

The important point here is simply that, according to expressivism, normative statements express special normative pro-attitudes ('wants'). Regardless of how this "special" character is specified, expressivists can always give an illuminating analysis of the distinction between demanding and recommending normativity, based on the account of ordinary demands and recommendations developed in the preceding section: They can say that moral sentences express *unconditional* normative pro-attitudes, while sentences about rationality express *conditional* normative pro-attitudes – or, equivalently, that moral convictions are unconditional normative pro-attitudes, while judgments of rationality are conditional normative pro-attitudes.[6]

As it turns out, this analysis is independently plausible: If you have the (genuine) moral conviction that we ought to help people in need, then (roughly speaking) you will *want* everybody to help people in need; but if you judge that it is rational for us to save money for our future, you need not want everybody to save money for his or her future (indeed, it may be that you positively want that some people – your enemies or competitors – do *not* save money, although you think that it would be rational for them to do so.) All that is required for your judgment of rationality to count as genuine is that you would want to save money for the future *if you were in the situation of any other person.* And this requirement seems quite plausible.

I conclude, therefore, that the proposed expressivist explanation of the difference between demanding and recommending normativity is likely to be on the right track.

---

[4]Cf. Gibbard (1990, p. 75).

[5]Cf. Blackburn (1998, p. 9).

[6]These conditional normative pro-attitudes are similar to Gibbard's "contingency plans," cf. Gibbard (2003, pp. 48ff). To my knowledge, however, Gibbard never uses the difference between unonditional and conditional pro-attitudes to explain the difference between morality and rationality. Instead, he accounts for the difference by bringing in norms for emotional reactions, cf. Gibbard (1990, pp. 40–45).

References

Bach, K., Harnish, R. M., 1979. Linguistic Communication and Speech Acts. MIT Press, Cambridge, MA.

Blackburn, S., 1998. Ruling Passions. Clarendon Press, Oxford.

Gibbard, A., 1990. Wise Choices, Apt Feelings. Clarendon Press, Oxford.

Gibbard, A., 2003. Thinking How to Live. Harvard University Press, Cambridge, Mass.

Kant, I., 1788/1968. Kritik der Praktischen Vernunft (Critique of Practical Reason). Kants Werke, Akademie-Textausgabe, Vol. 5. Walter de Gruyter, Berlin.

Pink, T., 2004. Moral Obligation. In: O'Hear, A. (Ed.), Modern Moral Philosophy. Cambridge University Press, Cambridge.

Searle, J., Vanderveken, D., 1985. Foundations of Illocutionary Logic. Cambridge University Press, Cambridge.

Only a Chapter of Psychology? The Defense of Normative Naturalism in Epistemology

Giedre Vasiliauskaite (Erasmus University Rotterdam)

vasiliauskaite@fwb.eur.nl

I Introduction

Philosophers of different convictions generally agree that epistemology is a discipline that gives us some guidance as to how make some order in the realm of our beliefs. It investigates what it takes to have knowledge rather than a merely true belief. It aims at stating the criteria that define which beliefs deserve to be taken as cases of knowledge. In relation to that, epistemology also strives to give some advice as to how arrive at our beliefs in the most adequate way and how to improve our ways of reasoning. These features lead philosophers to think that epistemology is a normative enterprise, and that the concepts of knowledge and justification are normative concepts. The emergence of the naturalist accounts in epistemology is threatening to dissolve the normative epistemic dimension into the descriptive empirical research. In this paper I will claim that the normative dimension can be preserved also within the naturalistic epistemic framework if it is interpreted in terms of epistemic supervenience.

I will start my presentation with portraying the ways in which normativity has been spelled out traditionally, and distinguish the

various elements of normativity. This analysis will serve as the background for comparing and evaluating the claim whether, and to what extent, we can see naturalistic theories as preserving the possibility of normativity. Later on I will proceed with discussing the origin of the naturalistic project that appeared in Quine's article "Epistemology Naturalized." There I will analyze the criticism of the traditional normativity and then discuss Quine's reductivist conception of epistemology as a sub-discipline of psychology. In the third part I will question reductionist account and show that if we regard the structure of the relationship between natural and normative properties as a kind of supervenience, such an account would allow the possibility of normativity in naturalist accounts.

II   The Traditional Account of Normativity

The main normative notion of the traditional account is that of justification. Traditionally, epistemic normativity is articulated in terms of duty and the rules of permissibility. Because of that this account is also called deontological. A belief that satisfies the rules can be regarded as justified and, if correct, also qualifies for the case of knowledge. The power of justifiedness in this account stems from the internal mental states of the cognizer. This account can be called aprioristic, as a belief gains justification by virtue of the internal conditions or the logical relationship between beliefs, independently of what is happening in the external world. In other words, if two mental states are identical in two different external circumstances, the two beliefs are nevertheless regarded as justified. Such a status is guaranteed by the assumption that the contents of the mind are immediately accessible to the subject. On the basis of this aprioristic evidence s/he is in a position to possess good reasons that perform the justifying function in epistemic inquiry.

To sum it up, the main features of traditional normativity are apriority, deontology, immediate accessibility and the requirement for good reasons to justify a claim.

But does this account propose a relevant and encompassing solution to all possible epistemic claims? During the last decades a strong rival voice started to emerge, especially after it had been shown that justification, as it is formulated traditionally, cannot account for certain specific counterexamples (Gettier, 1963). It is doubtful that apriority and rational reflection alone could account

for the concept of justification and knowledge. And it can hardly be the case that epistemology has nothing to do with empirical research and psychology. The main basic notion in epistemology is that of belief, which clearly stretches into the realm of psychology.

III   Quine and the Emergence of Naturalsitic Epistemology

Aforementioned criticisms gave the motivation to look for other ways to formulate a theory of knowledge. Quine's project of naturalized epistemology arises with the claim that the traditional account of knowledge was a failure. According to him, epistemology based on apriority and rational reconstruction deludes us with the fake image of what the mechanism of cognition looks like and psychology could give us a "real" empirical answer to that question. Why does Quine think psychology is more entitled to give this answer? To quote him, "the stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology?"[1]

Quine's proposal is to see epistemology as a study of natural properties of cognition. As he writes,

> Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentary controlled input – certain patterns or irradiation in assorted frequencies, for instance – and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meager input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology; namely, to see how the evidence relates to theory, and in what way one's theory of nature transcends any available evidence.[2]

So, as we see, epistemology now is regarded as a natural science with psychological characteristics of human cognition and its interrelation with the external world. Contrary to traditional approach, in Quine's opinion, the naturalist understanding of cognition closely

---

[1] Quine (1969, p. 75).
[2] Quine (1969, pp. 82–83).

resembles the real epistemic practices. Reciting notorious Neurath's urge that we have to rebuild our own boat while being onboard, Quine admits that it is exactly our situation as well as the situation of natural science and epistemology as a part of it.

To sum it up, in traditional epistemology reflective accessibility was prioritized. That eventually led to an idea that epistemic justification should be carried out by virtue of presenting rational grounds of belief or rational reconstruction, as Quine puts it. Since this kind of accessibility-based awareness is no longer counted as a prerequisite condition of justification, this situation opens up the room for other conceptions, such as, for example, an analysis of unconscious sensory perception as an input and the causal mechanism of belief-formation.

What conclusions can be drawn from this explication of ideas? First of all, Quine's idea of naturalization supersedes epistemology based on logical reduction or rational reconstruction with epistemology being just a chapter of psychology and hence a part of natural science. This idea confronts the traditional apsychologistic approach with the recognition of the relevance of psychology and cognitive science. The question that arises here is such – is there any normativity left in this new epistemological project?

IV   Why Not to Settle for Reductionism?

Quine's naturalistic approach, putting epistemology in place of a sub-discipline of psychology raises the question about the possibility of normativity in this context. Is there any place for it left within the naturalistic theoretical framework, or is the present situation completely dominated by descriptive epistemology?

The first thing that comes to mind with respect to this issue is the fact that Quine definitely abandons internalist and aprioristic interpretations of normativity. That is clear from his claim that the redefinition of the concept of observation is intersubjective and that instead of studying internal reasons for taking a belief as justified the focus of epistemological inquiry should be replaced by empirical psychological research of the elements of human cognition that are not necessarily conscious states the subject would be aware of. Although for Quine this naturalistic enterprise still deserves the name of epistemology, its task is much more radical: in fact Quine urges to abandon the whole classical justification-focused epistemology and

supersede it by purely descriptive natural science of cognition. That is indeed a challenging idea that turns upside down the flow of epistemological inquiry.

As philosophers such as Kim (1988) or Kornblith (1985) have noted, Quine asks us to withdraw justification from naturalistic epistemic activities. What then happens to knowledge, which is still a concept closely related to justification and can be considered a normative notion? Should it also be abandoned altogether?

Kim (1988) argues that Quine's project of new epistemology does not leave any room for the concept of knowledge. Such an argument can be supported by the observation that Quine is talking about "science," "theory" or "representation" but never uses the concept of knowledge or justification. Empirical research studies the causal ties between sensory perception and beliefs formed on the basis of it. Kim rephrases Quine and is arguing that "It is none of the naturalized epistemologists' business to assess whether, and to what degree, the input 'justifies' the output, how given irradiation of the subject's retinas makes it 'reasonable' or 'rational' for the subject to emit certain representational output. His interest is strictly causal and nomological: he wants us to look for patterns of lawlike dependencies characterizing the input-output relationships for this particular organism and others of a like physical structure."[3]

The main concern of new epistemology is to determine descriptive causal connections without establishing the conditions explicated in terms of good reasons. Although Quine emphasizes that concept of evidence and its relation to theory remains crucial in epistemology, this concept is understood in a different way. Naturalistic evidence based on lawlike connections is relative to a mechanism of cognition. On the other hand, traditional concept of evidence aims at determining the internal factors that contribute to justification. Because of that Quine's new concept of evidence is nonnormative, therefore one can claim that such kind of naturalized epistemology dismisses the concept of normativity from epistemology.

Furthermore, one can also question whether such a project can still be called epistemology. Recognition of the significance of psychological studies is no doubt important but does it mean that epis-

---

[3]Kim (1988, p. 390).

temology has to pay such a high price as to lose one of its most important features? As Kim points out, "for epistemology to go out of business of justification is for it to go out of business."[4]

So what could be the solution – is normativity really out of place in naturalistic discourse? Philosophers' views concerning this issue are far from unanimous. Some of them adhere to Quine's point of view that naturalism leaves no room for normativity (Lammenranta, 1998), while others (Kim, 1988; Kornblith, 1985; Goldman, 1986) argue that also naturalized epistemology is able to retain its normative character. Along with the latter authors, virtue epistemologists (Sosa, 1991; Plantinga, 1993) also see epistemology as an essentially normative discipline. The question that needs to be answered now concerns the possibility of naturalistic normativity and the peculiarities of its structure.

What would be the motivation not to accept reductivist view? One possibility is that it is even difficult to see any role for epistemology in purely descriptive context, and that threatens the existence of the very discipline. Of course, by introducing the naturalist approach, epistemology can partly be seen as losing its autonomy, yet is not necessarily has to disappear from the map of philosophy. Furthermore, if to regard the discipline as purely descriptive, then it can hardly offer any criteria and guidance as to how we ought to reason. Reductionist claims that the only things that really exist are natural properties and that the concept of normativity is rather ephemeral – it has no real existence as it has no point of reference. Consequently, all we can do is to investigate the former and forget about the latter. But why not to try just the opposite? As soon as we start to generalize with the intent of providing some benchmark for detecting justified beliefs, we step into the realm of the normative.

V Supervenience and the Possibility of Normative Naturalism

It is indeed an important task of epistemology to study our ordinary epistemic practices, yet such an activity does not fully exhaust the task of epistemology. In other words, it does not follow directly that normative claims should be reducible to the basic cognitive properties. I want to argue that it is possible to develop a concep-

---

[4]Kim (1988, p. 391).

174

tion that would combine both naturalism and normativity. For that we have to take a closer look at the cognitive mechanisms by dint of which humans acquire empirical information and process it into mental states.

One more distinction has to be noted here: as in normative ethics, in epistemology there should be a distinction between conditions or criteria and descriptive properties on the basis of which basic epistemic terms gain their normative force. The term "supervenience" refers to a particular kind of relationship, which involves two kinds of properties: the basic ones, and the others that follow or supervene on the former. First of all, the derivative properties can only exist on the basis of the fundamental properties. If the basic property has certain characteristics then the derivative property must also have the same characteristics. To put it another way, the quality of the basic property is transferred to the derivative property. By virtue of that transfer the derivative property acquires its normative force. However, the derivative property is at different level with respect to the basic property. Consequently, a belief is justified because of the existence of a different, i.e., factual or naturalistic, kind of properties, but cannot be reduced to them. It is worthwhile to refer to the area of normative ethics where the normative definition of a good action supervenes on descriptive properties of this action. If two persons commit the same kind of action under similar circumstances in a similar context, if the factual descriptions of these actions coincide, then both of them are qualified as good. For example, in order to evaluate a car as good we need to refer to the cluster of different properties, like the power of the engine, the durability of its parts, functionality of design and so on.

Similar parallel can be drawn with respect to epistemic normativity of naturalistic kind. If we recognize that reliable beliefs are generated by truth-conducive cognitive processes, such as reliable vision, perception, reasoning, memory, then naturalistic normativity is gaining its epistemic authority by virtue of the reliable cognitive faculties. Just to make one clarifying point, reliability of the cognitive faculties is articulated trough the high ratio of true beliefs produced by a certain cognitive process. In that sense reliability or truth-conducivity becomes the central notion from which the normative power of the justifying property is derived. Such a transition is accomplished due to the supervenience relationship between the

reliable faculties and the normative properties.

One may ask whether the naturalist standards of justifiedness are subjective, as it is the case with the concept of supervenience in the philosophy of mind: there the mental properties supervene on the subjective experiences or qualia. It is worth noting that in epistemology both the basic properties and the supervenient properties are objectively generalizable, i.e. can be measured and publicly demonstrated. That guarantees that the naturalistic concept of normativity does not fall into the trap of the subjectivist justification.

VI   In conclusion

The analysis of the belief-forming mechanisms is supposed to elucidate what and how humans "ought to believe." However, when supervenience comes into play, we can have several advantages with respect to normative naturalism: on one hand, it bypasses the criticisms that Quine has expressed in address of the traditional account of normativity; secondly, it recognizes the benefits of the empirical research but at the same time is able to refute Quinean reductionist-minded strategy. Finally, and the most importantly, such an interpretation of the structure of normativity preserves the main general goals and the normative dimension in naturalized epistemology.

References

Alston, W., 1989. Epistemic Justification. Cornell University Press, Ithaca.

Bishop, M., Trout, J., 2005. Epistemology and the Psychology of Human Judgment. Oxford University Press, New York.

Gettier, E., 1963. Is Justified True Belief Knowledge? Analysis 23, 121–123.

Goldman, A., 1986. Epistemology and Cognition. Harward University Press, Cambridge, Mass.

Kim, J., 1988. What is Naturalized Epistemology? In: Tomberlin, J. (Ed.), Philosophical Perspectives, Vol. 2, Epistemology. Rodgeview, Ataradero, Calif.

Kornblith, H., 1985. Naturalizing Epistemology. MIT Press.

Lammenranta, M., 1998. The Normativity of Naturalistic Epistemology. Philosophia 26 (3–4), 337–358.

Plantinga, A., 1993. Warrant: the Current Debate. Oxford University Press, New York.

Quine, W., 1969. Epistemology Naturalized. Columbia University Press, New York.

Sosa, E., 1991. Knowledge in Perspective. Cambridge University Press, Cambridge, Mass.