



SWIF Philosophy of Mind Review  
<http://lgxserver.uniba.it/lei/mind/swifpmr.htm>

Symposium on *Understanding People:  
Normativity and Rationalizing Explanation*  
by Alan Millar

Edited by Patrizia Pedrini

Vol. 6 - No. 1 - 2007

<http://lgxserver.uniba.it/lei/mind/swifpmr/0620071.pdf>

# *SWIF Philosophy of Mind Review*

<http://lgxserver.uniba.it/lei/mind/swifpmr.htm>

An on-line publication of the  
*SWIF Sito Web Italiano per la Filosofia*  
(*Italian Web Site for Philosophy*)

<http://www.swif.it/>

ISSN 1126-4780

*General editor:*

**Luca Malatesti**

Department of Philosophy  
University of Rijeka (Croatia)

The copyrights of the articles  
belong to their authors.

## Contents

<i>Sensitive to Norms, Caused by Reasons</i> PATRIZIA PEDRINI.....	1
<i>A Précis of Understanding People: Normativity and Rationalizing Explanation</i> ALAN MILLAR .....	14
<i>The Normative Dimension of Thought and Action: Alan Millar on Interpersonal Understanding</i> MARIO DE CARO .....	24
<i>Rationality, Animals and Causality</i> SIMONE GOZZANO.....	30
<i>Extending the Space of Reasons: Comments on Chapter Four of Understanding People</i> ADRIAN HADDOCK.....	41
<i>Understanding People Keeping Up Standards</i> PAUL NOORDHOF.....	48
<i>Replies to Authors</i> ALAN MILLAR	
Reply to De Caro.....	72
Reply to Gozzano.....	74
Reply to Haddock.....	78
Reply to Noordhof.....	81



# Sensitive to Norms, Caused by Reasons

PATRIZIA PEDRINI

(University of Florence)

## I

Philosophy has frequently been concerned with the apparent fact that people often seem to conform their thinking and behaviour to norms of various kinds. In so doing, they are shown to be responsive to reasons for their thinking and actions. This has brought about an intense debate that has in turn focused on: 1) clarifying how this can happen, and what else must be true of human beings for this to be possible at all; 2) what kind of norms we are sensitive to and how they provide reasons; 3) what this implies in terms of our mutual grasping of our actions and thinking; and also 4) whether this is genuinely so, that is whether we are really sensitive to norms, or if this appearance has to count as another among the innumerable misapprehensions due to the epistemic limitations human beings are condemned to.

The discussion became heated since the early 20<sup>th</sup> century and the main questions to which such concerns gave rise have reverberated onto most of the branches of philosophy, e.g. philosophy of science, philosophy of language, metaphysics, epistemology, metaethics, to an extent that one might try to work out the genealogy of most recent philosophical ideas in light of the question of whether we are sensitive to norms, and how.<sup>1</sup>

In what follows, I'll try to illustrate how Alan Millar's *Understanding People: Normativity and Rationalizing Explanations* securely belongs to this long-lasting and articulated context of investigation. I'll mainly do this by focusing on 3). Accordingly, I'll first outline how the background issues have historically matured, and I'll then say how the book takes a stand on some of the core theses stemmed from that specific tradition. Although expressly concentrated on clarifying what understanding people is and how it involves sensitivity to norms and responsiveness to reasons in ways to be specified, the author formulates and defends an original set of claims that show an insightful, wide-ranging conceptual appreciation of the import and interconnection of problems that have arisen, and still persist, in most of the fields involved in answering the question of the nature of sensitivity to norms and responsiveness to reasons. This is undoubtedly to count among the many virtues of the book.

I wish to thank Alan Millar for having been willing to write a précis of his book, in which he presents with the usual clarity the main claims he defends in the book; the commentators - Mario De Caro, Simone Gozzano, Adrian Haddock, and Paul Noordhof - for their engaging discussions on them; and, once more, the author for responding. I am especially indebted to Luca Malatesti, the chief editor of this journal, for his thorough support to my editorial work; and to Alessandro Pagnini, for useful discussions about the historical development and the significance of the philosophical problems I touch upon.

---

1 See Burge (1992).

## II

In a number of remarks dating from his lectures in Cambridge in the nineteen thirties, Ludwig Wittgenstein suggested that the explanation of what one did was liable to be regarded *either* in terms of the reasons one had *or* through the causes that determined one's deeds.<sup>2</sup> Roughly, whilst investigation of the former was mainly based upon what a person said - as unrepeatable, individual, and semantic-laden as such an event can be - the latter was associated with laws governing the allegedly uniform realm of a meaningless nature and could be explored through scientific experiment. Although specifically concerned on those occasions for the purpose of making a case against the allegedly scientific status of psychoanalysis, Wittgenstein's reflections passed into the philosophical literature, along with others about the nature of psychological narrative in general, as opposed to explanatory discourses where no psychological terms figure - typically, the accounts of physics. Further discussions then followed (or continued to follow) on the nature of psychological language and its relation to the scientific one. These debates were mainly due to neo-empiricists, who, given their commitments to scientific methodology and scientific language, were fascinated by the project of unifying the epistemology of comprehension with that of natural sciences. Moreover, a contribution to the philosophical discussion came from logical behaviourism, fervently engaged in clearing philosophy of mind from overt and hidden traces of Cartesian "mentalism". All these interests enforced the long-term and to some extent still unresolved distinction between *reasons* on which an agent acts and thinks and *causes* for her action and thinking. At the level of epistemology, they ended up strengthening a parallel widely accepted philosophical intuition: that *understanding* one person's actions and beliefs in light of the reasons she has for doing and thinking what she does have to be completely different a business from the attempt at explaining the physical world via the ever-increasing charting of its causal texture, which is the distinctive enterprise of science. More specifically, the debate has tended to be so set up that it has appeared plain to some philosophers to think of the parallel distinctions between reasons and causes, on the one hand, and between understanding and explaining, on the other, as dichotomies: if one *explains* a fact or an event<sup>3</sup>, one is invoking or focusing on *causes* that brought about that fact or event, whilst if one invokes *reasons* why a fact or event has taken place, one is properly involved in attaining *understanding* of the fact or event under scrutiny. Some grounded the dichotomies on the specificities of either methodology, independently of its subject-matter; others presented them as depending on the nature of the subject-matters: invoking reasons is epistemically appropriate to *human* beings and their deeds, while causes are the hallmark of *physical* world. Thus, one can make intelligible human facts or events as regarded in terms of their rational status, while investigation of nature via causes is necessarily silent on rationality, and has in itself nothing to do with rational intelligibility. For the sake of coining catchphrases, nature is in fact arational. Interestingly enough, however, whilst an arational nature is appropriately accessed through causes, according to some philosophers rationality of human beings would instead prevent us from unpacking their intentions and beliefs in terms of causes for the behaviour they bring about, since the rational connection between them and the resulting action and thinking would fail to be causal.

One might ask, however, whether such dichotomies should be valid at all. Since the Scientific Revolution at least, we have learned that the mechanical, causal point of view is appropriate for the investigation of nature. Thus, that natural facts or events are *not* to be searched through in terms of their "reasons" is common sense. After all, *scire est scire per causas*. But what about human facts or events? Most ordinary stories about human behaviour and thinking invoke reasons why people act and think in the way they do. We certainly understand (or fail to understand) why they acted as they did by taking into account the reasons they put forward to make their own behaviour intelligible, and by assessing the quality of these reasons<sup>4</sup>. But *why could not these reasons also be*

2 For a useful discussion of these notes on Freud, see Wollheim & Hopkins (1982).

3 The issue whether causal connections are between facts or events isn't settled. See Garrett (2006).

4 An interpreter, or the agent herself, may find that what the agent has cited as a good reason for her action fails to be

*causes* for what they make intelligible? Why should the connection between reasons and actions expressed in everyday, folk-psychological parlance fail to be causal, and thus unable to provide an explanation of the resulting behaviour?

The rationale for maintaining that reasons *cannot* be causes for what they “rationalize”, though fail to “explain”, was taken to be illuminated by typically illustrative considerations like those that follow.<sup>5</sup> Suppose that I say that I have raised my arm in a street in the attempt to call for a taxi driver’s attention, since I needed to catch a train in a few minutes and I believed the fastest way to reach the train station was to go there by taxi. My gesture is made intelligible to observers by my expressed motivation and intention, together with my current beliefs about the location of the train station and time, so that everyone can understand why I did so. The fact is, however, that I might have acted differently. It is intelligible that I raised my arm in order to call for attention, given my motivation, intention and beliefs, but nothing in the account given suggests that either everyone or I in similar circumstances would necessarily choose the same line of action to reach this goal. That is, it is not necessary for me to act in this way in order to reach my goal. Were there to be a causal connection between my states of mind and my action, such necessity should instead hold, and a good explanation should exactly be able to account for that necessity.<sup>6</sup>

The last point was believed to be a crucial one (others will be clarified later) to do justice to the dichotomies between explaining and understanding. The impossibility of reading into human behaviour “covering laws”<sup>7</sup> as exact and predictable as for instance those formulated in physics, and under which the fact or event explained can subsumed, has been for a long time perceived as the main obstacle to crediting rationalizing accounts with the status of genuine “explanations”. In other words, a standard for explanation was already in play: it came from the “hard” sciences, and it set the demanding constraints for other accounts in search for a scientific status. Any genuine causal explanation must also prove to be nomic.<sup>8</sup>

There are in fact other vocabularies that equally apply to the same sequence of gestures, for instance the vocabulary of physics, which seems to meet the requirement of issuing law-like connections. Typically, we can describe the sequence of gestures I made to call a taxi in a purely physical language, where no intentional states figure. The account specifies every proximal (bodily, neurological, or physiological) causes of every resulting bodily movement I made; more notably, it connects the states that figure in the chain of events as constrained by empirical *laws* discovered through experience and upon which predictions can be issued. But how exactly can an explanation that invokes laws be valid?

Hempel (1942; 1965; 1966) developed the idea that genuine explanations are, or are reducible to, accounts of facts or events that conform to the so called “Deductive-Nomological” model (henceforth DN), currently used by scientists to explain the phenomena that are typical of their domain. The model had already been formulated by Popper (1934), and shared by Oppenheim (Hempel & Oppenheim, 1948). Hempel decisively contributed to make it formally explicit what the model amounts to by providing a logical recipe for satisfactory explanation. The recipe specifies the

---

the actual reason for her action.

5 See Anscombe (1957) and Wittgenstein (1953).

6 It lies beyond the scope of this introduction a survey on whether the nature of causality is best caught by a Regularity theory of a Counterfactual theory. See Mackie (1993), Garrett (2006, Ch. 4), and Fulford, Thornton & Graham, (2006, chapters 13-15). I am aware that the philosophical debate on the scientific explanations has significantly changes during the eighties from the way the topic was framed within the Humean-Popperian-Hempelian discussion: see Salmon (1990). For an almost complete analysis of recent discussions on the scientific explanation see Ladyman (2002). At the same time, the debate between neo-Wittgensteinians and naturalists showed to be little affected by this shift. As to the the new definitions of causality recently developed in the philosophy of science, some tries to solve the dichotomy reasons-causes through important reconceptions of causality, see Peruzzi (2004) and Gozzano, (2007).

7 Hempel (1942; 1965, 1999).

8 It has been questioned whether causality and “nomologicity” must necessarily go hand in hand. See Garrett (2006, ch. 4).

#### 4 Sensitive to Norms, Caused by Reasons

logical structure in which the sentences that comprise an explanation have to stand. Thus, the DN model aims to provide necessary and sufficient conditions against which all explanations that aspire to be scientific can be tested. The DN model is meant to set a standard whose soundness is, by virtue of its logical properties, independent of the subject matter. This latter feature is what guarantees that the standard can be used across disciplines (whether “of man” or “of nature”).

Hempel (1966) lists four conditions of adequacy of scientific explanations: 1) the explanations must be a valid *deductive* argument; 2) the explanans must essentially contain at least one *general law*; 3) the explanans must have *empirical* content; 4) the sentences constituting the explanans must be *true*. Once all the four conditions are satisfied, the explanandum follows from the explanans by *modus ponens*, e.g.:

(a) At any location, the pressure that the mercury column in the closed branch of the Torricelli apparatus exerts upon the mercury below equals the pressure exerted on the surface of the mercury in the open vessel by the column of air above it. (b) The pressures exerted by the columns of mercury and of air are proportional to their weights; and the shorter the columns, the smaller their weights. (c) As P erier carried the apparatus to the top of the mountain, the column of air above the open vessel became steadily shorter. (d) (Therefore), the mercury column in the closed vessel grew steadily shorter during the ascent. (...) The explanations just considered may be conceived, then, as deductive arguments whose conclusion is the explanandum sentence, *E*, and whose premise-set, the explanans, consists of general laws, *L1, L2, . . . , Lr* and of other statements, *C1, C2, . . . , Ck*, which make assertions about particular facts. The form of such arguments, which thus constitute one type of scientific explanation, can be represented by the following schema:

D-N:

*L1, L2, . . . , Lr*                      Explanans sentences

*C1, C2, . . . , Ck*

---

*E*                                      Explanandum sentence

Explanatory accounts of this kind will be called explanations by deductive subsumption under general laws, or *deductive-nomological explanation*. (The root of the term ‘nomological’ is the Greek word ‘nomos’, for law.) The laws invoked in a scientific explanation will also be called *covering laws* for the explanandum phenomenon, and the explanatory argument will be said to subsume the explanandum under those laws (Hempel, 1999, pp. 301-2).

This definition of what an explanation is, however, proved to be problematic under various respects.<sup>9</sup> One is particularly prominent: if an explanation needs to rest upon the validity of a general law, it is clear that no one could successfully explain facts or events that fail to be subsumed under covering laws. There are sciences that clearly do not make use of general, exceptionless laws, but which issue explanations by virtue of mere generalizations that are inductively inferred from statistic frequency and contemplate exceptions. For this reason, Hempel had to modify his claim

<sup>9</sup> For the sake of giving one example that will be not explored in the remainder of this introduction, it is not clear that the model can distinguish between accidental universal connections (like “All the men wear shirts”) and laws of nature (like “All men are mortal”).

and acknowledge that DN needed to be coupled with a Inductive Statistical model (IS). Here, the explanandum isn't deduced from a premise that comprise a general law, but rather *inferred* from an empirical regularity of events that follow other events.<sup>10</sup>

And what about the extension of the model to rationalizing explanations? These latter clearly don't rest upon anything like a covering law. For example, my intention to raise my arm to call for attention is not covered by a general law according to which any time one raises an arm, one is calling for attention. By the same token, any "science" that makes use of explanations which don't follow from exceptionless laws cannot be said to provide satisfactory explanations of what it purports to explain. Historians, for example, when issuing explanations of historical events, don't operate in the way prescribed by Hempel's DN four conditions. One justification why this must be so is clearly related to a commitment to human freedom.

Maybe the IP model fares better: after all, it may be thought, rationalizing explanations could owe their validity to empirical generalizations on what people tend to do in certain circumstances. But whilst the acknowledgement that sciences which involve explanations by reasons do not conform to DN was shared by both those who favoured Hempelian account and its detractors, some defenders of the opposing view objected to the extension of IP to them as well. It is in fact not to be thought that the assimilation between rationalizations and explanations was barred solely by those who defended the scientific methodology as yielding valid explanations. On the one hand, the champions of scientific explanations as the only model for genuine explanation tried to extend their model to other sciences, and to show to what extent that could be done (i.e., to what extent rationalizing explanations refer to laws or generalizations, and thus are nomic); on the other hand, the proponents of the insuperable and unique specificities of rationalizing explanations forcibly resisted the assimilation to some extent welcome by the former. These latter put forward arguments for demonstrating that there are differences *in kind* between the two types of explanations. Such differences made it appear unlikely any reconciliation between them. Among the former are those who independently opted for a methodological unity of all sciences, e.g. neo-empiricists like Hempel and naturalists in general, whilst among the latter are those who held that any disciplines which study man and his deeds are founded and justified in a completely different way from those which study nature, e.g. anti-naturalists and Wittgensteinians.

However, the debate on whether there must be unity or tension between the science of nature (*Naturwissenschaften*) vs. science of man (*Geisteswissenschaften*) has a long history indeed. It is conventionally traced back to Hume (1748) and developed by Comte (1830-42) and Mill (1863), who maintained the virtue of such unity, and to Dilthey (1883), who gave instead a (neo-Kantian) foundation to the difference.<sup>11</sup> In the analytic tradition the contrast had already reappeared in keeping with questions raised within philosophy of history. Collingwood (1940; 1946), for instance, suggested that historians' job isn't to explain, but to penetrate the significance of past events by means of a "re-enactment": far from looking for regularities that can ground explanations as other scientists do, their method is to empathize with men in the past (re-enact those happenings), in order to give an interpretation of the facts *as if* they themselves had been involved in them.

Hempel relived the debate just against claims defended by philosophers of history. He reintroduced a defence of a Humean point of view that is still acknowledged as a seminal starting point for tackling the issue, as opposed to a line of thought that includes anti-naturalists reminiscent of Collingwood, such as Gardiner (1952), Scriven (1959), Dray (1957), Ernst Nagel (1961), Waismann (1965); and Wittgensteinians, whose trend I will spell out more in detail in a moment and whose main representatives are Ryle (1949), Anscombe (1957), Winch (1958), von Wright (1963; 1971), and that are echoed by Austin (1956), Strawson (1962), Hampshire (1959), Melden (1961), Davidson (1963), Taylor (1964), Apel (1982). Between the '50s and the '70s antinaturalist

<sup>10</sup> In that case, one must be independently committed to believing in the validity of the induction.

<sup>11</sup> For the commonality of interests, von Wright (1971, ch. 1) said that any supposedly division between analytic and continental philosophy was likely to become more and more senseless.

concerns earlier expressed in philosophy of history found a new idiom and new *raison d'être* in philosophy of action and analysis of ordinary language. From these latter fronts new arguments came for resistance to assimilating the two enterprises. These arguments are of paramount importance to let us approach Millar's view more closely.

The main line of challenge employed by anti-naturalists of a Wittgensteinian provenance who discussed Hempel's theories was associated with their commitment to what makes a rational connection valid in the first place, apart from establishing whether it must be rival to a causal one. They in fact maintained that no IP generalization could ever be explanatory of why people tend to raise their arm in order to call for attention. This is so because, however frequently that behaviour can be performed by humans, what makes it intelligible to an observer is *just* the logical link between having such an intention and acting in that way. That is, we understand this behaviour in light of that intention, and even in the absence of any generalization linking the two events. By the same token, knowing about the tendency of people to act in that way when they have the relevant intention doesn't contribute at all to help us understand the link between them. The connection is *conceptual*, not empirical - as instead a Humean interested in searching empirical regularities might wish - and it's not answerable to any regularity from which one can infer the reason why one acts as one does. The underlying thought here is twofold: a) it is an analytic truth that the action of  $\Phi$ -ing is intentional, that is, there's a corresponding intention of  $\Phi$ -ing that makes the action what it is - otherwise it would not be action at all; b) anyone who has the intention of  $\Phi$ -ing must be prepared to do whatever is necessary to  $\Phi$ -ing. This latter can be regarded as a principle that defines what an intention is. The principle is *constitutive* of the intention. As any principle, its validity depends on the norms it specifies, not on the frequency of their application. We *a priori* understand what an intention is by grasping its intrinsic nexus with following a norm like means-end principle, not by means of generalizations. This is the so called "Logical Connection Argument" which establishes the source of our understanding: it lies in our grasping a norm, not in generalizations or a laws grounded in any empirically recorded regularities. An alien landed on the Earth, who hadn't the capacity to grasp the normative dimension of our mental life, could not understand at all why we stop at traffic lights when they are red. He could realize that we actually do that, and presumably develop expectations, but he could not realize *why* we act in this way.

Wittgenstein (1953, § 325) had also argued against mistaking a mental mechanism that supports the transition between some states to others, or to behaviour, for the normative transition between there being norms, or "rules", that one is sensitive to and conforming to them. While the former is subject to factual judgments about how the transition *de facto* occur, the latter is answerable in terms of correctness / incorrectness, and the judgments that apply to it are evaluative, about how the conformity / unconformity is *de jure* attained.

But why could not the normative transition also be causal, when we understand it via norms? After all, one may say, the alien in the above-mentioned example could not only fail to understand, but also to explain why people stop at traffic lights. In other words, sensitivity to norms could have a causal import on behaviour. The reason why causality was barred was thought to be provided by an argument, the so-called "Argument from Verification", which is strictly related to the above considerations. In connection with the "Logical Connection Argument", this seems to threaten the very possibility of our being spelling out causes for actions when we rationalize them by citing the reasons the agent had. The thought runs as follows: Hume's lesson on causality was that it is requisite that one could identify the cause and its effect independently. That is to say, since the causal connection is an empirical matter, it has to be discovered and verified *a posteriori*. But then this feature of the causal connection rules out the possibility that the cause can ever logically entail its effect. The entailment is precisely that on which a reason for action or belief works. If I say that I call a taxi because I need to catch a train in few minutes, and I believe that I could do this by calling a taxi, my intention to catch a train by using a taxi already *logically* entails my action of calling a taxi. There's a logical connection between the intention that constitutes my reason for the action I

perform and the action itself. Such logical connection that holds between intentional states, like intentions, beliefs, and action itself is thought to be incompatible with the demand of the argument from verification, and thus to fail to be taken as causal. The very “logic”, or “grammar” of psychological terms, like intention, prevents us from considering them as causes.

It is not clear, however, that there is no viable route towards a resolution of the allegedly sharp dichotomies. Millar, as we will see, is in fact interested in a solution where such incompatibility should be overcome. For if these qualifications were thought to bring some light (and indeed did so) onto the specificities of scientific theories *vs.* theories fostered by the so called “humanities” – or, more loosely, onto the “linguistic games” played in either fields – a pressing question would legitimately loom: since human beings are part of physical nature, like any other physical entities, can the causal and the rationalizing points of view be completely detached? What kind of conceptual vocabulary is appropriate to account for their verbal and non-verbal behaviour? Perhaps both languages should happen to be necessary to do so, then what are we to conclude about the correct ontological nature of beings for which such a dual epistemology might prove to be indispensable? For the distinction between the two linguistic games – the causal and the rationalizing – doesn’t tell us anything on the metaphysics of the mind, while this interest is legitimate in itself, apart from any Wittgensteinian therapeutic avoidance of discussing such issues. Here is where eventually philosophy of science interfaces ontology of mental states and metaphysics. While Wittgenstein and the Oxford-Cambridge philosophy focused more on the differences between the intentional vocabulary *vs.* physical vocabulary, and assigned to philosophers the task of analysing ordinary language in order to rationally clarify mental concepts and their “logic”, many contemporary philosophers more ambitiously tried to say more on metaphysical issues as well.

The tension between norms and facts surfaces here again. The idea that mental *concepts* are normative has easily led to an interest in whether the corresponding mental *states* are normatively regulated *as a matter of fact*. It seems obvious that our representational capacities depend, among other things, on our having natural capacities that make them possible. When these capacities are investigated by psychological sciences as *de facto* mechanisms, they are described as merely conforming to causal patterns. Under a non-intentional language, however, these capacities seem to fall short of the normative dimension that ordinary language constitutively attaches to the relevant mental states. Perception provides good examples of the problem: we perceive our surroundings by virtue of our having sense organs and a suitable brain architecture. This structure operates by being caused by stimuli that impinge upon us. This however doesn’t explain how we sometimes attribute incorrect perceptual representations to others or to ourselves. There’s nothing in the perceptual mechanism itself that establishes such answerability to experience.

Similar problems are provided by language use: although it is widely acknowledged that we have complex disposition to language learning and verbal production, what establishes the correct use of a concept is not grounded in such capacities. Kripke (1982) made much of the work to show how language use cannot be a straightforward factual matter. As to the problems arising from analysis of perceptual knowledge, Quine (1960) famously tried to dissolve them by launching a program of reduction of psychological, intentional accounts, with the normative import they seem to contain, to natural patterns of stimulation; and Sellars (1956; 1963) forged the metaphor of two competing images of man that, respectively, science and intentional psychology offer: a “scientific image” according to which man is placed in the realm of law and a “manifest image” that seems instead to place him in the “logical space of reason”.

Davidson (1963; 1980) tackled the same problems. He claimed that our mental life is anomalous if compared to the rest of nature. First, it isn’t governed by law-like connections, though it seems plain that mental states are realized by brain states that are instead causally and nomologically related; secondly, it displays a normative character that has to do with our being rational and having propositional attitudes, like beliefs, intentions, etc. that are mainly governed by our reasons for

holding them. Though irrationality is a real possibility for human beings, there must be limits to the amount of irrationality we can tolerate in mutual understanding, since by using common sense psychological concepts we assume other people to be largely rational agents. (Were it to be the case that irrationality is overwhelming, then understanding would fail and the assumption of rationality would have to be given up). This however seems an *a priori* argument on how epistemology of mutual understanding works, and it is far from clear that it solves anything like the metaphysical tension between facts and norms we are facing - if anything because it restates the contrast contemporary philosophy has left us with. McDowell (1994) found it wanting in various respects and tried arguing that rationality and normativity must eventually find a free-tension place in the natural world: human beings are *naturally* placed within a space of reasons, and it is time to stop thinking philosophical questions in light of a merely assumed contrast.

Thus, the dichotomies between causes and reasons, between explaining and understanding, and between sciences of nature and sciences of man never cease to reappear, along with their intrinsic connection with sensitivity to norms and human rationality. Earlier discussed within philosophy of history, they then found a fertile ground in philosophy of science, philosophy of action, language and perception; in metaphysics of mind in general, and of mental causation in particular. Let us then see what, in Millar's overall project, is related to this tradition.

### III

Alan Millar's *Understanding People* is observant of the complexities and ramifications of the problems explored, and attentive to recent developments within this complex cluster of neighbour fields. Among the main sources of inspiration on normativity are Kripke, Davidson and McDowell. Like Kripke, Millar believes that any dispositional, non normative account of our use of concept and having propositional attitudes, such as beliefs and intentions, cannot help us with attribution of those states to others, which is what understanding consists of. Like Davidson, he views the practice of reciprocal understanding as governed by the assumption of human rationality: if this needn't set unrealistic standards for human thinking and behaviour, nonetheless normativity is implied by having propositional attitudes, such as beliefs and intentions. Like McDowell, he believes that the space of reason is where believing and intending, by their very nature, place us. The resulting picture is that human reason and its normative regulation is natural enough not to prevent us from describing and explaining – and not merely prescribing and understanding - how people on many occasions actually act, and by virtue of what. This also promises to put Millar in partial disagreement with those who state the peripheral nature of generalizations for interpersonal understanding.<sup>12</sup>

Millar believes that normative considerations that people are sensitive to are central to that very description. For this reason he maintains that rationalizing explanations are different from those used by science, where sensitivity to norms doesn't figure. At the same time, he wishes to save the idea that they are full-blown *explanations*. Thus, he is interested in a solution where the main hypothesis that long lay at the core of the tradition seen above, namely that reasons one has must not also be causes for one's thinking and action, is not taken as compelling. If sensitivity to norms has causal and thus explanatory import, the battle is over.

As to the distinction between explanations that are rationalizing and explanations that are not, he claims that the difference lies just in the much-thrown-around "normative dimension of our mental life", which largely coincides with our being *rational agents*. As to their common explanatory nature, he does an admirable job in clarifying how we can begin to rescue normativity and rationality into a perspective which is at once perceptive of naturalistic demands and unprejudiced towards the existence of human rational thinking.

---

<sup>12</sup> Millar, however, claims that they are compatible, not whether the generalizations contribute in themselves to understanding (von Wright's problem, seen above), though perhaps more would be desirable to know about whether he also believes that generalizations can specifically contribute to understanding, see Millar 2004, ch. 1 and 7.

As far as I read his book, Millar achieves this result in terms of a project that overall tends to present normativity as sufficiently “psychologized”. That is to say, normativity doesn’t drop from the skies, nor is it so alien to constitute an intractable problem for naturalism. Normatively regulated transitions can in fact be regarded as normal *psychological processes*. As Millar put it, “(...) any theory that represents believing and intending as lacking an intrinsic normative dimension fails to provide an adequate account of the psychological role of beliefs and intentions” (2004, p. 135). That being so, Millar’s view seems to suggest that there’s more factuality in our being sensitive to norms than one might have thought. Normative processes, however, are so constituted that:

- (i) understanding these processes is inextricable from grasping the *reasons* that guide people as to what they ought to think and do - that which they often in fact think and do;
- (ii) sensitivity to reasons is *constitutive* of attitudes like believing and intending, on which Millar mostly focuses, so that normativity is *intrinsic* to having those attitudes;
- (iii) any theoretical attempt to show that these processes are not intrinsically normative could never help us to grasp and ascribe such processes.

Pivotal to fully appreciating this set of claims is the notion of “normative commitment”. As to (i), Millar argues that when we ascribe mental states like beliefs and intentions to others, we essentially ascribe to them some basic *commitments* to what they ought to think and do. Human beings are suitably equipped to appreciate some basic, *constitutive* implications of what they believe and intend, and acquire beliefs and intentions in light of considerations about what ought to be thought or done, given what else they believe, desire, etc. For the attitude of belief, the constitutive principle is the *Implication Commitment Principle*, according to which:

For any  $x$ ,  $\pi$ , if  $x$  believes  $\pi$ , then for any  $\theta$ , if  $\pi$  implies  $\theta$ , then  $x$  incurs a commitment to believing  $\theta$ , if  $x$  gives any verdict on  $\theta$ .

Intending is instead shaped by the *Means-End Commitment Principle*, which states:

For any  $x$ ,  $\phi$ , if  $x$  intends to  $\phi$  then  $x$  incurs a commitment to doing what is necessary if  $x$  is to  $\phi$ .

Millar’s core claim is that such commitments aren’t “extrinsically” attached, as it were, to those states; rather, they shape the very phenomenology of believing and intending, by virtue of being *reflexively* incurred by those who believe and intend something. Were they to be treated as extrinsic to the processes, then the very psychology of believing and intending would be mistakenly depicted. If Millar is right about (i) and (ii), then (iii) has a point. And it is on the basis of (i) and (ii) that Millar can in fact object to competing theories about normativity which deny on various assumptions that normativity is intrinsic to our mental life, e.g. expressivism, deflationist strategies, and dispositionalism (see Millar, Précis, *this issue*, pp. 17-18 and Millar, 2004, ch. 1).

It is worth saying that Millar’s analysis of the relevance of normativity in making sense of human behaviour while also explaining it is judicious as to the scope of these explanations. His thesis doesn’t aim at ruling out cases where people actually happen to think or do things that don’t make sense by their own lights. His analysis of why human rationality needn’t compel one to give up normativity constitutes an excellent illustration of how defending normativity and the so called “high conception of propositional attitudes” doesn’t necessarily lead one to unrealistic conclusions on human mind. The same holds for reflexivity: although it plays an essential role in Millar’s notion of commitment, he takes that being reflectively aware of a commitment needn’t often be more than an unconsciously routinized matter. Being normatively sensitive is far from requiring an over-

intellectualized picture of the mind.

These core claims, other related and their defence have given rise to much debate among the commentators:

- Mario De Caro largely agrees with Millar, though raises the question whether Millar's view on mutual understanding is sufficiently protected against a sceptical challenge first directed against Davidson's interpretationalist theory, according to which the practice of interpersonal interpretation is grounded in the interpreter's assumption of "an acceptable degree of coherence and cogency" (De Caro, *this issue*, p. 25) in the reasoning and actions of those interpreted. But how can we be sure that they really tend to conform to rationality standards? De Caro believes that the challenge can essentially be recast to question Millar's account, since this latter is explicitly Davidsonian as well, in its appealing to human rationality as the basis for interpersonal understanding. Another point De Caro raises is whether the theory of normativity Millar puts forward in the book could be applied to solve the problem of "explanatory exclusion", according to which intentional, psychological vocabulary and the rationalizing explanations it licenses could prove to be superfluous in front of a complete account of nature formulated in physical terms.
- Simone Gozzano is interested in questioning whether Millar's notion of commitment is adequate to specify the available options for discharging such commitment, since there is more than one reasoning strategy to do so. These strategies should be specified along with the commitments incurred, in order to predict the upshot of our epistemic conduct. Furthermore, he asks whether Millar's notion of belief can "receive a naturalistic reading" (p. 35), so as not to contrast too sharply with psychology of non-human animals that lack our reflexive capacities. Although Millar's view is that there is a difference in kind between human and other animals' psychology, Gozzano is inclined to think that the existence and operation of sub-doxastic states in human psychology as well is an interesting bridge between the two groups of creatures. Another problem attached to this view, he argues, is that sub-doxastic states seem to be perfectly accounted for in non-normative terms, though this doesn't exclude their rationality and their playing their causal roles. Another merit of aligning animal and human psychology is to avoid overloading our cognitive life by requiring conscious sensitivity to constitutive norms for any instance of believing and intending.
- Adrian Haddock closely reflects upon the Implication Commitment Principle, which is crucial to Millar's account of belief. He considers three theses that Millar employs to give weight to the Principle, namely (i) a strong conceptual thesis, according to which the Principle is analytically true; (ii) a weak conceptual thesis, which says that grasping the concept of belief commits one to accepting the Principle; and (iii) a constitutive thesis, for which the Principle is a partial specification of the very nature of belief. He analyses the combination between them and provides an argument as to whether there are no other alternative explanations of why the Principle is valid, which rely on neither of them. He suggests that the very nature of rationality provides such an alternative. He also tackles issues of irrational beliefs and, more generally, of beliefs that are recalcitrant to rational reflection. They might threaten Millar's explanation as to why beliefs are incompatible with a contemplative stance.
- Paul Noordhof argues against Millar's idea that not all normative reasons for action are justificatory reasons. He believes, on the contrary, that if there's no justification for action, then "the attitude, and actions that follow from it, fall at least partly outside the ambit of

normative explanation” (p. 50). He also questions whether it is safe to claim that reasons – understood, as Millar does, as *considerations* in favour of attitudes and actions – always support intentional behaviour. He provides what he considers as genuine examples of behaviour that one may want to classify as intentional, though fails to be consideration-based. He has doubts concerning Millar’s idea that normativity arises from practices and his account of the demands the practices impose upon those who participate in them. Noordhof disagrees with Millar’s reading of dispositionalism, as allowing a contemplative stance that normative commitments incurred by believing and intending do not license. He argues that such a stance is possible in many cases, and that when it is not, this seems more due to endorsements of attitudes than to normative commitments. About the high conception of propositional attitudes that Millar supports, Noordhof expresses concern as to whether the reflexivity it requires may be excessively demanding. Finally, he engages in how Millar’s commitments can have an explanatory impact at all. Understanding this, he argues, would be of paramount importance to the project, that Millar seems to welcome, of integrating mentalistic explanations with physical explanations.

## References

- Anscombe, G. E. (1957), *Intention*. Oxford: Blackwell.
- Apel (1979), “The *Erklären-Verstehen* controversy in the philosophy of natural and human sciences”, in Floistad (1982), vol. II: 19-49.
- Austin, J. L. (1956), “A Plea for Excuses”, reprinted in White (1968).
- Burge, T. (1992), “[Philosophy of language and mind 1950-1990](#)”, in *The Philosophical Review* 101: 3-51.
- Collingwood, R. G. (1940), *An Essay on Metaphysics*. Oxford: Clarendon Press.
- Collingwood, R. G. (1946), *The Idea of History*. Oxford: Oxford University Press.
- Comte, G. A. (1830-1842), *Cours de philosophie positive*. 6 voll. Paris: Bachelier.
- Davidson (1963), “[Actions, reasons and causes](#)”, in *The Journal of Philosophy*, 60, 23: 685-700, reprinted in Davidson (1980).
- Davidson (1980), *Essays on Actions and Events*. Oxford: Clarendon Press.
- Dilthey, W. (1883) *Einleitung in die Geisteswissenschaften*, reprinted (1989) as *Introduction to the Human Sciences* (eds. R. Makkreel & R. Frithjof). Princeton, NJ: Princeton.
- Dray, W. (1957), *Laws and Explanation in History*. London: Oxford University Press.
- Floistad, G. (1982), ed., *Contemporary Philosophy. A New Survey*. 4 voll. The Hague: Nijhoff.
- Fulford, K. W. M. (Bill), Thornton, T., and Graham, G. (2006), *Oxford Textbook of Philosophy and Psychiatry*. Oxford: Oxford University Press.
- Gardiner, P.(1952), *The Nature of Historical Explanation*. London: Oxford University Press.
- Gardiner, P.(1959), ed., *Theories of History*. New York-London: The Free Press-MacMillan.
- Garrett, B. (2006), *What is This Thing Called Metaphysics?* London: Routledge.

- Gozzano, S. (2007), *Pensieri materiali*. Torino: UTET Libreria.
- Hampshire, S. (1959), *Thought and Action*. London: Chatto and Windus.
- Hempel, C. G. (1942), "The Function of General Law in History", in Gardiner (1959), pp. 344-56.
- Hempel, C. G. (1965), *Aspects of Scientific Explanation and Other Essays in Philosophy of Science*. New York-London: The Free Press-MacMillan.
- Hempel, C. G. (1999), "Laws and Their Role in Scientific Explanation", in R. Boyd, P. Gasker, and J.D. Trout, eds., *The Philosophy of Science*. Cambridge, MA.: MIT Press, pp. 299–335.
- Hempel, C. G. and Oppenheim, P. (1948), "[Studies in the Logic of Explanation](#)", in *Philosophy of Science*, 15, 2: 135-175.
- Hempel, C. G. (1966), *Philosophy of Natural Science*. London: Prentice-Hall.
- Hume, D. ([1748] 1975), *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Oxford University Press.
- Kripke, S. (1982), *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.
- Ladyman, J. (2002), *Understanding Philosophy of Science*. London: Routledge.
- Mackie, J. L. (1993), "Causes and Conditions", in E. Sosa and M. Tooley, eds., *Causation*. Oxford: Oxford University Press, pp. 33–50.
- McDowell, J. (1994), *Mind and World*. Cambridge, MA: Harvard University Press.
- Melden, A. I. (1961), *Free Action*. London: Routledge.
- Mill, J. S (1863), *Utilitarianism*, (G. Sher, Trans.), IN: Hackett Publishing, 2002.
- Millar, A. (2004), *Understanding People: Normativity and Rationalizing Explanation*. Oxford: Clarendon Press.
- Nagel, E. (1961), *The Structure of Science*. New York: Harcourt Brace.
- Peruzzi, A. (2004), (ed.), *Mind and Causality*. Amsterdam/Philadelphia: John Benjamins, 2004.
- Popper, K. R. (1934), *The Logic of Scientific Discovery*. London: Hutchinson.
- Quine, W. V. O. (1960), *Word and Object*. Cambridge, MA: MIT Press.
- Ryle, G. (1949), *The Concept of Mind*. London: Hutchinson.
- Salmon, W. C. (1990), *Four Decades of Scientific Explanation*. Minneapolis: Univ. of Minnesota Press.
- Scriven, M. (1959), "Truisms and the Grounds for Historical Explanations", in Gardiner (1959), pp. 443-475.
- Sellars, W. (1956), *Empiricism and the Philosophy of Mind*. Cambridge MA: Harvard University Press, 1997.
- Sellars, W. (1963), "Philosophy and the Scientific Image of Man", in W. Sellars, *Science, Perception, and Reality*. London: Routledge, 1963, pp.1-40.
- Strawson, P. (1962), "Freedom and Resentment", reprinted in *Freedom and Resentment and Other*

*Essays*. London: Methuen, 1974, pp. 1-25.

Taylor, C. (1964), *The Explanation of Behaviour*. London: Routledge.

Waismann, F. (1965), *The Principles of Linguistic Philosophy*. St. Martin's Press.

White, A. R.(1968), ed., *The Philosophy of Action*. London: Oxford University Press.

Winch, P. (1958), *The Idea of a Social Science and its Relation to Philosophy*. London: Routledge and Kegan Paul.

Wittgenstein, L. (1953), *Philosophical Investigations*. G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.). Oxford: Blackwell.

Wittgenstein, L. (1979), *Wittgenstein's Lectures, Cambridge 1932-1935*. A. Ambrose (ed.), Oxford: Blackwell.

Wittgenstein, L. (1980), *Wittgenstein's Lectures, Cambridge 1930-1932*. D. Lee (ed.), Oxford: Blackwell.

Wollheim, R. and Hopkins, J. (1982), eds., *Philosophical Essays on Freud*. Cambridge: Cambridge University Press.

Wright, G. H. von (1963), *Norms and Action. A Logical Inquiry*. London: Routledge.

Wright, G. H. von (1971), *Explanation and Understanding*. Ithaca, NY.

# **A *Précis* of Understanding People: Normativity and Rationalizing Explanation**

ALAN MILLAR

(University of Stirling)

## **1. The project**

A number of philosophers have suggested that it is essential to our mental life that it has a normative dimension. This *core idea* appears in one way in Saul Kripke's critique of dispositionalist accounts of meaning. For Kripke, meaning something by a word is not simply a matter of how one is disposed to use the word. That one means *plus* by 'plus' seems to have implications concerning how one *ought to* use the word. But it is hard to see how merely being disposed to use a word in certain ways can, by itself, have such implications. The core idea also comes over in John McDowell's thinking. McDowell (1994) has made much of the idea that to think, desire, intend, and so forth is to operate within the space of reasons. As understood by McDowell this implies that the possession of propositional attitudes is inextricable from governance by reason, which in turn implies that having propositional attitudes is not, as some would put it, a straightforwardly factual matter. On the contrary, it would be essential to our mental life that we are governed by reason in such a way that we *ought to* think and act in certain ways. In *Understanding People* I aim to clarify and defend the core idea and to consider its implications for the character of our understanding of people. In particular, I argue that understanding people is significantly different from scientific understanding, just because it views people as rational agents governed by reason.

## **2. Commitments in three key areas of the discussion**

The concept of a normative commitment is central to the perspective developed in the book. This concept figures in three areas of discussion.

### *(1) Believing and intending*

The first concerns the commitments of believing and intending (pp. 72-83). I take it to be intuitively plausible, and in keeping with common sense thinking, that believing something commits you to believing other things and that intending something commits you to doing certain things. The relevant concept of commitment is normative. It is not about what people are resolved to do or have committed themselves to do. Rather it bears upon what they ought to do. The issue is to explain how.

So far as belief is concerned a natural first thought is that if, for instance, I believe that Tom is going to the party, and I also believe that if Tom is going then Sally will go, then I ought to believe that Sally will go. Such a view faces two problems. The first is posed by the plausible assumption that while crazy beliefs incur commitments, it does not follow that those who have such beliefs ought to carry out the commitments they incur. It may be that they ought instead to abandon the

beliefs that incur the commitments. The second problem is that it is surely not the case that we could believe everything that our beliefs commit us to believing. It is impossible to believe all of the infinite propositions entailed by what we believe. I address these problems with the help of the notion of a verdict on a proposition. Think of the verdicts on a proposition as being belief, disbelief (rejecting the proposition) and suspension of both belief and disbelief. My suggestion is that the claim that believing that  $p, q, \dots$  commits you to believing that  $r$  amounts to the claim that you ought to avoid believing  $p, q, \dots$  and giving a verdict on the proposition that  $r$  other than belief. You can discharge this commitment in one of two ways—*either* by carrying it out, that is, by ensuring that *if* you give any verdict on the proposition that  $r$  then that verdict will be belief, *or* by giving up one of the beliefs that incur the commitment. I apply the conception of a commitment to intentions as well. Here the key idea is that intending something commits you to doing what is necessary to do what you intend. Obviously, in carrying out this commitment you will be guided by your beliefs about the necessary means, but the basic commitment incurred by intentions, I claim, is to do what is, as a matter of fact, necessary to do what is intended. There are derivative commitments incurred by intentions in combination with the beliefs you have about what is necessary. These may clash with the basic commitments. Nonetheless, the derivative commitments are explained by, and are in some sense for the sake of, the basic commitment.

## (2) Practices

Practices are activities that are essentially rule-governed. When one plays a game of football one is not merely subject to the rules of the game. Rather it is essential to the activity in which one is engaged that it is governed by those rules. Walking over a grassy area that happens to be a park, one might be subject to the rules of the park authority. But is it not essential to the activity of walking over that grassy area that one is subject to those rules. The area might never have come under the control of the park authority. By contrast, there is no possible world in which you are both a player of a game defined by certain rules and not subject to those rules. When you are a participant in a practice, how ought you to behave? It would be quite natural to respond by saying that you ought to follow the rules, but I claim it is better to represent participants as being *committed* to following the rules (pp. 83-92). The notion of a commitment that is applied here is that already introduced in connection with believing and intending. The idea is that participants ought to avoid continuing to be participants while not following the rules. The commitment can be discharged either by carrying it out—by following the rules—or by withdrawing from the practice. This helps to accommodate a point once made by Philippa Foot (1972) in connection with etiquette. Foot remarked that the mere fact that there are rules of etiquette does not give anyone a reason to follow them. I think this is right. We can accommodate the point if we think of rules of etiquette as governing practices and think of participants in such practices as incurring a commitment to following the rules. Perhaps for some such practices the rules are silly or even objectionable. It may be that participants ought not to follow them and ought to withdraw from the practice. Withdrawal would be one way of discharging the commitment incurred by being a participant.

Practices need not be governed by explicitly formulated rules. In this connection, I discuss a practice of giving and receiving undertakings (pp. 87-89). People sharing an apartment might engage in such a practice. Though they never formulate rules they may show, by the ways in which they interact, that a practice is in play among them. Suppose someone says that he will shop for food. If this is said in an appropriate context, for instance, in a discussion about the distribution of tasks, then it will be treated as an undertaking. In that case the others will expect that the job will be done, unless for some good reason it cannot be, and if it is not done, and no reasonable excuse if forthcoming, then they would regard this as making some kind of censure appropriate.

I apply the notion of a practice in connection with offices in institutions. Chief executives of companies and officers in armies occupy roles that are defined by certain institutionally devised duties. In effect, there is a practice of fulfilling the role of chief executive or the role of army

officer. The rules of these practices require that the duties defining the office should be carried out. Just in virtue of occupying the office one incurs a commitment to carrying out the duties that define it. The commitment can be discharged either by carrying out the duties or by withdrawing from the office. I offer some reflections on the role of our grasp of practices in understanding people on pp. 241-47.

(3) *Meaning and concepts*

The concept of a practice, and the associated conception of the commitments incurred by participating in a practice, move to centre stage in a discussion of meaning (Ch. 6). Discussions of normativity in connection with meaning have paid insufficient attention to a distinction between *two concepts of correctness*. So far as terms are concerned the distinction is between correctness in the sense of ‘true application’ and correctness in the sense of ‘being in keeping with, or faithful to, the term’s meaning’. An application of the term ‘cow’ to something would be correct in the first sense if and only if the thing is a cow. Being correct in the second sense contrasts with being a misuse. Suppose that I incorrectly (= falsely) apply the term ‘cow’ to an animal I see in the distance. It does not follow that I have misused the term in any sense that implies that I have not used it in keeping with its meaning. I may have mistaken the animal I see for a cow but I used an appropriate word to say what I wanted to say. Or I might have wanted to mislead someone into believing that there is a cow over there. In this case too I would have misapplied the term, but I have not misused it. I meant to say that the animal is a cow and I used an appropriate term for that purpose. Compare this with a situation in which I am in the early stages of learning English and, looking at a sheep that is in clear view, and I apply to it the term ‘cow’. The most likely explanation is that I have used the term as if it meant *sheep* and on that account have not used it in keeping with its actual meaning. In that case I have used the term incorrectly in the second sense.

I suggest that we should think of the meaning of words as being linked to practices in the sense indicated already. Being able to use a word in keeping with its meaning is a matter of participating in a practice of using that word. Just in virtue of being a participant in such a practice one incurs a commitment to following the rules of the practice. So far as terms are concerned I consider a rule the general form of which is: use the term in keeping with the conditions for its correct, in the sense of ‘true’, application. Following that rule does not require that *whenever* you apply the term you should do so correctly in the first sense on which correctness is truth. As I have already observed, an intentionally false application may be in keeping with the term’s meaning. It may accord with the rule just specified because it respects the conditions for the true application of the term. (If you want to tell someone falsely that something over there is cow you cannot do better than use a term that means *cow* (pp. 83-92).

Considerations similar to those that apply to the meaning of words apply to the employment of concepts (pp. 178-191). I show that the distinction between two kinds of correctness applies at the level of concepts. This requires me to make sense of the idea that someone may possess a concept and yet fail to employ it in keeping with its content. However, an important feature of my view is that the commitments incurred by possessing a concept are not to be explained in terms of practices (pp. 183-82). There are practices for the use of words and, correspondingly, rules for the use of words. A term T in English, say, has a meaning bound up with a practice that explains why it expresses the concept it does express. But the concept is not inextricably tied to that practice since it may have a role in the thinking of a person who knows nothing of the meaning of T and who does not participate in any practice of using T. It is the ‘logic’ of the concept that determines the commitments that are incurred by employing it.

**3. Explaining the commitments incurred by beliefs and intentions**

In twentieth century metaethics over the last century or so there has been much discussion of status

of moral claims. The issue was framed in terms of a distinction between, on the one hand, factual claims, conceived so as not to include normative claims, and, on the other hand, evaluative moral claims, understood as having normative implications. The idea that evaluative claims could be true or false was thought to be both metaphysically and epistemologically problematic. They were thought to be metaphysically problematic because it seemed puzzling that there should be facts that make such claims either true or false. They were thought to be epistemologically problematic because it seemed that they could not be established either empirically or *a priori* and it seemed implausible that we have some special cognitive capacity for discerning evaluative truths.<sup>1</sup> Parallel issues arise in the philosophy of mind. On the one hand, we have psychological statements which would be widely held to be factual in some sense that is taken to exclude their having normative implications. Ascriptions of belief and intention, for instance, would be widely held to be factual in this sense. On the other hand, we have normative claims, for instance, that one ought to take such-and-such means if one intends to bring about a certain end, or that one ought not to believe *P* and *Q* if *P* and *Q* are inconsistent. It is striking that so little attention has been paid to how exactly these kinds of normative claims relate to those psychological statements that are supposed to be non-normative.<sup>2</sup> This is despite the fact that at least some of the problems thought to attach to moral statements seem to be due, not to the fact that they are moral, but to the fact that they are normative.<sup>3</sup> My main concern in the book is not so much about metaphysical or epistemological problems that may attach to normative claims, but about how we should think of the relationship between certain psychological statements and statements bearing on how people ought to think or act.

In Ch. IV, which I regard as being at the heart of the book, I focus on the issue of how statements ascribing a belief or intention relate to the corresponding normative statements ascribing commitments incurred by the belief or intention. For instance, I ask how

(1) Sally intends to buy a present for her brother

relates to

(2) Sally incurs a commitment to doing what is necessary if she is to buy a present for her brother.

(2) is clearly normative. Plausibly if (1) is true then (2) is. I express this by saying that (1) has *normative import* which is captured by (2). The claim that (1) has this normative import is weaker than the claim that (1) is itself a normative statement. (1) would be normative, in the strict sense, only if it entailed some normative statement. I envisage that theorists who deny that (1) is normative will attempt to explain its normative import by invoking a principle connecting having the intention in question with incurring the commitment specified by (2). For the sake of having a definite target I discuss dispositional conceptions of propositional attitudes, though key points would apply to any view on which ascriptions of propositional attitudes are conceived as not being normative.

*Dispositionalism* about propositional attitudes is the view that propositional attitudes (at least conceived as standing states) are dispositional and that our concepts of the attitudes are dispositional concepts. It is to be understood that the relevant dispositions are conceived as analogous to the paradigm cases of dispositions, like flexibility, fragility and the like so that being

1 Such thinking informs emotivism and prescriptivism as well as the error theory. Mackie (1977) succinctly sets out the problems.

2 Allan Gibbard (1990) is relatively unusual both in homing in on the parallels between normative claims and factual ones and in extending an expressivist theory for moral statements to normative statements pertaining to rationality. Blackburn (1998) addresses the issue.

3 Mackie's argument from relativity seems to turn on considerations that are thought to make it plausible that moral beliefs are explicable in psychological and sociological terms, without the need to posit moral facts that suitably equipped people can discern. A similar line of thought may be found in Harman (1977). But his argument from queerness seems not to turn on assumptions about the distinctively moral content of moral statements.

in a dispositional state has no intrinsic normative import. Beliefs on this view would implicate dispositions to form further beliefs, depending on what else the subject believes, and dispositions to act in certain ways, depending on what the subject desires or intends.

Dispositionalists are among those who may wish to explain the normative import of statements like (1) above by invoking connecting principles, that is, principles connecting the non-normative to the normative. The strategy can be illustrated with the help of an example from ethics. Suppose that my neighbour is at my door, in distress, and calling on me for help. The consideration that this is so—call it *C*—seems to be non-normative. Prompted by it I may think that I ought to help because I take *C* to give me a reason to help that is not countervailed by other considerations. Thus, in effect, I treat *C* as having normative import. It would be *some* explanation of why I regard *C* as having normative import to point out that I accept, at least implicitly, a principle to the effect that the fact that one's neighbour is in distress and calling on one for help is a reason to help. This principle is a reason-specifying principle. It connects a certain type of non-normative consideration with a certain type of normative consideration. If the principle holds it goes some way to explaining the normative import of the consideration in question. In the light of my acceptance of the principle it is explicable that I should count the consideration as providing me with a reason to help my neighbour.

I envisage the dispositionalist applying a similar line of thought to explain why (1) has normative import which is captured by (2). The idea would be that we just need to invoke an appropriate connecting principle. The following principle, which I call the *Means-End Commitment Principle*, would fit the bill:

(3) For any  $x$ ,  $\phi$ , if  $x$  intends to  $\phi$  then  $x$  incurs a commitment to doing what is necessary if  $x$  is to  $\phi$ .

The explanation for the normative import of (1) which is captured by (2) would be that (2) follows from (1) and (3). Analogous considerations would apply to statements ascribing beliefs. In connection with these the following *Implication Commitment Principle* might be invoked:

(4) For any  $x$ ,  $\pi$ , if  $x$  believes  $\pi$ , then for any  $\phi$ , if  $\pi$  implies  $\phi$ , then  $x$  incurs a commitment to believing  $\phi$ , if  $x$  gives any verdict on  $\phi$ .

The question now is whether the strategy described is adequate. I suggest that it is not. Invoking a principle to the effect that the fact that an action would be a killing is, in the absence of special circumstances, a reason against doing it, provides no illumination on what makes killing wrong. Moral philosophers have aimed to provide deeper accounts. Consequentialists look for an answer in terms of the consequences of individual acts of killing for the general good, or in terms of the consequences of there not being widespread acknowledgement of a strong prescription against killing. Theorists inspired by Kant look for an answer in terms of the view that persons are ends in themselves who, as such, are worthy of respect. These strategies explain the wrongness of wrong killing in terms of a failure to have due regard to something which is held to be good in itself or worthy of some special kind of concern. For the consequentialist it is the general good. For Kantians it is persons viewed as ends in themselves. In each case the normative import of the consideration that an action would be a killing is traced to a basic assumption about what is valuable or worthy of special concern. It is not traced to non-normative considerations about what killing is but to normative considerations. The same problem arises from the attempt to explain the normative import of the fact that my neighbour is calling on me for help. The connecting principle that is invoked is not self-explanatory. If it is acceptable, its acceptability will be grounded in considerations about what is of value or worthy of regard.

I argue that invoking the Means-End Commitment Principle to explain the normative import of (1) will not do the trick. The principle itself provides little if any insight into why (1) has normative import. And if beliefs and intentions are conceived in non-normative terms it is hard to see what

deeper explanation is available. If Sally's intending to buy a present for her brother simply consists in her having various dispositions characteristic of an intention with that content, how is it that she incurs a commitment to taking the steps necessary to do so? The answer might seem obvious: if she does not take these steps then she will not carry out her intention. That is indeed so, but it does not supply what is needed. That she will not carry out her intention unless she takes the necessary means is clearly something that is of practical significance to Sally. But it has that significance for her only because she has incurred a commitment to travelling to Paris. This practical significance presupposes that the commitment has been incurred but does not explain why the commitment has been incurred.

The general problem about understanding the relations between ascriptions of intention and statements ascribing corresponding normative commitments can be made more vivid by thinking about a disanalogy between our intentions and our character traits. Suppose I am prone to being irritable. It would be sensible for me to do something about this since it would adversely affect my relationships with others that I care about. If that is so it is because it matters that I should have those good relationships. It is not because of something that is *intrinsic* to the character trait in question. And because it is not intrinsic to the trait it is conceivable, and indeed possible, that I should adopt a purely contemplative attitude to the trait from the standpoint of which I do not try to do anything about it, but simply observe how it manifests itself in various situations, and perhaps wonder what it will lead me to do in future situations in which I find myself. I might want to see what it would be like to observe how my troublesome trait affects me and others. I might want to write a short story based on the experience of taking such a stance. It may be irrational or foolish to adopt such a stance. It would be reprehensible from a moral point of view. What interests me is that it is possible that I should adopt such a stance.

If intentions were like character traits in that they consisted in clusters of dispositions to thought and action, then we should be capable of adopting a contemplative stance towards our own intentions. Yet it is hard to make sense of this. When I ascribe a current intention to myself I do not simply take myself to be in a certain dispositionally characterised state. The very thought by which I ascribe the intention to myself endorses that very intention and seems to involve an acknowledgement of the normative import of having the intention—an acknowledge that having the intention commits one to taking the necessary means to carry it out. I do not think that the dispositionalist can account for this peculiar feature of self-ascriptions of intentions. For if having an intention is merely a matter of being in a dispositional state then having ascribed to myself an intention it would, thus far, be an open question whether I should regard this as requiring anything of me. Other considerations—perhaps a principle like the Means-End Commitment Principle, understood as a principle connecting the non-normative with the normative—would have to be invoked to explain why, if at all, I regard my having of the intention as having normative import. I do not think that this strategy will help in explaining the normative import of (1). It is problematic whether the supposed status of the additional principle would be plausible if having an intention were not an essentially normative matter. Why suppose that the principle holds if intentions are not intrinsically normative? The connecting principle about killing is supposed to be explained, not just by consideration of what killing is in itself, but in terms of something of moral significance about killing a human being. Should we not look for some analogue in connection with the Means-End Commitment Principle? As I see it there is nothing extrinsic to intention that can explain the normative import of (1). If there were we could make sense of someone's accepting (1) yet not view it as having the normative import captured by (2). In practice we do not act as if it were an open question whether having an intention requires something of us. In practice we treat having an intention as inextricable from there being certain things that we have to do, where the latter is treated as a normative matter.

It is important to stress that none of this is at odds with the obvious fact that with regard to any intention that I have, I can raise the question whether it is wise to persist with that intention.

Suppose, for example, that I plan to travel to exotic places and have taken various steps to this end, like booking flights and accommodation, receiving vaccinations, etc. I start to wonder whether I should be doing this. If I raise this question then, it might be said, I am surely in a state in which it is both clear to me that I have the intention and yet not clear to me whether I shall carry it out. Would this not be a case of adopting a contemplative stance to my own intention? I do not think so. All of us may wonder whether we shall have the energy, the motivation, or the means to carry out intentions that we know we have. Even if we start to have doubts about a planned course of action it may still be true of us that we do indeed intend to pursue it to the end. A wavering intention is still an intention. But a contemplative stance to my intention would be one in which I think to myself that I (still) have the intention yet treat it as an open question whether this requires anything of me. I claim that a peculiar feature of our relationship to our own intentions is that we are not capable of adopting such a stance. The issue is how we should account for the phenomenon.

The phenomenon can be explained if we think of the concept of intending as being in important respects like the concept of promising and, correspondingly, think of intending as an intrinsically normative matter. Consider this principle.

*Promise Commitment Principle:* For any  $x$ ,  $\phi$ , if  $x$  promises to  $\phi$  then  $x$  incurs a commitment to  $\phi$ -ing.

This principle does not merely connect the non-normative to the normative in the way the Means-End Commitment Principle is supposed to do *on the dispositionalist reading* to which I have been objecting. It is, rather, a *constitutive* principle because it partially specifies what it is to promise something. It is also a *conceptual* principle. Grasping the concept of a promise commits one to accepting the Promise Commitment Principle. I take the Means-End Commitment Principle to be analogous to the Promise Commitment Principle, at least in the respect that it is constitutive and conceptual. If this is right then we should think of the ascription of intention (1) above, not merely as having normative import, but as being itself an implicitly normative claim. It is not just that if (1) is true then (2) is true, but that (1) entails (2). In thinking of Sally as intending to buy a present for her brother I thereby represent her to be in a commitment-incurring condition and I am myself committed to thinking of her as having incurred a commitment to do what is necessary to achieve that end. This has implications for the psychology of intending. Part of what it is for me to have an intention to travel is that I am prone to act in ways that I take to be relevant to travelling, *regulated by the commitment to taking the steps necessary* if I am to travel. So, for instance, if I find that I have some false beliefs about what is necessary I am likely to alter some of my detailed plans. To be regulated by the commitment is to aspire to avoid persisting with the intention while never getting around to taking the steps necessary to carry it out. The psychological role of an intention is shaped by the commitment it incurs. It is part of my wider picture that intentions perform their characteristic role through being recognised, and thus endorsed, as intentions. I pursue this, and a related theme concerning belief, in Ch. 5 of the book. But the point that I would like to emphasise here is that by treating ascriptions of intentions as being implicitly normative and, correspondingly, treating intentions themselves as being intrinsically normative, we can account for the peculiar character of our thoughts about our own intentions and of our relationship to our own intentions. The reason that we cannot adopt a contemplative stance towards our own intentions is that in thinking of ourselves as having an intention we endorse both the intention and the normative means-end commitment that having the intention incurs. This accounts for why it is paradoxical to think, ‘I intend to travel, but I shall not travel’ even although the content of the thought is not contradictory. In thinking the first conjunct of the thought I both represent myself as having incurred a commitment to doing what is necessary to travel and endorse that commitment. In thinking the second thought I resolve in effect to flout the very commitment that I have endorsed. That is a kind of incoherence. I do not think that a satisfactory explanation of the paradox is available to those who deny that intentions are intrinsically normative.

## 5. The Problem of Explanatory Relevance

In Ch. 7 I consider a problem that arises in connection with rationalizing explanations—explanations of, for instance, the formation of a belief or the performance of an action in terms of the subject's reason for forming that belief or performing that action. If rationalizing explanations are distinctive, *qua* explanations, then one might suppose that the explanatory insight that they yield has something to do with the fact that the relevant propositional attitudes rationalize the belief or action to be explained in the way they do. But if, as I claim in Ch. 1, the explanatory insight provided by rationalizing explanations is causal, and tied to appropriate generalizations, then it is not clear what is added by the rationalization. If the explanation is causal the insight it yields is provided by the information that the attitudes cited in the explanation figured in the aetiology of the formation of the belief or the performance of the action. That information provides explanatory insight of a limited sort. It implies that the attitudes cited in the explanation are such that, given suitable prompting circumstances, someone with attitudes of that sort would be liable to form a belief or perform an action of the sort being explained. That insight does not seem to depend on anything that has to do with reasons or rationalisation.

One reaction to the problem would be to deny that rationalizing explanations are causal and that the insight they provide implicates generalizations. I am interested in a solution to the problem on which rationalizing explanations are causal, the insight that they yield implicates generalizations, and yet the rationalization in question is, even so, explanatorily relevant to what is being explained.

My starting point for further reflection on this matter is that facts about the environment that people are in enable us to explain their behaviour without our having acquired independent information about what they believe. Knowing what others, given their situation, can hardly fail to know is an important guide to what they believe and therefore to what they will do. We can make sense of the movements of players in a game of football in part because we see the state of play as they do. We see how the players are positioned and orientated, where the ball is, which players are moving towards the ball, which players are waiting to receive passes, and so forth. But what enables us to connect the state of play to the behaviour of any given player? Obviously our understanding of the aim of the game and the rules that govern play is crucial. But a further, and just as crucial, part of our understanding of the behaviour of players depends on our grasp of what it makes sense for them to believe and intend, and what it makes sense for them to do given what they believe and intend. That more general understanding is not tied to any particular practice. It depends on our implicit understanding what is involved in believing this or intending that. It rests on requirements governing belief and intention in general—requirements that apply to any agent with beliefs and intentions. I pursue this theme to the conclusion that people are governed by normative considerations that they recognise and that, accordingly, understanding people requires us, as interpreters, to view them in the light of the normative considerations by which they are governed. This makes understanding people different from understanding states of affairs and events in the light of covering generalizations. It is not that generalizations have no role in our understanding of people. They obviously do. But our epistemic route to the generalizations is via normative considerations.

Further reflections on explanation and understanding are pursued in Ch. 8, in which, among other things, I take issue with simulation theory, and in Ch. 9 where I argue that there are limits to the scope of rationalizing explanation.

## 6. Normative commitments and normative reasons

I take normative commitments to be grounded in normative reasons. That is why I explore the notions of reasons for belief and reasons for action in Ch. 2. I claim that normative reasons for belief depend on belief's constitutive aim: to believe only what is true. I explore what it means to say that this is belief's constitutive aim (pp. 42-57). In this connection I make use of the idea that

believing is governed by the goal of believing only what is true. The story of how belief is so governed brings the notion of sub-intentional, goal-directed activity into the picture. I also argue that it is constitutive of the kind of believing that reflective creatures like ourselves go in for that we are not indifferent to whether or not our beliefs are true. Some of our believing may be sustained by mechanisms that serve psychological needs and deflect our attention from the truth or lead us to hold onto what is false. But the very fact that those mechanisms are in play testifies to the constitutive goal of believing only what is true.

Normative reasons for beliefs must justify those beliefs in the sense of making it appropriate to hold them given belief's constitutive aim. I argue in Ch. 2 that it is far less clear what normative reasons for action must be. From reflection on examples it is far from clear that we should think of all normative reasons for action as justifying the actions for which they are reasons. All normative reasons for action must favour or recommend the action for which they are reasons. But some may do so simply by conferring an instrumental point on the action. For instance, I may have an inclination to go for a walk on a sunny afternoon and, being free, intend to satisfy this inclination. Having this intention in some sense gives me a reason to go for a walk. But it does so only in the sense that it gives an instrumental point to my going for a walk: by going for a walk I can carry out an intention to do so that was prompted by my inclination to do so. I make two main points. One is that it does not seem appropriate to describe such a reason as a justifying reason for action even though it favours or recommends the action. The other is that not all actions even call for justification. In this respect actions differ from beliefs. It is in the nature of beliefs that normative reasons for belief should justify those beliefs, but it is not in the nature of action that normative reasons for action should justify those actions. No doubt some actions call for justification but we should not expect consideration of the nature of action to establish that all actions call for justification. Where justification is called for this will be because some particular evaluative perspective is in play. In the course of the discussion of these issues I consider the classical theory that the good stands to action as the true stands to belief (pp. 63-68). I argue that the classical theory is implausible since it fails to accommodate deeply perverse actions in a satisfying way.

## **7. Intricacies of normativity**

On my account, we incur a normative commitment to doing what is necessary in order to do what we intend. We can capture this commitment in terms of 'ought' by saying that we ought to avoid intending to do something while never getting around to doing the thing we intend. On the picture of reason-based normativity that I commend this 'ought' is ultimately grounded in a reason that is bound up with an ideal of rationality, in particular, the means-end ideal: avoid retaining an intention while never getting around to taking the necessary means (pp. 76-83). I call this an ideal rather than a requirement. To call it a requirement would suggest that any case in which we retain an intention while never getting around to taking all of the necessary means would be a failure of rationality. This is clearly not so. Such cases may be due to non-culpable ignorance of all of the means necessary to carry out an intention or to interfering factors that one is powerless to prevent (pp. 76-77). Nonetheless, there is reason to avoid retaining an intention while never getting around to taking the necessary means: to do so would put one in breach of the means-end ideal. The reason is strong enough to impose a requirement upon us: to strive to conform to the means end-ideal.

In Ch. 3, as well as setting out my view of the shape of various normative commitments, along the lines set out in section 2 above, I explore some tricky issues concerning criteria for being normative. A natural thought is that any statement is normative if it entails a paradigmatically normative statement. But what are paradigmatically normative statements? Consider the statement that there is milk in the fridge. This is presumably not normative. It entails that either there is milk in the fridge or I ought to give up smoking. Suppose that the second disjunct here is normative and that the disjunction is normative on that account. It follows that the statement that there is milk in the fridge is normative, which is not a very happy result. I argue that the disjunction is not

normative. Paradigmatically normative statements are statements to the effect that there is a (normative) reason for somebody to think or do something. A statement is normative if it entails such a statement. The disjunction just considered is no such statement. It does have normative subject-matter, as one might say, and thus is in some weaker sense normative, but it is important to keep the senses distinct.

## References

Blackburn, S. (1998), *Ruling Passions*. Oxford: Clarendon Press.

Foot, P. (1972), "[Morality as a System of Hypothetical Imperatives](#)." *Philosophical Review* 81: 305-306.

Gibbard, A. (1990), *Wise Choices, Apt Feelings: A Theory of Normative Judgement*. New York: Oxford University Press.

Harman, G. (1977), *Morality: An Introduction*. New York: Oxford University Press.

Kripke, S. A. (1982), *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.

Mackie, J. L. (1977), *Ethics: Inventing Right and Wrong*. Harmondsworth, Middlesex: Penguin Books.

McDowell, J. (1994), *Mind and World*. Cambridge, MA.: Harvard University Press.

# The Normative Dimension of Thought and Action: Alan Millar on Interpersonal Understanding

MARIO DE CARO

(University of Rome 3)

Not many contemporary philosophers are well acquainted with the work of the Greek poet Archilocus. A lot of them, however, are familiar with one particular Archilochean fragment, that saying, “The fox knows many things, but the hedgehog knows one big thing”. This is because that fragment was made famous by Isaiah Berlin, who used it for labelling what he thought were the two fundamental species of thinkers. The scope of Berlin’s taxonomy, however, can be instructively limited to contemporary philosophers.

In this light, we could say that, on the one hand, there are the hedgehog-philosophers, whose reflections turn around a single idea, which they analyze and discuss in all its facets and implications with a very rigorous (sometimes even hair-splitting) method. Many contemporary analytical philosophers belong to this group. On the other hand, there are the fox-philosophers, who are much broader in their interests, and tend to connect different areas in interesting (sometimes superficial) ways. Most Continental philosophers belong to this group, but also some analytic philosophers (see below). Berlin’s taxonomy, however, does not state the whole truth about what kinds of philosophers are around nowadays. There is, in fact, also a third, hybrid species, which exhibits the features of both the foxes and the hedgehogs (we can call them ‘hedgefox-philosophers’ – however puzzling this may sound from a purely zoological point of view). In this age of either hyperspecialization or superficiality, generally it is a healthy intellectual experience when, in exploring the philosophical ecosystem, one runs into a hedgefox-philosopher. They tend to be focused upon one specific issue, but they are also interested in studying the relevance and the implications that the issue has in different areas. As to their philosophical method, they are both analytic and synthetic.

Alan Millar is one of these philosophers. In *Understanding People. Normativity and Rationalizing Explanation* he studies, both in depth and in breadth, one of the hottest philosophical issues of the last decades, that which concerns the scope, modalities, conditions and limits of our ways of interpreting other human beings. According to an important tradition – which includes Gadamer, Ricoeur and Davidson – this question is the most crucial of all philosophy. Indeed, many difficult questions can be asked with regard to it. The method we use for interpreting adult mentally unimpaired human beings, by attributing them both thoughts and meanings their utterances they produce, is in fact very peculiar and enigmatic. How does it work precisely? How good, really, is it? Moreover, how does it fit in our general epistemic framework – dominated, as it is, by the nomological quest of the natural sciences?

Donald Davidson was perhaps the philosopher who did the most to put the issue of interpersonal interpretation in the centre of the philosophical stage. In doing so, he strongly connected thought and action with rationality: “To the extent that we fail to discover a coherent and plausible pattern in the attitudes and actions of others we simply forgo the chance of treating them

as persons (quoted, from Davidson's "Mental Events", at p. 222)<sup>1</sup>. However, Davidson worked on this issue in a 'hedgehogish' way. He developed very original views (about, for instance, 'radical interpretation', the role played by the principle of charity and the irreducibility of our interpretive practices to the explanations of the natural sciences) that had, and still have, a vast influence on the debate. However, he did not spend much time and energy in defining precisely the issue, in clarifying its controversial points and in making his arguments rigorous and sharp. Also, Wilfrid Sellars, John McDowell and Robert Brandom – the first two, at least, being philosophers also of a hedgehogish orientation – offered important contributions to understanding the issue. A lot, however, still has to be done.

In *Understanding People*, Millar argues that the kind of understanding implied when we aim at interpreting people is in principle different from, and irreducible to, scientific understanding. This is because, as is argued in Chapter 1, interpreting people requires that we see them as *rational agents*, that is, as entities whose actions are governed by normative reasons, so that, from our point of view, those agents tend to maintain an acceptable degree of coherence and cogency in their reasoning and acting. With Millar's words, "in thinking of the agent as having come to think something or do something for a reason, we think of the agent as taking his or her reason to be one in the light of which it makes sense to think or do that thing" (p. 230).

All this is very reasonable. However, there is, in connection with this, an important issue, which Millar does not discuss at length, which gave a lot of trouble to Davidson. Even if one grants that *from the interpreter's point of view* an interpreted speaker is reasonably consistent and coherent, how can we be sure that what he or she thinks is in fact correct? Could the interpreter and the interpreted speaker *both* have gone astray? Could they both be substantially wrong about how the world really is? This is, of course, a version of the classic scenario. In order to answer this challenge, Davidson first tried to appeal to the conceptual possibility of an omniscient interpreter who, in his opinion, *should* be largely in agreement with the non-omniscient interpreters, so proving that the latter cannot be largely wrong. This argument, however, sounded unbearably metaphysical to most critics, and later in his career Davidson tended not to repeat it. Instead, he proposed a view known as "triangular externalism", with which he offered an answer to the sceptical challenge by referring to an externalist view of thought. This solution, however, has been judged by many interpreters as partly circular, and therefore it is not really satisfactory. The sceptical challenge, therefore, still lurks in the background of any theory of interpersonal interpretation that is centred on notions such as coherence or consistency. This is because these are logical properties that *in themselves* do not seem able to guarantee a reasonable amount of correctness to the conceptual systems of both the interpreter and the interpreted speaker. I suspect that Millar sympathizes with the externalist attempts to answer that challenge (or to dismiss it as philosophically irrelevant!): in this light, for example, one can perhaps read Millar's remarks that in interpreting a speaker one has to consider his or her being or not being in touch with reality. This, however, does not solve the problem, since the question is how we, the interpreters, can be sure that the interpreted speaker is *actually* in touch with reality – even if this is what it seems like to us. I would like to know more from Alan on this issue.

Millar's view of interpersonal understanding, however, is actually a consequence of a more basic thesis that he defends in the first part of his book, according to which thought in general (and, therefore, also thought when we attribute it to other human beings) is governed by reason, as long as it is intrinsically bound by normative commitments. This also explains why understanding others is different from understanding things when one assumes a scientific perspective. In the latter case, one aims at (nomologically) explaining how things *are*; in the former case, it is necessary to consider how agents *should* think and act.

In Chapter 2, Millar argues the original thesis that there is an asymmetry between the

---

<sup>1</sup> All quotations are from Alan Millar, *Understanding People. Normativity and Rationalizing Explanation*, Oxford University Press, Oxford 2004.

justification of belief and that of action. While in the latter case, normative reasons have only an instrumental value, in the former case, normative reasons tend to offer a positive contribution to the justification of the belief.

In Chapter 3, Millar discusses why normativity is essential for beliefs and intentions, by describing two kinds of commitments associated with them: the “implication ideal” (which “provides a reason to avoid believing *P* while giving verdict to any implication of *P* other than belief”, p. 92) and the “means-end ideal” (which “provides a reason to avoid retaining an intention while never getting around to doing whatever is necessary if one is to do thing intended”, *ibid.*). In this light, beliefs and intentions should be considered to be, as Millar puts it, “psychological commitments” (p. 230).

In Chapter 4, the dispositionalist view of beliefs and intentions is accurately criticized, exactly because this view cannot account for the essential normative components of these propositional attitudes.

In Chapter 5, Millar defends a “high conception of the space of reasons”, which is at odds with many of the views that are fashionable nowadays. This is the idea that *reflexivity* (the capacity to think about our own “claims, beliefs, etc. and about the commitments and responsibilities that they incur”, p. viii) is central in thinking. Consequently, we can attribute a genuine capacity of thinking only to unimpaired adult human beings. “What about children, chimps and pets, then?” many would object. Here Millar’s analysis is very subtle (and clearer and deeper than those offered by other philosophers who have defended views similar to his, such as Davidson and Dummett). This chapter should be recommended to anybody who wants to keep thinking that full-blooded language and thought are essentially connected – as the founders of the “Linguistic Turn” supposed.

In Chapter 6, Millar argues, more generally, that in virtue of their conceptual content *all* propositional attitudes, not only beliefs and intentions, have an essential and irreducible normative dimension. This chapter begins with a detailed and stimulating discussion, from a normativist point of view, of Kripke’s critique of dispositionalism, and is centred on the idea that “meaning something” by using a linguistic term implies that, generally, a speaker participates in practices (conceived as *essentially* rule-governed activities) that determine what the correct applications of that term are. By paraphrasing Gibbard’s slogan “Means implies ought”, Millar summarizes his view by saying that “Means implies (normatively) committed” (p. 160). Interesting and original remarks about the kind of practices that are relevant for determining the correct applications of linguistic items are offered in the rest of the chapter.

In Chapter 7, Millar discusses the problem of the explanatory relevance of rationalizing explanations (this discussion is connected with that offered in Chapter 1, section 3). Millar argues for three theses, among which, *prima facie*, there seems to be tension:

- a) Rationalizations are a specific kind of causal explanation.
- b) Rationalizations implicate generalizations.
- c) The rationalizing component of these explanations is explanatory relevant, i.e., it is essential to such explanations that the agent acknowledge the normative relations that connect the *explanandum* (say, a belief or an action) to the *explanans* (some relevant reasons for having that belief or performing that action).

The challenge is that of explaining how c) is true once a) and b) are accepted (once, that is, we have a causally explanatorily relevant generalization that explains the formation of a belief or the performance of an action). Millar does an excellent job in proving that the rationalizing component adds a fundamental extra component to the merely causal component of the explanation. The kind of generalization implied here concerns the disposition of (rational) agents to recognize the normative reasons that rationalize his or her beliefs and actions. As Millar puts it in the case of a

formation of a belief, “the conclusion-belief [is] formed at least in part *because* the considerations constituting the contents of the basis-beliefs stand in a reason-giving relation to the conclusion” (p. 198).

This intrinsically normative component of rationalizing generalizations is what makes them different from scientific generalizations. As Millar puts it, as opposed to what happens in the case of the latter kind of generalizations, our epistemic access to rationalizations is *essentially* via normative considerations.

In Chapter 8, two theories of personal understanding are evaluated, the so-called ‘theory-theory’ (according to which commonsense psychology works as a proto-scientific theory – a very good one for some philosophers and a very bad one for others) and ‘simulation theory’ (according to which, in order to understand people’s actions and thoughts, we have to simulate them). Millar argues that both these theories are much less satisfying than his normative approach, according to which understanding others requires the interpreter to take into account what the interpreted agent is rationally committed to doing. It is true that the upshot of considering what people are committed to doing may coincide with the results we would obtain in case we tried to simulate them (that is, in case we tried to put ourselves in their shoes). However, as Millar notices, “putting rationality considerations centre-stage in relation to personal understanding does not commit us to putting simulation centre-stage” (p.225), since “it is not similarity considerations that drive attempts at understanding, but normative considerations” (p. 229). Millar also moves a strong and well argued attack against the scientism which generally is implicit in the theory-theory. As he notes, “the theory-theory conceives of theories as natural science conceives of theories. Crudely speaking, theories do their work by representing there to be (non-normatively specifiable) uniformities in nature, and explaining and predicting in terms of this” (p. 223). According to Millar, on the contrary, “normativity is written into the content of our ascriptions of propositional attitudes and actions ... and normative considerations have an indispensable epistemological and explanatory role in personal understanding” (*ibid.*).

Chapter 9 is the most surprising of the whole book. Given Millar’s insistence on the relevance, peculiarity and irreducibility of rationalizations, one would expect to see him in the company of those who extol the allegedly vast explanatory and predictive power of rationalization. The most famous member of this party is probably Jerry Fodor, who wrote:

It’s not possible to say, in quantitative terms, just how successfully commonsense psychology allows us to coordinate our behaviours. But I have the impression that we manage pretty well with one another; often rather better than we cope with less complex machines (Fodor, from *Psychosemantics*, quoted by Millar at pp. 241-242).

Millar thinks that Fodor is way too enthusiastic in this respect. He notices, for example, that there may be desires that are not motivated by reasons, but that still have explanations lying outside of commonsense psychology. Millar also notices that the hyper-enthusiastic defenders of the allegedly formidable predictive and explanatory powers of commonsense psychology tend to underestimate the relevance that knowledge of the routines, practices and roles, which are inherent to institutions, has for understanding and predicting other people’s actions and thoughts. In general, in his opinion, there are not only some kinds of situations in which it is typical, and unsurprising, to be unable to find acceptable rationalizations, but there are also situations in which no good rationalizations are available at all where one could reasonably expect some to exist.

However, consistent with the previous chapters of the book, Millar’s criticism of the Fodorian view does not ally him with Paul Churchland’s eliminativist perspective with regard to commonsense. According to Churchland (in “Eliminativist Materialism and the Propositional Attitudes”, quoted by Millar at p. 231), commonsense psychology does not adequately explain many mental phenomena, including, “the nature and dynamics of mental illness, the faculty of mental illness, the faculty of creative imagination, .. the ground of intelligence differences between

individuals... [Moreover it totally ignores] the functions of sleep, the ability to catch a ball on the run, the variety of perceptual illusions, and the ‘miracle of memory’”. As is well known, from these premises Churchland draws a very dismissive conclusion about commonsense psychology, since in his view it is “a stagnant or degenerating research program” (*ibid.*).

Millar disagrees with Churchland as much as he does with Fodor. First of all, as said in Chapter 8, he refuses Churchland’s (and Fodor’s, as to this) theory-theory approach to commonsense psychology. Moreover, and more relevant in this context, he argues that Churchland overstates his case. Millar has to be praised for his anti-ideological way of arguing such a delicate issue. In fact, he defends three very sensible theses, one of a Churchlandian tone, one of a Fodorian tone, while the last is the one that Millar carefully argues for in this book. They respectively are:

1. There are important limitations to the explanatory and predictive powers of commonsense propositional-attitude-based psychology.
2. The framework of commonsense psychology has to be preserved.
3. The conceptual framework of commonsense psychology essentially encompasses normativity.

An interesting point that Millar discusses in arguing for the first thesis is that of an agent who makes a choice that is perfectly reasonable for him/her, but could have also made a different but still reasonable choice (this is not only true of trivial choices, but also of important ones, such as what university a student decides to go to). In general, Millar’s point seems to be that there is no reason to think that in such cases *there must be* a (potential) intentional explanation of a choice based on offering its *sufficient* conditions. If so, we can very plausibly infer the falsity of psychological determinism. Millar also grants that these kinds of choices could be perhaps be accounted for by explanations that would be “at the interface between physiology and thought” (p. 240). This, however, does not mean that such explanations have to exist; and even less, that they could deterministically account for those choices.

With regard to this issue, a couple of questions I would like to ask Alan concern the kind of relations, both in the ontological and in the epistemological sense, he sees between the space of reasons and the physical world. As to the ontological issue, as said, he refuses to consider eliminativism of the mental as a real option. Once this reasonable move is made, however, the crucial point is to establish what kind of *real* autonomy, if any, the space of reasons has with regard to the world of the nature, which is nomologically described by physics and the other natural sciences. Such autonomy is not obvious: if it is too small, the space of reasons risks to be only given an instrumental value or to be regarded a pale shadow of the natural world; if it is too wide, supernaturalism immediately lurks. As to the first case, for example, one can consider Davidson’s anomalous monism, which (as convincingly showed by Jaegwon Kim) fails in its attempt at portraying the mental as really autonomous from the physical. The point is that Davidson tries to conjunct the (tendentially anti-reductionist) claim that the mental is anomalous with the (tendentially very reductionist) principle of the causal closure of the physical world, and the result this attempt produces is a view in which the mental is nothing more than epiphenomenal, since no mental property is left any causal power. On the opposite side of the philosophical spectrum, there are philosophers, such as Jonathan Lowe, who get all the mental causal power one could desire by simply dismissing naturalism altogether. But this is a move that I assume Millar is not ready to make. What then?

Notwithstanding Davidson’s failure, one could still try to work on a view in which the mental supervenes on the physical. In that case, however, the question is whether and how epiphenomenalism can really be avoided. An alternative is to conceive the relation of the mental with the physical as one of ‘constitution’ (as proposed by Lynne Rudder Baker) or one of

emergence. But of course these options have their own difficulties: most scientific naturalists, for example, insist that these views have the unacceptable cost of contradicting the conservation principles of physics. Indeed, there are several reasons (and some are, in my view at least, pretty convincing) to suspect that the ontologically pervasive interpretation of these principles offered by most scientific naturalists, and especially by physicalists, is not compelling. However, I wonder what Alan thinks of this crucial issue.

From an epistemological point of view, one could wonder what, in Millar's perspective, the relation between the intentionalist explanation of an action and the (very often only potential) naturalistic account of the related bodily movements (Benjamin Libet's famous, and controversial, experiments are a typical case-study here) is. Playing the devil's advocate, one could ask what would happen if one day, physicists, geneticists, neurophysiologists and evolutionary psychologists, by joining their scattered efforts, were able to realize the dreams of all reductionist philosophers, by obtaining a complete, and perfectly predictive, theory of our physical movements. What would we think, at that point, of our commonsense-based intentionalist account of actions? Would we still recognize any epistemic autonomy in them? Possibly this scenario is not really intelligible (and surely is absolutely implausible from an empirical point of view). However, it perhaps raises a problem that anti-reductionists cannot simply avoid.

Summarizing, in this important work Alan Millar explores the issue of interpersonal understanding. In doing so, he shows a remarkable competence in at least four different philosophical areas – practical philosophy, epistemology, philosophy of mind and philosophy of language – which nowadays are, on the contrary, more and more often kept surgically, if unwisely, separate. This unusual quality is another reason to recommend this book.

# Rationality, Animals and Causality

SIMONE GOZZANO

(University of L'Aquila)

Human beings are quite talented in understanding, predicting and explaining the behaviour of their fellows. Such talent reveals itself not only in familiar situations, but also in novel scenarios. How can we make sense of this ability, granting that this is the best way to conceive of it? What are its basic features, and how are these connected to one another? Is it a natural endowment or is it rather a cultural by-product? To put the same point in a more dramatic form: is this ability exclusively human, or are other species so endowed? These are the fundamental questions that everyone interested in the analysis of so-called folk or common sense psychology should consider, the same questions that form the core of Alan Millar's book *Understanding People*.

The answer that Millar provides is as clear as the questions themselves: the fundamental elements of our psychological life are beliefs and intentions, and these should be conceived of in normative terms, that is, as imposing on people commitments of various sorts. Beliefs, at the heart of our epistemic life, entail our undertaking epistemic commitments to other beliefs and, I would add, desires; intentions, which are fundamental for action, impose on us the planning of future courses of action in order to realize the intentions themselves. There is no point in having a belief if no knowledge is derived from it, if some knowledge deriving from it is needed; and there is no value in having an intention if one never tries, either through abstract planning or by practical attempts, to realize it.

Millar's first problem is to cope with the so-called "rationality assumption". On what basis does the game of ascribing beliefs and intentions to others rest? According to many authors, from Quine to Grandy but notably Davidson and Dennett, the ascription of propositional attitudes is based on the assumption that the "system" (to use a neutral term made famous by Dennett) to which intentionality is attributed is, by and large, *rational*. Very nice indeed, but how should we interpret this predicate? One of the important features of Millar's book is a detailed interpretative reading of this concept *vis a vis* the work it plays in helping us to understand people.

## 1. The rationality assumption

Before getting into the rationality debate it is useful to consider how intentional states are construed in Millar's book. These states are taken, in accordance with Russell's classical analysis (1918), as comprising a proposition within a psychological attitude, thus forming propositional attitudes. The individuating conditions of the two items are different. The attitudes reveal the psychological category of the systems' thoughts, whether beliefs, desires, intentions or what not. Such attitudes are, at least partially by Millar, individuated by considering the causal role they play in the overall psychological life of the system. So a belief, in general is the result of some perceptual state or the ending state of an inference and generates further beliefs or desires. An intention or a desire, vice versa, can be brought about by a belief and generally leads to action. Searle (1983) spoke of "direction to fit" to illustrate that beliefs are fixed by the way the world is, while desires are tantamount to having the world "fixed" by the way the desires are. What is to be

fixed or to fix, however, is given by the content of the propositional attitude, that is, by the proposition.

Here we come to the way propositional contents are individuated, and such individuation is where the rationality assumption gains traction. For the content of any given intentional state is the result of its “logical” relation to other contents. Here Millar is drawing on Davidson: “These relations between the attitudes are essentially logical: the content of an attitude cannot be divorced from what it entails and what is entailed by it” (Davidson 1995, p. 232). And again, “... a belief is identified by its location in a pattern of beliefs; it is this pattern that determines the subject matter of the belief, what the belief is about” (Davidson 1975, p. 168). However, the entailment relations among propositional attitudes are not independent from the intentional system (to use a Dennettian expression again): if the system believes that  $p$ , and  $p \rightarrow q$ , then the system will believe that  $q$  only if it has to take any verdict toward  $q$  (pp. 73-4). This is a shareable way of avoiding logical omniscience. Millar is quite cautious not to use too high a conception of rationality. Consider Dennett: “one gets nowhere with the assumption that entity  $x$  has beliefs  $p, q, r, \dots$  unless one also supposes that  $x$  believes what follows from  $p, q, r, \dots$ ; otherwise there is no way of ruling out the prediction that  $x$  will, in the face of its beliefs that  $p, q, r, \dots$  do something utterly stupid” (Dennett 1971, p. 229). This is fine, but how far should we go? Dennett has this problem in mind when he says:

A system’s beliefs are those it *ought to have*, given its perceptual capacities, its epistemic needs, and its biography. (...) A system’s desires are those it *ought to have*, given its biological needs and the most practicable means of satisfying them. (...) A system’s behaviour will consist of those acts it would be *rational* for an agent with those beliefs and desires to perform. (...) (Here) ‘ought to have’ means ‘would have if it were *ideally* ensconced in its environmental niche’ (Dennett 1987, p. 49).

Now, I think that Dennett’s contribution to setting the agenda which Millar pursues in his work is as important as Davidson’s, even if Millar is not taking Dennett as a primary source of inspiration.<sup>1</sup> Such an agenda can be depicted as having the following items set for future discussion: first, ascribing propositional attitudes entails getting into a holistic net of interdependent states. Second, these states enforce the, so to say, ascribed systems in assuming some commitments just by virtue of having them. Third, the relation between the states and the system determines the rationality of the system itself. Dennett’s view is that the three items in the agenda can be linked by taking the systems to be *ideally* ensconced in their environment: in such a case any system would act in accordance with its best beliefs and intentions given its knowledge and its needs. Davidson’s own way to construe the rationality assumption has a more transcendental flavour: ascription entails the mastering of the concepts involved in the contents of the states ascribed. The standards by which one has to conduct him or herself are established by various principles such as the requirement of total evidence of inductive reasoning and the like. So, these are the limits of the rationality assumption.

The fact is, however, that none is ideally ensconced in its niche, whether this is construed in purely naturalistic terms or not; nor does anyone strictly follow the principle of total evidence or any other strategy like means-end analysis. Everyone is, at most, sub-ideally ensconced, and the degrees and ways in which one is fall short of this ideal may, and in fact do, vary from individual to individual and, in the same individual, from time to time. So, we are back to the original problem: how should we characterize the rationality assumption?

Millar’s way consists in weakening the assumption without abandoning it: “The rationality assumption is short-hand for a number of more specific considerations having to do with coherence,

---

<sup>1</sup> Possibly, because Dennett adopts the rationality assumption and takes a naturalistic stance, one which Millar wants to avoid. This, in passing, shows that taking such an assumption is not enough to take issue with respect to naturalism.

cogency, and being in touch with reality” (p. 25) So, his way to cash out the rationality assumption is a *divide et impera*: splitting it into a number of different and, hopefully, more tractable issues. The unifying theme is that the rationality assumption is the way in which normative considerations get into the picture. This is why I considered Dennett’s words as particularly telling (as in this case Millar does as well): they show that to be rational is to be (en)forced to act according to one’s own internal states. Millar acknowledges that it is tempting to treat this normative link between internal states and other states or external behaviour as having to do with logic, with appropriate judgements or with regular patterns of behaviour, possibly based on the awareness of the psychological states motivating it.<sup>2</sup> But he refuses to adopt such guidelines: “I take normative reasons to be constituted by considerations, and I take normative reasons to apply only to creatures with the capacity to treat considerations as reasons”. (p. 32) So, *in* Millar’s view our psychological life is basically a commitment-making life based on the ability to conceive our psychological states (beliefs, desires, intentions) as considerations in favour of or against other psychological states or some course of action. (p. 120)

Let us consider beliefs in more detail. These are properly taken by Millar as aiming at truth. So, having the belief that  $p$  means taking commitments for the reasons that support the belief that  $p$ , the reasons that enforce us to take  $p$  as representing a truth.<sup>3</sup> These reasons, either perceptual or epistemic states, may operate at different levels. In one case we may have the sensitivity of our iris to light, which determines its shrinking when light increases. This is a sub-intentional capacity of the system in that the need to maintain a constant amount of light on the retina *strongly governs* the behaviour of the iris (cf. p. 53). The iris cannot do otherwise unless it is malfunctioning, and saying that the iris “believes” that this amount of light asks for this degree of contraction is a mere *façon de parler*. In another case, we have systems, such as ourselves, that can consider the aiming at truth of some of their internal states from a first-person perspective, ranking them with respect to their goal and interests: “... the agents in question are guided by *what strikes them* as being necessary to achieve the goal.” (p. 54) Our epistemic life works more or less in the same way: given any belief  $p$  I have, I incur commitment in accepting (and so believing) whatever  $p$  entails and whatever is entailed by  $p$ , given what else I do believe, if I have to give any verdict on the entailed propositions (cf. p. 73 *et passim*). This gets to an ideal of reason that Millar considers without embracing uncritically. This principle says: “For any  $\pi$ ,  $\Phi$ , if  $\Phi$  is implied by  $\pi$ , then avoid believing that  $\pi$  while giving a verdict on  $\Phi$  other than belief” (p. 76). As he says, the idea of taking a verdict is his way of discharging many problems concerning crazy beliefs and the like (Précis, p. 15).

One of the crucial elements of the principle is the universal quantifier “for any”: I think it creates some problems for the way in which the principle should be interpreted. Our epistemic conduct, the way we proceed from one thought to another, may follow very different strategies. We can reason deductively, inductively, abductively, we may perform non-monotonic reasoning, frame reasoning, prototypical inferences and so forth. All these methods have their own value and justification. It seems quite difficult, given a set of beliefs, to say in advance what is entailed by this set and what this set entails without specifying what kind of epistemic conduct we are to follow and what is the aim of our reasoning. Saying that having a belief entails having many of them, the holistic assumption, says nothing about the “logical role” that is in play among these beliefs. Given a set of beliefs, I can infer  $p$  or *not*  $p$  depending on my reasoning strategy and my interests: I can keep holding that *all birds can fly* after discovering that penguins cannot and taking them as an exception (so concluding that *all birds, excluding penguins, can fly*) or abandon the general belief and revise it accordingly, so concluding that *most birds can fly*. In this latter case, though, I won’t be impressed by the discovery that newborn birds cannot fly, and I will not modify the revised

<sup>2</sup> Authors that inspire such perspectives are Churchland, Kahnemann and Tversky, Dennett.

<sup>3</sup> I say “representing a truth” and not “as being a truth” because, even if propositions are abstract entities, we grasp them in a sentential form. That is to say, I think we have sentential attitudes, more than propositional, in that every thought we have has a mode of presentation that, as such, represents a truth, one and the same that can be grasped through a different mode of presentation.

belief. The reason I do not revise the belief again (so not getting to the belief that *most adult birds can fly*) depends on the prototypical assumption according to which capacities are shown in adulthood. Now, the way in which we cope with exceptions is at the heart of non-monotonic logic, and our considerations of whether to use non-monotonic reasoning or not may vary from context to context. Affirming that having a belief  $p$  entails believing, if a verdict should be taken, what is implied by  $p$  when it is not specified what the accepted logical relations are, is tantamount to affirming the holistic thesis that beliefs are placed in a net of other epistemic states, and cannot survive in isolation. But how can we say something specific about what is acceptable to infer? What I am suggesting is that there is not a further fact of the matter that determines acceptability. Piling up other beliefs by the subject would not create this further fact; it would simply blur the epistemic situation.

Millar presents an additional constraint: the entailment relation holds if it holds *without further ado*, that is, without the help of additional assumptions (p. 93). Such a restriction, however, does not avoid other potential difficulties: a logical intuitionist rejects the entailment having the form  $\text{not}(\text{not } p) \rightarrow p$ , while a logical classicist accepts such an entailment. Such an entailment holds without taking stock of other considerations and assumptions, so the problem surfaces again. Is Millar saying that it does not matter which method you follow as long as, once you have taken one, you follow it thoroughly? That is to say, is Millar ready to confine the validity of his thesis to contexts and domains? This option, however, does not make Millar's point, for in such a case how can we be sure that an individual who appears irrational to our eyes is in fact consistently following a method that we cannot even understand because, say, it requires very difficult logical structures that are beyond our cognitive capacities? The "projectivist" (cf. Fodor and Lepore 1990) strategy inaugurated by Davidson and elaborated by Millar is fine as long as normative considerations are limited.

These problems have an echo in the thesis by Arthur Prior (1961) according to which the idea of having the meaning of a word determined by its inferential rules raises difficulties because one may introduce a new logical constant, such as  $\text{TONK}$ , in a language  $L$ , by adopting very explicit rules that, as a matter of fact, allow the generation of paradoxes. Nuel Belnap (1962) has argued that introducing a constant by fixing the inferential rules for it is not enough: these rules must be a *conservative extension* of the rules governing  $L$ , where *conservative extension* requires that the new elements neither modify the logical consistency nor the expressibility of  $L$  itself. At the heart of the Prior-Belnap discussion there was the conviction, by the latter, that paradoxical languages are devoid of meaning and hence must be avoided. However, Cesare Cozzo has forcefully argued that "it is a fact that we are capable of constructing and of using languages that are paradoxical. ... important examples like *set theory* and the *calculus* show that in the course of the history of science paradoxical languages were fruitfully and lastingly used even if their paradoxicality was *well known*. Moreover we mostly have *no guarantee* that the languages we *now* use are not paradoxical." (Cozzo, 1994, p. 144-45) On the same score, we have no warranty that folk psychological talk is paradox free, and we keep using it or base on it normative requirements, as Millar does.

The same point can be somewhat applied to action, where the ideal of reason goes as follows: "For any  $\Phi$ , avoid intending to  $\Phi$  while never getting around to doing what is necessary if you are to  $\Phi$ "<sup>4</sup> (p. 76). In such a case, you cannot tell what the agent is going to do given her intention to  $\Phi$ : there are many ways of buying a present or preparing to travel, if this is what you intend to do. None of them is necessary for having the present bought or the trip arranged; rather they are all sufficient. So, if the ideal amounts to saying that if you want to  $\phi$  then do whatever you think can bring about  $\Phi$ , then it is equivalent to giving the *grammar* of realizing an action, but it cannot provide any substantial help for giving a normativity condition for an intention, i.e. for delivering any correctness condition for meeting the intention to  $\Phi$ . In fact, if what you intend to do is to buy a present and what is necessary to do to buy a present is preparing yourself to go shopping and the

4 Belief and intention are not perfectly parallel, an interesting difference that Millar notes.

like, then this is a truism more than an ideal of reason. However, Millar thinks that the only other option is dispositionalism, which he deems insufficient to capture the significance of having an intention to  $\Phi$  (more on this later).

These critical considerations on the rationality assumption are quite wide-ranging and can be summed up as follows: imposing normativity as a constraint that holds on the entailment relations among intentional states, or intentional states and action, means that there are logical, or practical, liaisons that are correct and liaisons that are not. However, the logical relations among intentional states can be enforced by very different methods, each of which could give rise to very different results. These differences may surface not only in case of very complex concepts or words, but also in basilar inferences, having to do with fundamental logical principles, or in the case of logical constants. If the situation is as portrayed, how can we gain any normative force by the mere acknowledgement that a system or a subject has followed, in a given context or situation, this or that method in her/is reasoning, if in other contexts s/he can legitimately use a different one? If we cannot give any judgement of correctness on the individual having followed this or that method, the normativity requirement seems idle. I am not saying that we should discard the rationality assumption, but it seems to me that the allegation of insufficiency addressed to other accounts can be made to Millar's view as well.

Millar has a (partial?) way out from the previous considerations when he focuses on normativity as a primary feature of statements. In so doing he distinguishes between statements that have just a normative import and those that are fully normative, the object of his chapter IV. Before getting into this I would like to point out that Millar says that "Normativity ... is primarily a feature of, for instance, judgements, beliefs, statements, claims - the sort of thing that can be true or false" (p. 92). Including beliefs in the list is somewhat ambiguous or potentially circular: statements and claims are clearly linguistic entities and their normativity would be a semantic feature of language. It is an open question whether the semantic value of beliefs (with respect to judgements I leave the issue open) is due to their intrinsic participation in the linguistic domain or is rather a basilar feature independent of language. This point has very important consequences for the issue of non-speaking creatures, a point I will explore later. So, I will treat "semantic entities", such as beliefs and statements, as those to which the different degrees of normativity apply.

Back to normativity. Millar thinks that there are beliefs/statements with a normative import and fully normative ones. He seems to accept that there are also beliefs/statements that are not normative, as is the case with the belief that Washington, D.C., is the capital of the United States (p. 93-4). So, we have three kinds of beliefs: the purely descriptive, those with a normative import and the purely normative, that is, those that somewhat explicitly mention reasons for the agent to believe or intend that  $p$ . Here are three examples on the belief side:<sup>5</sup>

- 1) John believes that it is past nine o'clock
- 2) John believes that he is arriving at the party later than nine o'clock
- 3) John has reason to believe that he is late for the nine o'clock party.

The normativity - absent in 1, indirectly present in 2, explicit in 3 - is a reason-linked feature of these statements. However, also 1 can be considered as bringing into play reasons: John is ready to give us his reasons in support of his belief, as Millar seems to recognize. Moreover, I have set the examples to show that 3 can be based on 2 which, in turn, entails 1. So, what is the difference, exactly? This lies in that in moving from 1 to 3 the role that reasons play in the fixation of the belief are more and more important. Consider a belief (statement) like this:

---

<sup>5</sup> Most of Millar's examples on this point are on the intention and action side, more than on belief and knowledge. I think this is not by chance: his thesis seems more compelling when referred to actions and their ethic values.

0) John believes that *this* [pointing] is red

Such a “zero level” belief (statement) has a very reduced degree of support in other beliefs (statements). However, a Davidsonian like Millar would say that the concept of *red* is not for free: you can have it only if you have some idea of being coloured, even if you do not need much else. Quite different is the case with *being nine o'clock*, a very abstract relation. When we consider something like 2, we are implicitly relying on other beliefs that John must have: that there is a party time, that arriving late is a disvalue etc. In case of 3, the focus is directly on the reason, more than on the propositional content that *he is late for the nine o'clock party*, so the impact this reason has on the belief ascription is stronger, but eventually Millar’s strategy is that of having two kinds of intentional states. Let me clarify.

Millar thinks that the principle or ideal of reason mentioned (“For any  $\pi$ ,  $\Phi$ , if  $\Phi$  is implied by  $\pi$ , then avoid believing that  $\pi$  while giving a verdict on  $\Phi$  other than belief”) is acceptable only if one takes beliefs and intentions as normative in the sense of being reason-linked (p. 118). In turn, one can have such beliefs and intentions only if one can directly or explicitly master the reasons for the beliefs and intentions ascribed. Here is Millar:

It is in the nature of the psychological commitment in which believing that  $p$  consists that the way it impacts on my thought and action is shaped by the implication commitment that it incurs –the normative commitment to believing any implication of  $p$  on which I give a verdict. The normative commitment can have this shaping role only if subject has the appropriate reflective capacities –including a capacity to reflect on the implications of things believed. ... My point is that, since beliefs in the realm of personal understanding ... implicate reflective capacities through the exercise of which the impact of those beliefs is shaped, we should acknowledge a distinction in psychological kind between those beliefs and any belief-like states the impact of which is not so shaped. (p. 123)

I have mentioned Millar at length because in raising another important topic (do non-human animals have beliefs?) he is making evident a common strategy among philosophers such as Davidson, McDowell, and Brandom, among others. According to them, there are two kinds of epistemic states: the “real” propositional attitudes of language speakers and the somewhat diminished states of dumb creatures. In the background the issue is: how much can the notion of a belief receive a naturalistic reading? This is quite a debated topic in the present literature in epistemology and in the philosophy of mind. So, let consider such a question.

## 2. Non-human animals and reflective thoughts

The issue of animal intentionality is taken up by Millar in order to show the difference between belief in which there is no normative import and one that is essentially individuated by its normative commitments. A few paragraphs ago I noted that Millar considers statements and beliefs on a par as to their normativity. If this approach can be somewhat accepted for human beings, it is risky to form a *petitio principii* in case of animal thinking, because it is exactly the linguistic status of belief (and intentions) that is at stake. So, where does the difference between human and animal beliefs lie? Millar thinks that animals can be credited with some sort of sub-doxastic states that guide their behavior in a way directly linked to environmental stimuli, a kind of state that we may have as well (see Stich 1978; Bermúdez 2003). However, belief-intention psychology can be credited only to us, because it is solely humans who have the appropriate reflective capacities, those that allow us to engage in deliberative thinking about our own beliefs and intentions. So, there seems to be some sort of irreducible gap between sub-doxastic states, those information-bearing states that manage our motor behavior in driving the car on a busy road or helping us to catch a thrown ball, and beliefs and intentions, those states whose content is consciously elaborated in a linguistic way (p.135; 150). However, consider what has happened to anyone when learning to

drive the car. We were conscious of our intentions, in a way that seems to meet the conditions set by Millar for the high-view of intentions. For instance, we were not only conscious that it was necessary to be smooth in relaxing the gas pedal, but we were conscious of the reasons for acting that way: a smooth acceleration would have resulted in a well conducted car, in not losing control of it, in not having the engine choking or eventually stalling. It may turn out that in those moments we controlled the car quite well. It may turn out that our behavior in that occasion was perfectly comparable with the one we presently have.

However, despite the behavioral similarity in our control of the car, now that we are skilled drivers, we do not engage in deliberative thinking about car-conduct any more. While driving we may engage in deep philosophical discussions (well, up to a certain extent, I must say...). What has changed? It seems that some intentional states are passed from a fully epistemic dimension to a sub-doxastic one or, to put it in Millar's terms, that there has been a passage from the normative dimension to the non-normative dimension, where non-normativity, or sub-doxasticity, is here individuated by the epistemic isolation of the relevant states. So, the two dimensions are not completely apart, at least not in us. Notice, moreover, that the intentional flow, so to say, may go in the other direction as well: when the situation asks for my full reflective capacities, I stop my philosophical discussion and engage myself in thinking what should I do in driving, so letting my sub-doxastic states resurface in the conscious dimension.<sup>6</sup> A further thing to be noted: a system with sub-doxastic states could be, in some sense, perfectly rational: it could employ some means-end strategy and some implication strategy even if it is completely unaware or unconscious of such strategies.

This example shows, I think, that the view according to which reflective thoughts are independent from, or at most based on, sub-doxastic states is not necessarily true: we can have sub-doxastic states developed from, and based on, reflective thoughts. Another impact this kind of examples may have for Millar's view can be found in his discussion of the simulation theory, in Chapter 7. Here, in tackling a line of reasoning by Stephen Stich, Millar notes that there can be two perfectly functionally specifiable systems sub-serving full propositional attitudes and sub-doxastic states respectively (p. 228). These two systems, one is invited to think, are entirely isolated from each other, so it seems that there is no way to transfer some epistemic or sub-epistemic content from one to the other. Again, such a view does not match with the example given.

Can we support the idea that, as to intentional states generically considered, there can be more than two kinds? I think we can. The view licensed by Davidson and Millar himself seems the following: animals' behaviour is directly driven by states that are not mediated by any kind of representation. In this sense the states are sub-doxastic: epistemically isolated and directly linked to the perceptual stimuli. Our behaviour, on the contrary, is always based on reflective capacities, also in case of epistemic states of perceptual nature: if John believes that *this* [pointing] is red, he believes so in a fully conceptual way, rather than in non-conceptual terms as animals may do.<sup>7</sup> The epistemic divide between humans and non-humans cannot be any larger. Now, I do not want to deny that we mainly have fully conceptual and reflective thoughts; rather I want to point out that there are other epistemic states beyond the non-doxastic ones as these have been construed by Millar.

A very simple non-doxastic state is one that drives the behaviour of the system endowed with it in a quite rigid way. My iris "knows" the amount of light in the surroundings, but I cannot derive any knowledge from the knowledge of this (sub)system of mine (which makes this a sub-doxastic knowledge) and I cannot actively do anything in order to correct it. If the system gives me incorrect information, this can be due solely to a malfunction, and this malfunction can be repaired just by

6 During the Sixties many chess programs in the artificial intelligence area have been developed by asking chess masters to make explicit their unconscious heuristics. See Simon (1982) for references to early studies and recent applications to economy.

7 In support of the non-conceptual, see Peacocke (1992) and Crane (1992); against it see Brewer (1999).

modifying the circuitry, either chemically or physically, of the system itself.

Consider now the case in which I see, as in the case made famous by Malcolm and Davidson, a cat that runs toward an oak tree but, at the last moment, swerves and disappears up a nearby maple. I believe the cat is on that [pointing] tree, and my dog can do the same. We are both wrong, though. Should we invoke a cognitive difference between me and my dog Fido in order to account for our mistake? True, I can believe of the tree that *it* is the oldest tree in the town, a kind of belief not available to Fido, but this would miss the point. The point rather is that our perceptual systems, mine and Fido's, are perfectly functioning and if a mistake occurred it cannot be repaired by modifying the perceptual apparatus circuitry; what I and Fido need is more information, and such information need not necessarily be fully conceptual. Spotting the cat on the maple would suffice, both for me and for Fido. I have called these states with content of type 2, distinguishing them from the non-doxastic states previously mentioned (content of type 1) and the fully conceptual (and semantically opaque) states (content of type 3) (see Gozzano in press).

So far I have been considering, with a critical look, the way in which animals' beliefs are conceived, and the contrast between sub-doxastic and reflective thoughts, on which the normative constraint can be applied. A further question, though, can be the following: is the way in which reflective thoughts are conceived fully defensible?

According to Millar (p. 138), to have the belief that  $p$  ( $Bp$ ) entails (in general) knowing that one has such a belief ( $KBp$ ). That means that  $Bp$  entails  $KBp$ . Similarly, for intending that  $p$ , in which case  $Ip \rightarrow KI p$ . At the same time Millar has to endorse that  $Kp \rightarrow Bp$ , a quite accepted thesis.<sup>8</sup>

This shows that, for Millar, since  $Kp \rightarrow Bp$  and  $Bp \rightarrow KBp$  (Millar's thesis) then  $Kp \rightarrow KBp$  (by transitivity). I think that this result is somewhat unrealistic. It entails that, for every proposition we know, we should have reflective knowledge on our believing it, and this overloads our cognitive capacities by doubling every knowledge we have, explicit or implicit, with a second order belief, either implicit or explicit. Secondly, if there is no constraint on the logical form of what is known, it is possible to generate a regress of this form: if  $KBp$  is a knowledge as is  $Kp$ , then it entails (by the accepted thesis) the belief  $BKBp$  that, in virtue of the same reasoning, gets to  $KBKBp$ , and so forth. This surely is an unpalatable consequence. So, at least a closing clause is needed.

Let now consider the "ideals of reasons". Millar thinks that if one has to take verdict on a proposition  $q$  ( $Vq$ ) then, if s/he believes that  $p$  ( $Bp$ ) and knows that there is an entailment relation between  $p$  and  $q$  ( $Kp \sqsupset q$ ) then s/he ought to believe  $q$  ( $OBq$ ).

So, let's suppose that I have to take verdict on  $q$  and that the above described thesis holds. This can be formalized as follows:

$$Vq \rightarrow (Bp \& Kp \sqsupset q \rightarrow OBq) \quad Vq$$

By *modus ponens* we get to

$$Bp \& Kp \sqsupset q \rightarrow OBq$$

Let us now suppose that as a matter of fact we

$$Bp \& Kp \sqsupset q$$

By *modus ponens* again we get to

$$OBq$$

Now, let's introduce the implication discharging the first premise  $Vq$

$$Vq \rightarrow OBq$$

---

<sup>8</sup> This is so because if  $Bp \rightarrow KBp$ , if I believe that  $p$  and I know that  $p \rightarrow q$  then, if I have to give verdict on  $q$ , I can believe that  $q$  ( $Bq$ ) only if  $K(p \rightarrow q)$  entails that I  $B(p \rightarrow q)$ , which is to say that, in general  $Kx \rightarrow Bx$  for any  $x$ .

Finally, let's introduce the implication by discharging the second added premise

$$(Bp \& Kp \supset q) \rightarrow (Vq \rightarrow OBq)$$

So, if one believes that  $p$  and knows that  $p$  is somewhat entailed by  $q$  then, if one has to take a verdict on  $q$  one ought to believe it. Modularity here determines a problem. For, suppose that the issue you have to take a verdict on is the content of Müller-Lyer (M-L) illusion, or any other kind of persistent illusion. The persistency of some illusions is due, so the story goes, to the informational encapsulation of the modules that treat them. So, independently from our knowing that the two segments of the M-L illusion have the same length, they appear to us as different. These cases show, in general, that our inferences may mix beliefs, as mental states with conceptual content, with beliefs as mental states with non-conceptual content, one along with the other in a hybrid succession of thoughts and proto-thoughts. So, a *prima facie* reading would be that of taking proposition  $p$  as “the segments have different length” and proposition  $q$  as “the segments have same length”. It is difficult, though, to understand how there can be an entailment relation between  $p$  and  $q$  other than contradiction; it would be unreasonable, then, to suppose that by believing that  $p$  one should be led to believe that  $q$ .

Another option is to take  $p$  as “the segments seem to be of different length”. In such a case, even if it seems to be logically odd to insert the *appearance* within the propositional content of the belief, it would be possible to have an entailment from  $p$  to  $q$ . So, if I were to believe that  $p$  and knowing that  $p \supset q$ , I ought to believe that  $q$  were I to take verdict on it. Here the crucial step is the kind of epistemic state  $p$  belongs to. In fact, the  $p$  content, inasmuch as it is informationally encapsulated, could be appropriately considered as a sub-doxastic state of the same kind non-speaking animals are endowed with; at the same time, content  $q$  is a fully conceptual propositional content, as those that are proper of human beings alone. Now, if the entailment relations among propositional contents were supposed to isolate only those contents on which the normative discourse is in force, then there should be no entailment between  $p$  and  $q$  because  $p$  is not in the normative domain. On the other hand, if one wants to bite the bullet and to consider  $p$  as fully conceptual and hence normative, since there is no way to modify it permanently, we have to give up, or at least weaken, the idea that all propositional attitudes are connected in a net of logical relations and thus individuated.

The general problem that this argument brings to the stage is the extent to which Millar adheres to a coherentist view of belief and intentional states. Davidson and McDowell have both endorsed the view according to which our beliefs are justified and supported solely by other beliefs. So, every belief we have is the result of an inference. The status of perceptually fixed belief, though, is vexed. So, it seems that either adherence to coherentism is complete, and then it must be explained why some belief cannot be modified as is the case with illusions, or it is limited. In both cases, however, the idea of individuating intentional states through their logical relations seems in jeopardy.

A somewhat related theme is the attack on dispositionalism, a sort of important detour. Millar thinks that the only other game in town is the view that the ideals of reason, both the one that concerns belief and the other on intentions, can be construed in terms of complex dispositions. So, to set a trip to the sea is to be ready to take a bathing suit and a towel and set off for the shore (see Ryle, 1949). Dispositionalism, though, is not a viable option because it can license a contemplative view of our mental life, analogous to the attitude we can have toward our character traits. The basic intuition behind dispositionalism is that beliefs and intentions are dispositional properties, like fragility. A dispositional property is causally set off by specific triggering conditions; once these conditions are in play, the property gets activated unless other properties prevent it from manifesting. If a system has a dispositional property  $D$  and the system is in the appropriate conditions then the property shows its causal effects. Secondly, dispositional properties can be considered as supervening on categorical properties (the fragility of the glass supervenes on its molecular structure) (cf. Armstrong 1968 and Mumford 1998 for clarifying such a view). In such a

case, dispositionalism would accept the contemplative interpretation of the dispositions in the sense that dispositions as such would not have causal powers over and above those of their categorical bases, as is the case with sub-doxastic states. However, some systems can avoid the purely contemplative stance toward the property by setting themselves into a condition that would prevent a given property from setting off. I long for a cigarette, so I go to a place where smoking is forbidden, forcing myself to respect my promise not to smoke. It seems that one cannot make sense of such a possibility without endorsing in some way Millar's view that we have capacities that transcend the purely dispositionalist view. However, it seems to me that taking dispositions as supervening on categorical bases does not entail having a contemplative view toward them. Dispositionalism, as specified, can be taken as a thesis concerning beliefs formation: I see a red car and I come to believe that there is a red car in front of me. My perceptual apparatus is structured in such a way that if there is something red in front of me I come to believe that something red is in my visual field. Millar resists calling these beliefs; rather he takes these as purely information bearing states. If so, the problem lies in the concept of *belief* rather than in the dispositionalist view, a concept whose subtleties have been the matter of my previous concerns.

### 3. Last remark

In closing his book, Millar briefly surveys two different naturalistic approaches to common sense psychology. Both assume that there is a conditional thesis to the effect that if folk psychology is true, then there must be some physical properties realizing what is posited by common sense psychology. However, the two approaches part company on how to face such difficulties. On the one hand, Jerry Fodor (1987) thinks that if we cannot confirm the expected properties it is necessary both to analyze again the generalizations of folk psychology and to look for deeper structures of the mind, in line with researches in computational linguistics. On the other hand there are those, like Paul Churchland, who think that if the expected physical properties cannot be confirmed, so much the worse for commonsensical generalizations: these should be abandoned. Even if Millar recognizes, with Churchland (1981), that there are serious limitations to folk psychology, he denies the very conditional that is held by naturalists. The propositional attitudes posited by common sense psychology should not necessarily be conceived of as mirrored by specific physical properties. "On the conception developed in this book, there are dispositions – albeit normatively specifiable dispositions – that are characteristic of the attitudes. But the constraints that the dispositions place on the physical states of the subjects are loose". (p. 234) However, at the beginning of the book we are told that attitudes get their individuation, at least in part, because of the causal role they play (p. 8). So, if these dispositions do not causally force the structure of our psychology, the how and why we act and think, it seems that Millar is taking the causal role of intentional states as loose. If this is so, then either the individuation conditions of the attitudes are loose, but in such a case how can these attitudes support deep normative considerations, or the causal efficacy of them is lost, and in this case how can we be said to be acting and thinking by virtue of our intentions and beliefs? In both cases the prospects are difficult. However, I think that Millar has been very able in refining and ameliorating the normative view of propositional attitudes, and every naturalist has to confront her or himself with such a deep and mature new theory.<sup>9</sup>

## References

Armstrong, D. (1968), *A Materialist Theory of the Mind*. London: Routledge.

Belnap, N. (1962), "[Tonk, Plonk, and Plink.](#)" *Analysis* 22: 130-134. 2

<sup>9</sup> I would like to thank Patrizia Pedrini and Luca Malatesti for helpful comments on a previous draft. I am deeply grateful to Donald Challenger for his editing of the paper.

- Bermúdez, J. L. (2003), *Thinking Without Words*. Oxford: Oxford University Press.
- Brandom, R. (1994), *Making it Explicit*. Cambridge Ma: Harvard University Press.
- Brewer, B. (1999), *Perception and Reason*. Oxford: Oxford University Press.
- Churchland, P. M. (1981), "[Eliminative Materialism and Propositional Attitudes](#)". *Journal of Philosophy* 78: 67-90.
- Cozzo, C. (1994), *Meaning and Argument*. Stockholm: Almqvist & Wiksell.
- Crane, T. (1992), "The nonconceptual content of experience" in T. Crane (Ed) *The contents of experience*. Cambridge: Cambridge University Press: 136-57.
- Davidson, D. (1995), "Donald Davidson", in S. Guttenplan 1995. *A Companion to the Philosophy of Mind*, Oxford: Blackwell: 231-6.
- Davidson, D. (1975), "Thought and Talk", now in D. Davidson 1984. *Inquiry into Truth and Interpretation*. Oxford: Clarendon Press: 155-170.
- Dennett, D.C. (1971), "[Intentional Systems](#)". *Journal of Philosophy* 68: 87-106. Now in D. Dennett 1978. *Brainstorms*. Cambridge Ma: Mit Press.
- Dennett, D.C. (1987), *The Intentional Stance*. Cambridge Ma: Mit Press,
- Fodor, J. Lepore E. (1990), *Holism*. Oxford: Blackwell.
- Fodor, J.A. (1987), *Psychosemantics*. Cambridge Ma: Mit Press.
- Gozzano, S. (In press), "The beliefs of mute animals." in M. De Caro, F. Ferretti, M. Marraffa (Eds) *Cartographies of the mind*. Dordrecht: Kluwer.
- Mumford S. (1998), *Dispositions*. Oxford: Clarendon Press.
- Peacocke, C. (1992), *A Study of Concepts*. Cambridge Ma: Mit Press.
- Prior, A. (1960), "[The Runabout Inference Ticket](#)". *Analysis* 21: 38-9.
- Russell, B. (1918), "The Philosophy of Logical Atomism." *Monist* 28: 495-527; *Monist* 29: 32-6; 190-222; 345-380.
- Ryle, G. (1949), *The Concept of Mind*. London: Hutchinson.
- Searle, J.R. (1983), *Intentionality*. Cambridge: Cambridge University Press.
- Simon, H.A. (1982), *Models of Bounded Rationality*. Cambridge Ma: Mit Press
- Stich, S. (1978), "[Belief and subdoxastic states](#)". *Philosophy of Science* 45: 499-518.

# Extending the Space of Reasons: Comments on Chapter Four of *Understanding People*

ADRIAN HADDOCK

(University of Stirling)

1. Wilfrid Sellars employs the metaphor of the space of reasons to express a certain conception of knowledge: “in characterising an episode or state as that of *knowing* ... one is placing it in the logical space of reasons, of justifying and being able to justify what one says” (1956: 289-9). A growing number of philosophers employ the same metaphor to express a conception of at least some (other) mental states: in characterising a state as that of *belief*, or *intention*, for example, one is placing it in the same logical space. The burden of Alan Millar’s characteristically careful and thought-provoking book is to tell us what this conception amounts to, and to argue for its truth.<sup>1</sup> Its central claim is that the concepts of belief and intention, and what they are concepts of, are (in a sense to be explained) *normative*. Chapter four – “the heart of the book” in Millar’s view (*this issue*, p. 17) – is devoted to explaining, and defending this claim.

In these comments, I will focus on what Millar’s claim has to say about belief. I will suggest that the claim breaks down into three indistinct theses (§2), examine his argument for these, and claim that it is, at best, incomplete (§3). I will then develop a further problem for his claim (§4) and, finally, suggest that there might be a way to make sense of a conception of belief as a standing in the space of reasons that is not wanting for argument, and does not face this further problem (§5).

2. The three theses into which Millar’s claim appears to divide are: a strong conceptual thesis, a weak conceptual thesis, and a constitutive thesis. Each gives weight to the following principle.

*The Implication Commitment Principle:* For any rational subject *S*, and any proposition *P*, if *S* believes *P*, then for any proposition *Q*, if *P* implies *Q*, then *S* incurs a commitment to believing *Q*, if *S* gives a verdict on *Q* (p. 109).<sup>2</sup>

According to the strong conceptual thesis, the Implication Commitment Principle is analytically true. The claim that *S* believes that *P* brings into play the concept of believing that *P*, and “without further ado” – that is, simply on account of the content of the claim, and “without the help of additional assumptions, other than ones that merely make explicit features of [this] content” (p. 93) – implies that *S* incurs a commitment to believing any implication of *P* upon which she gives a verdict.

According to the weak conceptual thesis, “grasping the concept of belief commits one to accepting the [Principle]”. It is not possible for *S* to believe that she believes that *P*, but not be committed to believing that she incurs a commitment to believing any proposition *P* implied upon which she gives a verdict. And it is not possible for *S* to believe of another rational subject that she

---

1 All page references are to Millar (2004) unless otherwise specified.

2 The restriction to rational subjects is important. A rational subject is a subject in possession of capacities to reflect on her own states of mind, and Millar’s claim only concerns the species of belief that such subjects can enjoy, and its concept. (There is more on the idea of a rational subject in §5.)

believes that *P*, but not be committed to believing that this rational subject incurs the same commitment.

According to the constitutive thesis, the Principle is a “partial specification ... of what it is” (p. 118) to believe *P*; in other words, it is a partial specification of the nature of belief. It is part of what it is to believe *P* that *S* (who believes *P*) is committed to believing any proposition *P* implies upon which she gives a verdict.

What is the relation between the strong thesis and the constitutive thesis? If the Principle is an analytic truth – as the former maintains – how can it *not* be the case that the Principle gives a (perhaps only partial) specification of what it is to believe *P*? If<sup>3</sup> it is analytically true that bachelors are unmarried males, and unmarried males bachelors, it goes without saying that what it is for *X* to be a bachelor is (at least in part) for *X* to be an unmarried male. And it goes without saying not only because it is evidently true, but also because, in typical circumstances, the point of saying that a proposition specifies what it is to be something is to bring out that the proposition is, at the least, necessarily true, but not necessarily analytic. This seems not to be Millar’s point, because he seems to think the Principle *is* analytic. Perhaps he bothers to state the strong thesis *and* the constitutive thesis because he wants to reject the idea that his claim concerns only the concept of belief and not belief’s nature. But it is not entirely clear what this rejection amounts to. The strong thesis concerns the concept of belief; but because it purports to state an analytic truth, it cannot but concern the nature of belief as well. What kind of claim could concern the concept of *X* but not *X*’s nature?

One answer is: the kind of claim enshrined in the weak thesis. The weak thesis states that being committed to accepting the Principle is a condition for grasping the concept of belief; and, on the assumption that anything one must accept in order to grasp a concept must be true thereby implies the truth of the Principle. But, on the further assumption that the Principle can be true and the constitutive thesis false – an assumption that Millar seems to accept, as we will see – it does not follow that the constitutive thesis is also true. As a result, the weak thesis appears to be a claim that concerns the concept of belief, but neither makes nor implies a claim about belief’s nature. In saying that his claim concerns the nature of belief as well as its concept, Millar seems to be insisting that he is committed, not only to the weak thesis, but to the constitutive thesis as well.

This tells us why Millar bothers to state the constitutive thesis *and* the weak thesis. But it does not tell us why he bothers to state the constitutive thesis *and* the strong thesis. Perhaps Millar is not committed to the strong thesis after all. If he was, why would he bother stressing that he is *also* committed to the constitutive thesis, as opposed simply to reminding us that he is *obviously* committed to it, simply on account of his commitment to the strong thesis? This suspicion is compounded by the fact that, in these post-Quinian days, it is a little *recherché* – and so a little surprising – to find a philosopher boldly claiming to have laid his hands on an analytic truth, especially one that has been unacknowledged hitherto. Even so, Millar does *seem* to express a commitment to the strong thesis. In what follows, I will assess Millar’s argument for its success in establishing both the strong thesis and the constitutive thesis, where the latter is understood on the assumption that the Principle states a necessary, but not analytic, truth. As we will see, it will not matter in the end whether Millar is committed to the strong thesis or not.

Whichever thesis (or theses) Millar is trying to argue for, his argument attempts to show its (or their) superiority over an alternative position that rejects it (or them) but endorses the Principle. This argument can only get off the ground if it is possible to do what this alternative position claims. In what follows, I will consider whether it is possible, *en route* to suggesting that the argument seems not to succeed.

---

3 Why ‘if’? Well, which of the following are bachelors: a recently divorced male who has previously been married fifteen times; an unmarried male who has lived with, and been in a loving relationship with, the same partner for over sixty years; an unmarried male who is in a civil partnership with another unmarried male; an unmarried male who has never married and lives on his own; a married male who belongs to a society very different to our own, and was forcibly married at the age of one-and-a-half? Is this supposed to be a matter that a concept – and it alone – can decide? See Travis (forthcoming).

3. In the case of the constitutive thesis, the alternative position claims the following: even though the Principle is a necessary (but not analytic) truth, incurring the commitment specified in the Principle is not *part of what it is* to believe that *P*; in other words, it is not *constitutive* of believing – contrary to what the constitutive thesis claims. However, on the assumption that the alternative position is not simply a version of the strong thesis, and so does not maintain that the Principle is analytic, what exactly does this claim of constitution add to the claim upon which the constitutive thesis and the alternative position agree: that, necessarily (but not analytically), if *S* believes that *P*, then *S* incurs the specified commitment? In other words, is there is a difference between the constitutive thesis and the Principle, where the latter is understood as stating a necessary but not an analytic truth?

It is not obvious that there is. I am going to consider two ways in which it might be thought that there is, and suggest that neither is of help.

According to some philosophers, even though it is necessarily the case that, if I was not alive, then I would not be able to wiggle my fingers, nevertheless my being able to wiggle my fingers is not *constituted* by my being alive, because it is not in virtue of my being alive that I have this ability, but rather in virtue of my brain and body functioning in a certain sort of way.<sup>4</sup> We ought to be able to explain why this is true (if it is). And one explanation seems to be this: having a suitably functioning brain and body is necessary *and sufficient* for possession of the ability, but being alive is merely necessary. Applying this suggestion to the present case, we can say the following: according to the Principle, where *P* implies *Q*, and *S* gives a verdict on *Q*, *S*'s incurring a commitment to believing *Q* is necessary for *S*'s believing *P*; so, according to the constitutive thesis, where *P* implies *Q*, and *S* gives a verdict on *Q*, *S*'s incurring a commitment to believing *Q* is also sufficient for *S*'s believing *P*. But this is surely absurd. The proposition that it never rains in Southern California implies the proposition that it did not rain in Southern California on Friday 18<sup>th</sup> March 1983; and I can (of course) be committed to believing the latter without believing the former, for I might simply believe the (more modest) proposition that it did not rain in Southern California during the week beginning Monday 14<sup>th</sup> March 1983. Incurring a determinate commitment to believing *Q* cannot, then, be sufficient for believing *P*. To ensure that the constitutive thesis remains both a going concern, and distinct from the Principle, we need a different account of what constitution amounts to.

We can arrive at one such account by reflection on the case of dispositions and their grounds. Consider an aspirin: it has a disposition (to relieve headaches), which can be recorded in terms of a conditional (if *S* takes the aspirin, then *S*'s headache is relieved), and which has a ground (the chemical structure of the aspirin) that both explains why the aspirin has this disposition, and, importantly, can be specified without employing the concepts that figure in the relevant conditional. It is not simply that the conditional holds of the aspirin; it is also that there are features of the aspirin that can be specified from outside the conditional's concepts, and which explain why the conditional holds. This allows us to draw a distinction. It is one thing for it to be necessarily so that if ... then ..., and quite another for there to be features of the thing of which this holds that explain "from outside" why this is so. In the latter case, but not in the former, the conditional is not only a necessary truth but also a (perhaps partial) specification of what it is to be the thing in question. We can now say the following: if the Principle is to be a partial specification of what it is to believe that *P*, there must be an explanation of why this Principle holds that appeals to features of belief that can be specified "from outside".

Millar explicitly rejects the possibility of providing an explanation of why belief is commitment-incurring that satisfies this "from outside" requirement. He notes that certain ethicists attempt to explain moral principles by appealing to its consequences for a feature that can be specified "from outside"; as when consequentialists attempt to explain why killing is wrong by

4 Ram Neta (2008) is one philosopher who argues in something close to this way.

appealing to its upshot for the general good. But he thinks that attempting to do something similar in the present case – by appealing to some kind of independently specifiable value that will be attained if we conform to the Principle – is “odd”, and “leads to a dead-end” (p. 117). So, it looks as if, by Millar’s own lights, we cannot make sense of a distinction between the constitutive thesis and the Principle, if we take the aspirin case as our model.

But perhaps we can still make sense of the distinction. The constitutive thesis purports to specify part of the *nature* of belief, and says that it is part of this nature that, if *S* believes that *P*, then *S* incurs the commitment specified in the Principle. In so doing, it seems to contrast with a different thesis, according to which it is not part of the nature of belief, but part of the nature of *rationality* that, if *S* believes that *P*, then *S* incurs the specified commitment. (Perhaps the nature of rationality consists of the totality of relevant normative principles, of which the Principle is one). If this is right, then there is a difference between the constitutive thesis and the Principle, because the thesis does, and the Principle by itself does not, take a stand on whether the Principle specifies (part of) the nature of belief, or (part of) the nature of something else (for instance, rationality). It is, of course, possible for the Principle to specify the nature of both rationality and belief. But, whether or not it does, it looks as if an alternative to the constitutive thesis – which accepts the Principle but denies the thesis – *can* be made out.

An alternative to the strong conceptual thesis is easy to see: the Principle is *not* an analytic truth. The claim that *S* believes that *P* implies that *S* incurs a commitment to believing any implication of *P* upon which she gives a verdict only if this claim is conjoined with the Principle. The Principle is necessarily true; but it does not specify the nature of the concept of belief. It is not even part of the nature of this concept that, if *S* believes that *P*, then *S* incurs the specified commitment.

So, does Millar’s argument establish the strong thesis, or the constitutive thesis, or both? If we were to use Millar’s argument in support of the strong thesis, it would run as follows. Short of an explanation of why the Principle holds, it is mystery why it does so; but it is not a mystery why it does so; so, there must be an explanation; and the strong thesis provides the only such explanation: the Principle holds because it is part of the nature of the concept of belief that it holds; therefore, the strong thesis is true.

The problem with this argument is that, even if it is valid, one of its premises is surely false. If the strong thesis suffices to explain why the Principle holds, it seems the constitutive thesis can do the same: the Principle is true because it is part of the nature of belief that it is true. Even if Millar’s explanatory demand can be met by the strong thesis, it looks as if there is no need for it to be met by the strong thesis, for the constitutive thesis will do just as well.

To ensure that his argument establishes all three theses, Millar needs to build a bridge from the constitutive thesis to the others. He might be able to move from the constitutive to the weak thesis if he could make good on his claim that there is no more to the nature of belief than is “specified by ... the concept” of belief *P* (p. 102). I am not entirely sure what it means for the nature of something to be “specified” by a concept, but it is not implausible to suppose it means at least the following: if *S* grasps the concept of *X*, then she believes all the truths there are to believe about the nature of *X*. If this is right, then belief in the constitutive thesis is a condition for grasping the concept of belief, and something very close to – in fact far stronger than – the weak thesis would be ensured.

However, it is not obvious that we can move from this new version of the weak thesis to the strong thesis. To ensure that we can, Millar would need to free from criticism the thought that any belief required for the grasp of a concept must have an analytic truth as its content. And that thought is not evidently true. Donald Davidson (1974) is one philosopher who claims that we can reject the very idea of an analytic truth, but nevertheless maintain that possessing appropriate beliefs is a condition for grasping (at least some) concepts. For Davidson, it is considerations of overall interpretative charity that determine which beliefs are appropriate, and not analyticity. If Davidson is right, then the beliefs required for grasping concepts need not have any analytic truths

as their contents.

But, even if Millar could free the thought from criticism, trying to do so would surely be a forlorn project, for it seems we already have in hand an explanation of why the Principle holds that does not appeal to the constitutive thesis: namely, that it holds because it is part of the nature of *rationality* that it does so.

The upshot, then, is that it seems there is another explanation of why the Principle holds, which relies on neither the constitutive, weak, nor strong, thesis. And if that is so then it looks as if, even if Millar can plug the lacuna in his argument that I have identified, he has not given us reason to prefer any of the theses over a position that denies them, and accounts for the truth of the Principle in the way I have described.

4. I now want to develop a problem for the Implication Commitment Principle itself, by examining Millar claim that “it is not possible for [us] to adopt a contemplative stance towards” (p.124) our own current beliefs.

Taking a contemplative stance towards one of our beliefs is a matter of registering that we have the belief, “without taking that fact to have any normative import” (p.111) for us, where taking the fact to have such import is a matter of thinking that the fact ensures we incur the commitment specified in the Principle. It is not obvious that we cannot take this stance towards our beliefs.

Gareth Evans (1982) is famous for discerning a certain transparency in second-order belief: our justification for believing that we believe that *P* (or, do not believe that *P*) is our justification for believing (or not believing) *P*. We consider the reasons whether or not *P* and either conclude that *P* is the case, and we believe *P*, or conclude that *P* is not the case, and we do not believe *P*. We might think that all second-order beliefs are transparent in this way. But Richard Moran, from whom Millar takes the idea of a contemplative stance, writes that if “I have some reason to believe that some attitude of mine is not “up to me”... then I cannot take the question regarding my attitude to be transparent to a corresponding question regarding what it is directed upon” (2001: 67). And it seems that Moran is right about this.

I might be a committed believer in socio-biology and, as a result, believe that, as a member of the species *homo sapiens*, I have various psychological tendencies that flow from my evolutionarily determined nature. One of these is a tendency to believe racially discriminatory thoughts. Of course, when I reflect on whether or not these thoughts are true, I see no reason to believe they are. Nevertheless, I see plenty of reason to believe that I believe they are, and as a result, I form this second-order belief. In such a case, transparency fails.

If this socio-biological story is correct, my first-order belief derives, not from reflection on whether or not its content is so, but from something over which it seems I have no control: namely, my biological nature. It is a belief I am simply landed with, and not one I formed on the basis of reasons. It is also one I am unable to give up. Or, rather, the only way I could give it up is by altering my biological nature. Perhaps it is possible for me to do that. But, even if it is, it surely cannot be rationally required that I abandon the belief. Consider an analogy. It cannot be rationally required that NN conform to basic social norms if NN no longer possess the ability to do so because she suffers from Alzheimer’s Disease, even though it may be possible to find a cure for this disease (one day), and thereby restore her ability. ‘Ought’ implies ‘can’, at least in this case, for even though there may be a sense in which even an Alzheimer’s sufferer can conform to such basic norms, they surely cannot do so in the sense that matters to the present ‘ought’. The same goes for the subject of the socio-biological story: it cannot be the case that they are rationally required to abandon the belief.<sup>5</sup> The problem is that this is precisely what the Principle requires.

<sup>5</sup> It might be objected that the Alzheimer’s disease sufferer is not relevant here, because she is not in any sense a rational subject. I think this is disputable, if we employ the conception of a rational subject that I sketch in §5 of this paper (which Millar seems to endorse). But the objection is not relevant in any case, for the point of the analogy is to remind us of a sense of ‘ability’ on which a sufferer who lacks a certain ability is on that account not subject to a certain ‘ought’, and so long as the very same sense applies to rational subjects’ inability to reject certain beliefs on

If I believe that *P*, and *P* implies *Q*, I am committed to believing *Q*, if I give a verdict upon *Q*; and – as Millar makes clear – one way of discharging this commitment is to stop believing *P*. That is what I am rationally required to do if *P* implies *Q*, and I believe *P*, but give a negative verdict on *Q*. So, if I were to believe the racially discriminatory thoughts, but to give a negative verdict on whether or not their consequences are so, then – by the Principle – I would be rationally required to abandon my belief in the thoughts. But, in the relevant sense of ‘cannot’, this is something I cannot do. And, for this reason, it cannot be rationally required that I do so. In such a case, taking a contemplative stance towards my beliefs is something I can do, and ought to do.

If this is right, then it seems that Millar has a choice. He could restrict the Principle so that it applies only to beliefs that are “up to us”; in other words, to beliefs that (in the relevant sense) we can reject. Or, he could deny that the story I have just told, and others like it, could possibly be true, and claim that all beliefs are, necessarily, “up to us” in this sense.

But, if beliefs are located in a causal nexus, how can there be an obstacle in principle to their being so located as to place them outside our control? Millar does not want to deny that beliefs have causal location; indeed, he wants to insist that a conception of belief as a standing in the space of reasons is compatible with them having such location. And yet, if he were to say that all beliefs are necessarily “up to us”, he has to claim that there is a certain kind of causal location that beliefs cannot have.

I do not know how Millar will respond to this problem. But, however he does, it seems to be a problem that he has to face.

5. In these comments, I have suggested that Millar does not have a compelling argument for his theses. I have also presented a problem for the Principle to which the theses adhere. Does this mean that we should reject a conception of belief as a standing in the space of reasons? I want to end by suggesting: perhaps not. Perhaps there is a way of understanding that conception on which it is undeniable, and avoids these objections.

Millar’s claim confines itself to the beliefs of rational subjects. There are various ways of understanding the idea of a rational subject, but one way is surely as that of subjects whose doings – intentional bodily actions, for instance, and mental actions such as thinking and intending – are, perhaps not always, but certainly sometimes, explicable by appeal to considerations that display them as things that subjects rationally ought to go in for, in the relevant circumstances (McDowell, 1986). In such cases, the considerations count as both *reasons for* subjects to do what they do, and *reasons why* they do so. And these ‘reasons for’ are, in many if not all cases, *things that the subject believes*. So, the beliefs of rational subjects must be capable of providing reasons for their doings.<sup>6</sup> That they are so capable is a truth that holds necessarily of beliefs in this class.

This description of rational subjects offers one way of unfolding the idea that belief is a standing in the space of reasons. It is also a description that anyone who understands this paper should recognise, for it is surely a description of us. That is something I cannot see how we can hope to deny. And, if we cannot, then we have reason to accept a version of the conception that Millar wishes to defend.

## References

Brandom, R. (1998), *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass.: Harvard University Press.

---

account of their biological nature, it does not matter whether the Alzheimer’s sufferer is a rational subject or not.

<sup>6</sup> They may only be able to do so when they act in tandem with other states of mind (such as desires). But there is no need for me to take a stand on this issue here.

- Davidson, D. (1974), "On the Very Idea of a Conceptual Scheme", in his *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, 2001.
- Evans, G. (1982), *The Varieties of Reference*, ed. John McDowell. Oxford: Oxford University Press.
- McDowell, J. (1985), "Functionalism and Anomalous Monism", in E. Lepore and B. McLaughlin eds. *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell.
- McDowell, J. (1994), *Mind and World*. Cambridge, Mass.: Harvard University Press.
- Millar, A. (2004), *Understanding People*. Oxford: Oxford University Press.
- Millar, A. This journal. "Précis of Understanding People".
- Moran, R. (2001), *Authority and Estrangement*. Princeton: Princeton University Press.
- Neta, R. (2008), "In Defence of Disjunctivism", in A. Haddock and F. Macpherson eds. [\*Disjunctivism: Perception, Action, Knowledge\*](#). Oxford: Oxford University Press.
- Sellars, W. (1965), "[Empiricism and the Philosophy of Mind](#)", in H. Feigl and M. Scriven eds. *Minnesota Studies in the Philosophy of Science*, vol. 1. Minneapolis: University of Minnesota Press.
- Travis, C. (Forthcoming), "Aristotle's Condition", in P. Greenough and D. Pritchard eds. *Williamson on Knowledge*. Oxford: Oxford University Press.

# Understanding People Keeping Up Standards

PAUL NOORDHOF

(University of York)

Alan Millar's book, *Understanding People*, is a defence of the role of the normative in the proper understanding of people. In broad outline, the defence has three principal components. The first is an explanation of what an appeal to the normative potentially provides that other explanations lack. Here the foil is reductive dispositionalist. The second addresses a worry about the explanatory relevance of the normative. Even if appeal to the normative promises to provide explanations which other things cannot, the promise may not be fulfilled if it has no purchase on our mental life. Millar outlines how he thinks it does have such a grip. The third addresses the question of whether we should understand the explanations offered in terms of the theory theory or simulation approach to the understanding of others. Millar's principal claim is that those who privilege the normative need not be committed to a simulation approach.

Cross cutting these three components, there are three issues that play a crucial role in the development of the picture. The first is the correct identification of the normative. Here Millar's distinctive claim is that there are a class of non-justificatory reasons - or *commitments* - that have an important role in our understanding of other people.

The second is the proper account of their nature: what is the source or basis of these non-justificatory reasons. Millar identifies three *prima facie* different sources.

In the case of the commitments we incur through having intentions and beliefs, he argues that they are derived from ideals of reason. The commitments are not identical to the ideals because there can be cases where we fail to live up to the ideals without irrationality, for instance, because doing so ceases to be in our control (Millar 2004, pp. 76-78, 120). Because these propositional attitudes serve to characterise our ideal of reason, the commitments derived from the ideal are part of the nature of the states. The commitments are reflected in the normative character of our concepts of propositional attitudes. When we self-ascribe propositional attitudes, we represent ourselves as having commitments to other propositional attitudes and, as a result, seek to meet our commitments.

In the case of commitments we incur concerning the meaning of terms, Millar's answer is that they are derived from the constitutive rules of practices or institutions to which we are committed by participating in them. In the case of the commitments we incur concerning concepts, they are, as with words, derived from the constitutive rules governing the correct uses of these concepts. However, these rules are not conceived as constitutive of practices or institutions because a particular concept could not be governed by different rules and, thus be susceptible to a different practice (Millar 2004, pp. 183-185).

The third cross-cutting issue is a defence of a 'high' conception of propositional attitudes according to which they are only properly attributed to subjects with the capacity for reflection about their commitments and what they ought to do (Millar 2004, pp. ix-x, 14).

These three issues will be the main focus of my discussion. Millar's overall approach is not to demonstrate that the correct understanding of propositional attitudes gives a substantial explanatory

role to the normative. It is rather that, call them what you will, there is a distinct way of understanding other people as a result of various facts about them that involves appeal to the normative. If we recognise this, the battle is over. The terminological dispute as to what to call them - propositional attitudes, propositional attitudes in humans, normatively committed attitudes or whatever - is of little interest.

On this last point, which I may have expressed a little more forcefully and crudely than Millar (!), I am in complete agreement. I also think that the debate is significantly advanced by the systematic and subtle way Millar has gone about discussing the issues. Certainly, the line of resistance I express below would not have been possible for me to formulate, if I had not benefited immensely from reading Millar's book. Nevertheless, however pleasant it may be to hear, agreement does not require restatement. Thus, I will emphasise the ways in which I find the defence of Millar's position wanting.

In brief, I will argue that we don't yet have clear cases of non-justifying reasons, nor a clear motivation for recognising them. I will argue that, in any event, Millar's characterisation of their nature makes them ill-fitted to play the role which he says that reductive dispositionalism cannot play and that they are supposed to, namely explain why we cannot adopt a *contemplative* attitude to certain mental states. I will argue that the high conception of propositional attitudes is not justified by the phenomena to which Millar appeals. Finally, I will argue that Millar's response to the integration challenge - that is, whether, we can have explanations with a genuinely normative explanatory force along side complete causal explanations concerning the very same entities - doesn't really get to the heart of the issue. I will also take the opportunity briefly to comment upon Millar's discussion of theory theory vs simulationist approaches to the understanding of others.

My own views on these matters will come up at certain points in the discussion though, for reasons of space, the exposition will be brief and I will rely upon references and defence elsewhere. In the conclusion, I try to sketch an alternative picture. Regarding the extent of the normative, my line will be that there is genuine normativity and then keeping to a standard. The latter should not be taken to be an instance of the former and is not an extension of its domain. It cannot play the explanatory role identified for genuine normativity and, indeed, a hybrid rather than a univocal approach to explaining the various phenomena to which Millar adverts is more likely to be successful. I hope that, by this focus, some of which raises moves which will be familiar to readers of the literature, I will draw out from Millar further justification and elaboration of his significant work in this area.

## 1. Reasons which do not supply justification

According to a picture endorsed by Donald Davidson, an agent's reasons for an action are a combination of a desire plus a means end-belief (Davidson 1974, 1980, p. 233). Millar seems happy to allow that these *may* constitute motivating or explanatory reasons although he prefers a different picture (Millar 2004, p. 10). More important for him is that they do not constitute normative reasons (Millar 2004, p. 11). Normative reasons are the sort of things people put forward as reasons: *considerations*. Thus they are not beliefs but the contents of beliefs, not desires but whatever rationally leads us to have a particular desire (Millar 2004, pp. 12-13). If reasons are simply propositions, then there is a reason to believe that ships will fall off the end of the world, namely the proposition that the world is flat. Apart from such awkwardness, to some, propositions are the wrong kind of thing to be reasons. Most of the time it is what propositions are about, rather than the propositions themselves, which support what we believe, or justify our acts (Dancy 2000, pp. 114-116). If reasons are states of affairs, then, it seems, we fail to have reasons when our beliefs are false. Assuming that my son did not beat up somebody at school, I have no reason to be angry with him even if I believed he did. This may lead us to suppose that reasons are possible states of affairs or to recognise that there are two different classes of cases which we talk of collectively in terms of an agent's reasons. I shall not settle this matter here and, indeed, Millar says no more about it.

A consequence of taking reasons to be considerations is that many reasons are not characterised in normative terms even though they may possess normative properties. The reason for believing that it is raining now is the possible state of affairs of it raining now. The rain is not normative (by anybody's lights). Millar gives, as a reason for regret, that a particular remark caused offence and, as a reason for wanting to see a movie, that it was recommended by a critic whose judgement he respects (Millar 2004, pp. 11-12). Neither of these seems particularly normative unless it turns out that respect and offence are propositional attitudes with commitments. Even so, these commitments will not concern whether regret or desire is appropriately formed in some other subject. If reasons are normative, it is in virtue of their non-normative character that they are.

Although it seems that reasons must have normative properties in virtue of their non-normative character, it is deeply puzzling how they could have such properties given the character in question. Just what is it about the rain falling down that gives it an 'oughtiness' with regard to the formation of a belief that it is raining. It is tempting to think that the situation is rather like this. The reasons themselves are not normative. However, the character of various propositional attitudes gives them this role. For instance, if the subject aims to have true beliefs, the rain falling down will, relative to this aim, determine that the subject ought to have a belief that the rain is falling down. Note that this point does not derive from thinking that the normative could not be part of the world independent of ourselves. It rather derives from an observation about what we actually count as reasons and how those things might obtain their normative properties. Those considerations for which it is most difficult to identify a type of propositional attitude with a distinctive aim are those where, if we take the reasons to be objective, we feel impelled to attribute normativity to the outside world: morality being a case in point.

There is a potential threat to the extent of the normative explanations Millar favours. In providing considerations in favour of one attitude or another, reasons are naturally thought of as justifying the attitude in question. If it seems that justification is not available for a certain kind of attitude, then there are no reasons and the attitude, and actions which follow from it, fall, at least partly, outside the ambit of normative explanation that Millar is seeking to defend. His rejection of the link between providing considerations and justification promises to extend the application of normative explanation further.

Millar also argues that it helps to deal with a dilemma. Consider the following two principles:

*The Motivation Principle:* For any agent  $x$ ,  $x$ 's motivating reasons for a belief, action, etc., are reasons that  $x$  takes (rightly or wrongly) to be adequate normative reason for that belief or action (Millar 2004, pp. 42, 68).

*The Intention Principle:* For any agent  $x$ , and action  $\Phi$ ,  $x$   $\Phi$  s intentionally if and only if  $x$   $\Phi$  s for a reason (Millar 2004, p. 69).

They seem in conflict if actions include apparently arational actions like doodling, kicking a car when it won't start, scratching out the eyes of a love rival in a photograph and so forth (Anscombe 1963, sec. 17; Hursthouse 1991). The concern is that no adequate justifying reasons are available for these actions. So, if the reasons mentioned in the Intention Principle must be adequate justifying reasons, then either the Intention Principle is false - there are intentional actions for which there is no reason - or we must claim that there are justifying reasons for these actions after all (Millar 2004, pp. 68-71). If reasons need not be justifying, then we can cite the agent feeling the urge to lash out as if the car deserves a kick or to damage the photograph respectively as reasons. The more pointless something seems - for instance, an agent's doodling - the less it seems like an action.

The success of this treatment depends upon two things. First, that our inclination to suppose that there are no reasons why we do these things rests on the assumption that whatever is cited must justify doing them. Second, there is no way of capturing the fact that action has a point without appealing to a reason. Both are questionable.

Regarding the first, the challenge of the examples is not removed by allowing that the reasons need not justify the agent. The point was rather that the things that might flash before the agent's mind resulting in the action couldn't possibly be taken as a consideration in favour of the action in any sense: can't stand the way he looks at me; teach the car a lesson etc. Strictly speaking, they cannot give a point to the action because, although there is a clear association, the action isn't to the point that they would suggest. In the photo, the rival is not looking at you. You know only too well that cars don't receive lessons by kicks. Doodling is even a clearer case since there seems no point to it other than doing whatever is going on.

Urges may causally explain the actions but the actions aren't done to express those urges. They are not the consideration why the agents acted in the way that they did. Indeed, if urges were the reasons for the actions, then the agent would appear to have a justification after all. If there is no consideration against defacing the photo or kicking car, what better ground for doing so than expressing an urge and, thereby, relieving oneself?

Regarding the second point, it is by no means obvious that all actions require a reason for them to have an aim. Certain expressions of emotion are cases in point. There is a continuum here. At one end, there are frowns and smiles, laughter and tears. Here I have in mind not those we put on but those which are elicited from us as simple natural responses. These are not actions (Goldie 2000, pp. 136-137). However, there are a range of other cases in which, although it would be incorrect to *deny* that the agent wanted to respond in the way that he or she did, attributing the desire is not to attribute something which the action was aimed to satisfy. Closest to the non-actions are stylised expressions of emotion such as jumping for joy, punching the air and so forth (Goldie 2000, p. 138). These are close to being acts in so far as we modify this type of behaviour over time and, indeed, monitor it when we engage in it, and yet still they are relatively automatic. We don't have a desire to jump up and down, or punch the air, and then satisfy it, yet the point of the action is to express a certain emotion. This is settled by the way in which the action relates to the emotion. It is not so much that we desire to express these emotions as that we lack the desire to suppress them. Philosophers should think long and hard before they commit themselves to understanding such cases in terms of a desire to express an emotion and a belief that (say), if it is joy that one is seeking to express, then jumping up and down would be a good way of doing it. It might invite ridicule.

Then there are cases like the ones described above relating to the photograph and the car kicking. Here, as I have already noted, strange desires would have to be postulated to explain these actions. It doesn't seem to be any help to suggest, as Michael Smith does, that such desires are part of the nature of the emotions we are undergoing, for instance, jealousy involves the desire to scratch out the eyes of a rival (Smith 1998, pp. 158-161). The issue is not the way in which we attribute the desires (as part of an emotion-package or individually) but simply that it seems implausible to attribute the desires to the agent in the first place outside of the hope of fitting the action into a standard model of action explanation.

Peter Goldie has suggested that behaviour such as scratching at the eyes in the photograph is to be explained by a desire to scratch out the rival's eyes plus a tendency to imagine oneself doing this (Goldie 2000, p. 129). I think it is unlikely that the agent genuinely did want to scratch the rival's eyes out, that is, *really* blind and disfigure them. Few of us are that unhinged by love. Nevertheless, even if this were the desire which explains the action, it seems that it cannot provide us with a characterisation of its point. So it is no defence of Millar's claim that the point of an action must derive from a reason. Scratching the eyes out of a love rival in a photograph is no means by which to secure that the rival is blind and disfigured.

Finally, at the other end of the continuum, there are the complex expressions of emotion in art works. Such expression can happen in various ways and I am not trying to argue that all art works must involve the expression of emotion in one or more of these ways. It is just that Collingwood describes a particularly striking way in which emotions may be expressed which is a nice illustration of the point I am trying to make:

When a man is said to express emotion, what is being said about him comes to this. At first, he is conscious of having an emotion, but not conscious of what this emotion is. All he is conscious of is a perturbation or excitement, which he feels going on within him, but of whose nature he is ignorant. While in this state, all he can say about his emotion is: "I feel ... I don't know what I feel." From this helpless and oppressed condition he extricates himself by doing something which we call expressing himself ... the emotion expressed is an emotion of whose nature the person who feels it is no longer unconscious (Collingwood 1938, p. 109).

Emotion guides the expression without it being clear to the agent what is being expressed or how it will be expressed. Yet the guidance is not to be thought of as an automatic response to a stimulus. It is something over which the agent seems to have control. It is active rather than passive. Again, although it is not wrong to say that the agent desired to express the emotion, this is because the agent did express the emotion and failed to have the desire to suppress it. Moreover, the desire in question would not be determinate because, by the nature of the case, the agent did not know what the emotion was.

My conclusion is that, while actions have goals, these may not always be derived from reasons the agent had. An agent's reasons are just one - perhaps particularly important - way in which an action may have a goal. Other elements of our mental life may supply goals too if the appropriate mechanisms of feedback and guidance are in place. In his discussion of self-deception, Millar acknowledges that there may be sub-intentional purposive formation of beliefs and that these are properly designated *activities* (Millar 2004, p. 157; Johnston 1988). This invites the question of whether there may be intentional activities, which derive their purpose from the same kind of factors these sub-intentional activities do, and yet which are intentional in that they are controlled to a greater extent by the agent. It is hard to see why not. To deny this possibility seems to rest upon a tendency to take propositional attitude psychology to apply *en bloc*, manifested, for example, in the insistence that, if something is intentional, there must be a reason which supports the intention. I question whether this must be so. It may be appropriate to allow for the application of certain elements - for instance, the fact that something is intentional - without bringing with it the whole package. If that's right, then there is no need to introduce the notion of non-justificatory reason to avoid the dilemma expressed by the two principles with which I began this section. The cases cited provide reason to suppose that the second principle is false. That does not mean that Millar is wrong to recognise instances of non-justificatory reasons. Just that the general motivation for recognising them identified above is not effective. In his book, in contrast to his earlier article, he places more emphasis on intuitive and theoretical considerations which arise from reflection upon a range of cases (Millar 2002, Millar, personal communication).

The first type of case involves action where, it seems, there is no objective matter concerning the right thing to do. One illustration is the thorny question of what to do with one's leisure time. Millar imagines a trip to Madrid in which he considers visiting the Prado. There are considerations in favour of visiting the Prado but these don't establish that it is the right thing to do over every other leisure activity. Suppose that this is the case. It is not clear why the considerations fail to count as non-conclusive justifications in favour of one course of action or another. Justification does not require commitment to there being a uniquely right thing to do.

There can fail to be a uniquely right thing to do in, at least, two ways. According to the first way, there is no determinacy in this area. There are one or things which one might do, none of which are *the* right thing to do, even from a certain perspective. According to the second way, there are one or more right things to do, from one perspective one would be right, from another, another would be right. The latter would be a kind of relativism. A third option would be to take judgements about the uniquely right thing to do as expressive of sentiments, as the expressivist recommends. I set aside this option because Millar does so too and I am interested in the options which work within his assumptions (Millar 2004, pp. 34-36).

If there is a link between truth and justification, then the second way in which there may be no uniquely right thing to do presents no problem. The truth-justification link would just be a link to a relativistic truth (if this is coherent). The first way is more challenging. Insistence on the justification-truth link would have the consequence that the considerations in favour of visiting the Prado couldn't be counted as justifications. However, the insistence is in need of defence. It is one thing to say that, where there's a truth to be had, justification must be linked to truth. It is another to insist that justification must be linked to truth when there isn't.

Instead, we might take justification to be linked to the avoidance of falsehood or, more generally, less good options. It seems undeniable that, while there is no right answer about what one should do in one's leisure time, there are several options which are not right. Staying in the hotel room with the curtains drawn watching repeats of lottery draws from the UK National Lottery while in Madrid would be one example. The considerations in favour of going to the Prado or searching out tapas bars would be considerations which speak against this kind of error about what to do with one's leisure time.

The second type of case Millar discusses to support his position concerns means-end reasoning. Consider the conditional: If I intend to buy a newspaper, then I *ought* to go to the newsagent.

Millar correctly notes that the 'ought' is non-detachable. It is not true that I ought to go to the newsagent. Morality or prudence makes no such demands upon me. Rather, it is simply that *if* I intend to buy a news paper, I ought to. The conclusion Millar draws from this is that the 'ought' which characterises the consequent of the conditional should not be taken to indicate the justification of a particular action: going to the newsagent (Millar 2004, p. 59).

This type of case does not show that normative reasons are independent of justification because here we don't have a distinct type of normative reason at all. Rather we have a justificatory ought which governs the conditional: it has 'wide scope'. It says that the following combination of mental attitude and (in)action is not being as we ought to be: intending to buy a newspaper, not going to the newsagent. The proper representation is not  $Ip \rightarrow Rq$  (where 'R' represents non-justificatory reason) but rather  $O(IP \rightarrow q)$  ('O' for the justificatory ought). In John Broome's terms (from whom I take this move), there is a rational requirement rather than a case of non-justificatory reason conditional upon the antecedent being realised (Broome 1999).

Millar is aware of this alternative account of the matter and notes that he recognises a type of reason that Broome does not (e.g. Millar and Bermudez 2002, p. 6; Millar (2002), p. 123 footnote 15; Millar 2004, p. 79). However, he seems neither to offer an argument against Broome's scepticism about these kind of reasons nor offer an argument in their favour. On the sceptical side, it seems implausible that one can generate reasons simply by having intentions. Suppose I have no reason to slap myself on the thigh but intend to do so. Do reasons come so easily that I now have a reason to slap my thigh? It doesn't seem so (Broome 2001, pp. 98-99). Suppose, to return to the original example, I ought not to intend to get a newspaper. There are far more important things to do. Do I have the slightest reason to get a newspaper because I intend to? Is there now something to be said in favour of it even if it does not justify doing it? It is plausible that the answer is no (Broome 1999, pp. 406-410).

I presume that Millar's response is that, if we take his notion of 'reason' simply to be that of commitment, then it is plausible that I am committed to slapping my thigh or getting a newspaper, given I intended to (hereafter, I shall talk of commitment-based reasons). As Millar recognises, though, Broome's rational requirements can capture the idea of commitment while rejecting the existence of commitment-based reasons. So appeal to commitment does not seem to do the job of justifying his kind of reasons. Moreover, as we shall see, accounts of the bases of these commitments make it even less likely that they will always be reason-generating.

Recent work by Niko Kolodny may appear to assist the development of Millar's position. Kolodny has argued that 'narrow scope' reasons - those where 'O' or 'R' just governs the consequent of the conditional - are required to capture the fact that certain ways avoiding of

irrationality allowed by the wide scope reading are not open to an agent. Commitment-based reasons are a subcategory of narrow scope reasons. The relationship between intention and action is not a particularly good illustration of Kolodny's point. The following candidate requirement of reason is a better one.

(WS) *O*(If you believe that you ought to do *A*, then you intend to do *A*) (Broome 2007, p. 360).

The requirement would seem to suggest that, if you believe that you ought do to *A* and currently fail to intend to do *A*, then there are two ways of being as you ought: either give up the belief or form the intention. But, Kolodny argues, it is not rational to give up a belief simply because you fail to have an intention. So, to capture this asymmetry, we need

(NS) If you believe that you ought to do *A*, then you *ought* to intend to do *A* (Kolodny 2005, pp. 527-530 - or, more clumsily, it ought to be the case that you intend to do *A*).

In fact, Kolodny's observation is ill-suited to support the attribution of commitment-based reasons. First, commitment-based reasons cannot capture the intuition that you should not revise your belief but have the intention. If you are committed to something given that such and such is the case, one option is always to change what is the case, if you can, so that you are not committed (a point on which Millar is agreed, Millar 2004, pp. 74-75). There may be considerations which rule this out but the simple presence of a commitment-based reason is not one of them. Second, and relatedly, the attribution of the relevant commitment-based reason cannot capture why the revision of the belief is not an option for you. Yet, once we have explained this, the commitment-based reason seems redundant.

The reason why it seems unacceptable to remove the belief that you ought to do *A*, rather than form the intention to do *A*, is that we think of the belief as having rational support. There is nothing particularly bad about getting rid of a belief for which there is no support given that we fail to have the intention (put aside for the moment the point about whether we can). So we should break down reasoning into two components: requirements of reason which take the form Broome recommends and lay out the options; and reasons, which exclude some of the options to be taken. In the case at hand, then, we have it that, as far as the requirements of reason are concerned you ought either not to believe that you ought to do *A* or intend to do *A*. The reasons in favour of believing that you ought to do *A* exclude the first option. Hence reasons plus requirement imply that taking the second option is the only way to be as you ought. It is not that we have reason to intend to do *A* simply from believing that we ought to. It is that the reasons in favour of believing that one ought to do *A* rule out abandoning the belief as a way to secure rationality. If you believe that you ought to do *A*, and you have reason to have this belief, then the only way you can avoid irrationality, and be as you ought, is by having the appropriate intention. It is compatible with this that, in fact, one's reasons for believing that you ought to do *A* are misleading - one oughtn't to do *A* - and hence that there is no reason for you intending to do *A*. If the reasons for the belief are not misleading, then one may have reasons for the intention but these will be both justificatory and derived from the reasons for belief.

We may be psychologically unable to give up a belief even if we lack reasons for it. It is a simple fact that we believe it and have no reasons against it. In those circumstances, it cannot be that one of the options open to us is giving up the belief. This does not mean that *O*(If you believe that you ought to do *A*, then you intend to do *A*) is incorrect. It is simply that the first option - of giving up the belief - is ruled out for us. So, for us, the only way we can obey the rational requirement is by implementing the second disjunct: forming the intention. But that doesn't mean that we reason as follows. I ought either to give up the belief or form the intention. I cannot give up the belief. Therefore, I ought to form the intention. There is no reason for forming the intention

(given that we lack reasons for having the belief). It is rather that I fail to have the option of giving up the belief. Examples of these type of cases are apt to be contentious and often involve conflicted agents. Yet, to illustrate, consider a married woman who holds down a job and can see no reason why she ought to do 80% of the domestic tasks. Nevertheless, she may still believe that she ought to do them and feel guilty if she does not. She would not be as she ought to be if she had the belief and yet failed to have the intention to carry them out. Nevertheless, in recognising that she cannot abandon the belief and yet can find no reason to hold it, it does not follow that she has reason to have the intention.

Millar's third type of case involves actions which even the agent can see no virtue in undertaking and yet they do. Examples include smashing your squash racket into your successful opponent's face or killing your children to spite your ex-spouse. The actions have a point and yet, the thought runs, not only from the objective viewpoint but also from the agent's viewpoint they are wholly negative (Watson 1975, p. 210; Stocker 1979; Velleman 1992, pp. 117-118; Millar 2004, pp. 63-67).

Again it is questionable whether we actually have reasons in play here as opposed to the workings of rational requirement. It is not that our strongest desire provides a non-justificatory reason for a particular action. It is simply that if, say, our strongest desire is to smash our racket into an opponent's face and yet we fail to have the relevant intention, we would not be as we ought to be.

These third type of cases also rest upon an illegitimate assumption concerning how the point of action will be revealed. The thought seems to be that, if the object of our action is not conceived to be good, then the point of action cannot be the good and, hence, reasons for action cannot be seen as *justifying* through pointing out how something is conducive to the good. It helps to compare matters with belief. If I believe that  $p$  is true, it does not follow that the aim of belief is truth. If it did, then my desiring that  $p$  is true would establish that the point of desires is to be true too (Velleman 2000, pp. 247-248). If I regard the proposition I believe as true, it does not follow that the aim of belief is truth either. When I imagine that  $p$ , I imagine regarding the proposition as true (Velleman 2000, pp. 248-250). Yet our imaginings do not aim at the truth. David Velleman suggests that belief involves accepting a proposition with the aim of accepting a truth (Velleman 2000, p. 251). By analogy, doing something involves having a favourable attitude to the success conditions of the action with the aim of favouring something which is good. There is no requirement that we should *conceive* of these conditions as concerning something good. Thus when we do what we conceive to be bad, it does not follow that the point of action fails to be the good.

Velleman argues that, if aiming at the good were the aim of desire, then we must conceive of that which we desire as good or desirable. It is a necessary condition for having this as its aim (Velleman (1992), p. 118). If the point of actions partly derives from the desires behind them, then it follows that, if the point of action is the good, it should be so conceived. The points made above regarding action taking this connection to be a sufficient condition rather than a necessary one.

However, it is not clear that Velleman is right in making conceiving something to be good a necessary condition. Although the state of those who desire what they conceive to be bad is puzzling, it does not seem to exclude the possibility that their desires are still aimed at the good. Suppose, for instance, that being aimed at the good was the biological function of desires. This would remain so even if there were occasions in which subjects conceived that what they desired was bad. Part of the mechanism by which desires tracked the good would be the subjects' assessment of what was desired. One way in which this mechanism could malfunction is by judging to be good what was, in fact, bad. Another way in which this mechanism could malfunction is by failing to have the proper impact upon what one desires either by being ineffective or by being conversely effective: what we conceive to be bad being hitched up to what we desire. So long as these latter cases are plausibly viewed as cases of malfunctioning, there is no threat to the thesis that desires are aimed at the good. The point is a general one and can apply to any account of the

constitutive aim of desires, beliefs or, the case in point, action. Note how unnatural it would be to say that the point of the act, in smashing a racket into your opponent's face, or killing your children to spite your spouse, is to do something bad. At best, the plausibility of such cases derives from whether they undermine the thesis that the point of desires is to reveal the good. If what I have argued here is correct, though, the cases don't.

The analogy with belief is two edged. I used it to suggest that the way we regard a proposition may come apart from the aim of the state which has the proposition as its object. Actions may be aimed at the good without us regarding what we do as good. Velleman resists this position partly because the corresponding possibility for belief seems to make no sense. How can we believe that  $p$  while regarding  $p$  to be false? (Velleman 1992, p. 118). I suggest that cases of self-deception sometimes have exactly this feature but that they are rare because of an important connection between attentive consciousness and the formation of belief which I will touch on in the next section and which I have defended elsewhere (Noordhof 2001). Another consideration stems from beliefs or acts where the subject, either through immaturity or lack of reflectiveness, does not regard what is believed or done in either of the ways said to be distinctive: as true or as good. These beliefs and actions will be classified as such because of other substantial similarities with the reflectively appreciated beliefs and actions, for example, causal profile. These similarities will also justify extension to cases in which the conception of the objects is opposite to that which is expected in the proper functioning cases e.g. as false or as bad.

The interpretation of cases of the third type is also delicate. Rather than suppose that the agent sees nothing good in the acts it seems plausible to suppose that what they do see as good - the injury of an opponent, the suffering of somebody who they think merits it - is far outweighed by all the things which are bad about the acts. Part of the anguish which is characteristic of agents' view of these acts stems from the fact they are drawn to them, they see something good in them which, because they also have a reasonably clear perception of how bad they are, becomes itself a source of reproach. If there is something good in the acts - though completely outweighed by all that is bad - then there is no problem with supposing that there are reasons derived from the good even though these reasons are outweighed by all that is bad.

I conclude that we have yet to find a case for the existence of commitment-based reasons. So far, on the one side we have justificatory, but often non-normative, reasons and, on the other, we have requirements of reason. The latter specify the combinations of states in which we are as we ought to be, and that is, as people who have preserved whatever reasons exist for further reasoning. The former may include, as a sub-class, justificatory reasons which, due to malfunctioning, in fact, reveal what is conducive to the bad.

Another case for which Millar argues that there are non-justificatory reasons relates to the use of terms and possession of concepts. Millar claims that, although their meaning or content gives us reasons to use them in particular way, these reasons should not be counted as justificatory. Again, I have my doubts about whether we have reasons here at all. In contrast to the previous cases, this is much clearer when we consider Millar's account of the source of these reasons. So rather than consider the case further here, I will go on to discuss Millar's account of the basis of non-justificatory reasons. My conclusion will be that reasons derived from the meaning of terms or the content of concepts do not constitute a homogenous group. There will be justificatory reasons and Millar-style reasons whose character I question.

## **2. The Source of Non-justificatory Reasons**

I said at the outset of my discussion that Millar has three *prima facie* distinct accounts of the source of commitment-based reasons. In the case of propositional attitudes, the idea is that they have constitutive principles governing their role. These constitutive principles derive from the nature of the states. The principles in question are derived from ideals of reason characterised in terms of the role of the various states in question. In the case of meaning, our commitments stem from

participation in a practice. In the case of concepts, the constitutive principles are derived from the nature of the concepts (so, in this sense, they are like propositional attitudes) but not in turn from ideals of reason (so, in this sense, they are like meanings).

However, behind this heterogeneity, there is more homogeneity than may first appear. In all three cases, there is something which is understood in terms of constitutive principles: propositional attitudes, concepts, practices. In each case, we grasp this through the concepts we have of them. Millar takes the constitutive principles to be part of our concepts of these things and, hence, identifiable *a priori*. Self-attribution is the vehicle by means of which the commitments are recognised (Millar 2004, pp. 118-125). We attribute to ourselves a certain intention (for example) and, thereby, are committed to act upon it, and so on. If that's right, then the explanation Millar offers of our commitments, in general, is not substantially different in character from the explanation which takes some of our commitments to derive from participation in a practice. Variation in the account derives from whether it is practices, concepts or mental states that are attributed the constitutive principles. The fundamental mechanism thereafter is broadly the same. Something Millar explicitly recognises when he indicates that his account of the basis of normative commitments in general may be understood as on a par with the commitments which are based in the institution of promise keeping (Millar 2004, p. 119).

In this section, I shall first discuss Millar's argument for the practice-based account of our commitments regarding the meaning of words. This helps to introduce my concerns with the whole approach. I'll then go on to consider whether this type of account in general is adequate for the explanatory role of commitments Millar favours.

Millar distinguishes between conditions of use and those of application. The conditions of application of a term are those conditions in which, if the term were applied to the object, the resulting sentence would be true. By 'applied' the idea is roughly that of predicating the term of some object (e.g. concatenating it with a referring term for the object to form a simple sentence) (Millar 2004, p. 161). The conditions of use are broader. They can involve how a term is correctly used in interrogative and other non-indicative sentences. The difference also arises where a term is correctly applied but based upon a misapprehension of meaning. Millar's example is of a ritual which is both ancient and arcane being described by a speaker as 'arcane' on the mistaken understanding that the term 'arcane' means ancient rather than known only to the initiated (Millar 2004, pp. 162-163).

With this understanding of the circumstances in which a term is correctly used - rather than applied - Millar points out that standard accounts of the origin of the normativity of meaning don't work. For instance, Paul Horwich's claim that the meaning of words determines how we ought to use them because it is desirable to believe truths so our actions succeed, and we will not believe truths if we use them in nonstandard ways, fails (see Horwich 2005, pp. 81-82, 116-117). There are rules governing what a term means, without being connected with truth, and we can arrive at truths without following the rules. The account has no way of explaining why we ought (or are committed) to using the term according to these rules (Millar 2004, pp. 176-177).

Millar's cases of truths stemming from misuses are accidental, for instance, not all ancient rituals need be arcane. So we might try to base the normativity of meaning upon the importance of *reliably* believing the truth. Two issues arise in this context. The first is the strength of the association between the conditions of application of two distinct terms. For instance, if there is a nomologically necessary or metaphysically necessary relationship between the two, then it is not simply accidental that the rule for one predicate enables subjects to arrive at a truth regarding the application of the other. The second is the proper specification of reliability, for example, does it require *sensitivity* to the truth of proposition believed or *safety*, that is, that the belief could not easily have been false (Williamson 2000, pp. 127-130, 147-163). Obviously these two issues may interact. The tighter the connection between the conditions of application of two distinct terms, the less likely that misuse will be unreliable in whatever sense.

Suppose we allow that reliability comes in different strengths. Across one dimension, for instance, sensitivity is stronger than safety. It may be that a belief could not easily have been false and yet, if the proposition believed had been false, we would still have had the belief. Similarly, care over the rules which govern the use of a term will, in general, make us more reliable, and less reliant upon whether their conditions of application happen to be tied closely together even if, in particular cases, care does not matter. A natural way of putting these two points together is to suggest that the rules of application of terms can be understood as part of the characterisation of *an* ideal of reason - something to which Millar appeals in characterisation of the normativity associated with propositional attitudes (Millar 2004, pp. 72-78). Because sets of terms used to classify the world may vary, there will be different determinate ideals corresponding to different sets.

Not all propositions have potential pragmatic consequences. For those which don't, the appeal to the desirability of successful action won't work. As a result, Horwich suggests that believing truths is non-instrumentally valuable. He characterises the value as a moral one (Horwich 2005, p. 118; Horwich 2006, pp. 351-352). Developing the point within the framework above, the thought would be that we value the ability to be reliably and non-reliantly moral. On the assumption that moral requirements don't apply to all rational subjects but only for those which are morally sensitive, the ideal of reason identified earlier becomes an ideal of morality and reason.

Even this retreat from instrumental or pragmatic value cannot deal with Millar's other point: the existence of proper uses which are not applications (Millar 2004, pp. 176-177). Nor, Millar argues, can our commitments be derived from intentions to use terms with their received meanings. Young children fail to have such intentions yet, according to Millar, they still have the relevant commitments (Millar 2004, pp. 172-173). Equally, according to Millar, we should not seek to derive the normativity of meaning from a desire for successful communication, as a result of which you keep meaning fixed. He suggests that it is important to distinguish our commitment to participating within a practice with the commitments following from this participation (Millar 2004, pp. 176-177).

These observations have force if a univocal account of the normativity of meaning is to be expected. An option that Millar does not consider is that different attitudes and values are the basis for different elements of our commitments or obligations. This is a point to which I shall return at the end of the section. I turn first to Millar's own approach because it promises a univocal account. If it is successful, then that would be a significant consideration in its favour.

Millar's preferred alternative is to view our use of terms as participation in a practice. As a result of participating, we are committed to the terms being used in a certain way (Millar 2004, pp. 166-175). Millar explicitly makes the comparison with Searle's familiar work on speech acts and the practice of promise-keeping in particular (Millar (2004, p. 119). Searle argued that he could derive an 'ought' from an 'is' because, in brief, when subjects utter the words 'I promise *X*' in certain specific circumstances and, thereby, participate in the promise-keeping practice, they place themselves under an obligation and, hence, ought, in certain circumstances, keep the promise. The important point is that the specification of the circumstances in which uttering 'I promise *X*' is placing oneself under an obligation is entirely non-evaluative regarding promise-keeping. It is simply a matter of linguistic facts, facts about the practice of promising-keeping, the speaker's sincerity and so on (Searle 1969, pp. 177-179). Some of these facts may be more generally normative but they do not concern whether or not we ought to keep our promises.

The standard response to Searle's approach is to distinguish two senses of 'ought': the internal and the external. Internal to the practice of promise-keeping - that is, what the practice says we ought to do given that we are participants in it - it follows that we ought to keep our promises. However, it is always possible to adopt an external view to the practice of promise-keeping and ask whether we really ought to do as we ought in the internal sense (Hare 1964, pp. 140-144; Mackie 1977, pp. 66-73). There are, at least, two ways of understanding the internal sense. One is participation-relative. Somebody participating in the practice of promise-keeping simply ought to

keep their promises (or, better, ‘ought’ unless there are outweighing considerations allowed by the practice). The second way of understanding the internal sense is that it simply records the principles laid down by the practice for its participants, what the practice says about how you should act when you have promised. Call this a practice-relative sense. Respondents to Searle’s position have no problem allowing that he can characterise our situation regarding promise-keeping in terms of the second way of understanding the internal sense of ought. They deny that anything stronger has been established.

The same point applies to Millar’s favoured formulation regarding practices concerning the meaning of words although, initially, the formulation of his position may seem to avoid the difficulty in two related respects. First, Millar doesn’t speak of ‘ought’ but rather ‘commitment’ which is explicitly distinguished from ‘ought’ by being non-justificatory. Second, he characterises the situation as:

*O* (either withdraw from the practice *OR* be committed to the rules of the practice) (Millar 2004, pp. 168-169).

For instance, regarding the practice of promise-keeping, if this holds and one doesn’t withdraw from the practice, then one is committed to keeping one’s promise. In a personal communication commenting on a previous draft of this review, Millar claims that the second disjunct should read simply ‘follow the rules of the practice’. This doesn’t seem to reflect the fact that Millar, in contrast to Broome, holds that practices supply non-justificatory reasons for doing something and do not simply rule that certain combinations of states are not being as we ought to be. Broome, after all, denies that there are reasons (justificatory or otherwise) simply generated by doing or failing to do something.

The distinction between internal/external senses of ‘ought’ goes across to commitments. Talk about ‘commitment’ rather than ‘ought’ is primarily meant to leave open the possibility that one might withdraw from the practice and hence one is not obliged to follow the standards laid down by the practice. The question still remains whether, if you are a participant in the practice, the practice-relative commitments become, for *you*, participation-relative commitments: commitments you have by participating in the practice. Or, can you be a participant and yet have no (normative) commitment to the rules?

I don’t see how Millar has established that we should be stuck with taking practices so seriously. For all Millar has argued, what is wrong with playing along, and hence participating in the practice, and yet eschewing the practice-relative commitments as not applying to you. Not all practices are good and yet good can be brought about by playing along until the crucial moment. Suppose that in some dystopian future only those who swear allegiance to, and learn the arts of, human huntsman obtain access to the killing grounds. Each human who survives a hunt in those grounds is freed from being hunted for ever. The good huntsman swore allegiance and learnt the arts in order, by sabotaging hunts, to increase the chance of the hunted humans surviving. I don’t see that the good huntsman was committed to hunting unless he or she withdrew from the hunt. Obviously, this is a deliberately extreme case but it suggests that we need a lot more than participation to transform the alloy of practice-relative commitments into the gold of participation-relative commitments for you.

Millar appreciates that participation-relative commitments may be overridden or undermined (Millar 2004, p. 86). However, that is not sufficient to answer the charge that his position is committed to taking practices too seriously. The case of the human huntsman is not one where, though the huntsman has commitments, other things count for more. Nor is it a case where the particular nature of some his or her commitments may be undermined by the considerable cost in human suffering. Instead, the nature of the practice fails to generate participation-commitments at all.

The question of whether or not the practice-relative sense of commitment - as just recording the rules of a practice - is a genuine sense of commitment is of little interest. More important is whether

this sense of commitment captures what Millar needs. I see little problem for Millar regarding the meaning of terms. The rules which govern their use characterise a way in which a speaker can go wrong which we can describe as failing in one's commitments in the use of the term (if we want to). This is compatible with allowing that there might be a very good reason for failing in this way and, indeed, rejecting any kind of attribution of participation-relative commitments generated from these practice-relative commitments.

Nor is there a problem when we turn from practice-relative commitments to what we might term *concept-relative* commitments, those recorded in the constitutive rules grounded in the nature of concepts. In the case of concepts, apparently radical failure to keep to the rules, so that there is not even a partial specification of the property the concept purportedly concerns, makes attribution of the concept incorrect (Millar 2004, pp. 182-183). Outside of that, there can be departures from the correct rules of application distinctive of a concept without that concept failing correctly to be attributable (Millar 2004, pp. 184-187). We can set aside the question of the circumstances under which it is appropriate to attribute a concept. A variety of different opinions are compatible with the general thrust of Millar's approach. Bracketing this, practice-relative commitments are sufficient to make it appropriate to describe a subject's use of a concept as either correct or incorrect in terms of the rules without having to ratchet up the accusation to saying that they have not behaved as they ought.

The difficulty arises when we focus on Millar's argument against dispositionalism and in favour of a normative characterisation of propositional attitudes. As I noted at the beginning of this section, Millar supposes that beliefs, desires and other propositional attitudes have constitutive principles governing their role in our mental lives. These constitutive principles concern the aims of these states. For instance, beliefs, guesses and conjectures all have truth as their constitutive aim, though, Millar argues, it governs their roles in distinctive ways (Millar 2004, pp. 23, 46-47). These constitutive aims reflect conceptual truths about the states in question and our recognition of our commitments with respect to these aims are ensured by, and grounded in, our reflective self-attribution of the states in question.

For the reasons I have indicated previously, I shall not consider in any depth whether it is correct to suppose that propositional attitudes should be partly understood in terms of constitutive aims. The issue is rather the significance of this. Divide the causal role of these states into two components: their downstream role, characterising the kind of states they bring about, and their upstream role, characterising the kind of things - other mental states or items in the world - required to cause them in the first place. If the states have a certain constitutive aim, and perhaps a distinctive way in which the aim governs their presence, then it is plausible that this will show up as a distinctive causal profile of the state. Some have argued that we cannot understand belief in terms of its downstream causal profile alone because we would be unable to distinguish between belief and imagination (e.g. Velleman 2000, pp. 255-263). I do not think this is true (Noordhof 2001, p. 253). However, we need no such claim to defend the interest of individuating mental states in terms of a causal profile which includes both downstream and upstream causal roles. The important question is whether there are any grounds for supposing that a normative characterisation of our mental lives is essential and whether, by recognising this, we can understand why it is of interest to individuate mental states in terms of both their downstream and upstream causal roles. It is here that I have my doubts about Millar's approach.

His approach rests upon the observation that we cannot adopt a purely *contemplative* view to the causal role of our propositional attitudes. He illustrates this as follows:

Suppose I am prone to irritability... I could adopt a purely contemplative stance in which I simply register that I am easily irritated. I could be indifferent or just curious about how I will react in situations I am about to encounter. In any case, the matter would be of merely theoretical interest (Millar 2004, pp. 111-112).

By contrast, Millar urges, if we currently intend to do *A* but do nothing to ensure the implementation of *A*, we don't really have an intention (Millar 2004, pp. 112-113). Similarly if we believe that something is the case but fail to be moved by its implications being false, we don't really have the belief (Millar 2004, p. 117). Dispositionalism about any mental state allows for the possibility of adopting a purely contemplative view to the state. Hence the impossibility of a purely contemplative view to beliefs and intentions reveals the inadequacy of dispositionalism. Here Millar is appealing to Richard Moran's point that our own propositional attitudes, such as beliefs and intentions, concern what we endorse, what we are engaged to do, and bring with them commitments to do it: they have roles in a first person essentially practical perspective (Moran 2001, p. 79).

The problem is this is not what we would predict given Millar's account of the source of the commitments relating to our propositional attitudes. Suppose I believe that I promised to do *A*. Then, given the distinction between practice-relative commitments and the rest, I can still reflect on whether I ought to do *A*. I could be indifferent to doing *A* because I don't accept the obligations of the practice of promise-keeping or I might be just curious to see whether the practice of promise-keeping manages to make me feel guilty about whether or not to keep my promise even though I don't recognise the obligation or I might carry out its requirements experimentally to see whether good or bad things follow. Of course, I can't quite see the outcome as of theoretical interest alone because I am considering what to do. Herein lies a potential disanalogy with being irritable. Irritability does not just govern what we do but also behavioural responses that are less plausibly characterised as actions. The disanalogy does not establish the connection between practice-relative commitments and the inability to adopt the contemplative stance though. It can simply be explained in terms of the way in which making a promise is more practically salient. Moreover, recognising that we are irritable, can be highly practically salient too. We can reflect upon whether our beliefs and intentions are steered by this disposition and take steps against it.

I don't mean to suggest that, when we self-attribute intentions or beliefs, we can adopt the external perspective to them and dismiss whatever commitments we seem to associate with them. It is very plausible that, when we have an intention, we strive to implement what is intended and, when we have a belief, we draw the appropriate consequences. These observations point to the character of the complex dispositional properties attributions of each pick out, in this respect different from that of irritability. If detachment is not an option, then this is a feature the characterisation of the causal role must respect. My point is that, if the normative character of our beliefs and intentions involves commitments which are simply the result of constitutive aims and principles grounded in the nature of these attitudes - just as the practice of promise-keeping has certain constitutive rules concerning the practice-relative commitments of a promise - then we should expect that it is possible to adopt the external perspective to our intentions and beliefs as well. I should also make clear that I'm not claiming that our attitude to making promises is as commitment-free as my remarks might seem to suggest. We are not indifferent to the promises we make. The point is rather that, if our commitments were routed in practices in the way that Searle and Millar suggest, that would be our situation. The fact that it is not suggests that they have identified the wrong source of our commitments or, at best, given an inadequately differentiated account of its source.

As I have already noted, Millar holds that we recognise the commitments involved in being in certain mental states through attributing them to ourselves. Thus, ascribing to oneself the belief that *p* is a way of endorsing *p* and being committed to what follows from this. However, it is the endorsement element of the story which is doing the work. Successful self-attribution may require that we are disposed to act upon the state attributed, either as the basis of our recognition of what state we are in, or as a requirement for its successful representation. In neither case, does it follow that there are distinctive commitments attached to the states attributed or, more to the point, if there are commitments, that they are the basis for our inability to take a contemplative attitude to these states. A state's constitutive principles do not make the contemplative stance towards that state

impossible unless these principles are endorsed and, of course, if a subject has endorsed the principles, then contemplation is over.

Millar argues that reference to commitments is essential in the specification of all of this because the dispositions characteristic of a belief are dispositions to discharge our commitments (Millar 2004, p. 128). I didn't see the force of this line of thought. We may need to grasp the commitments of a belief in order to be able to specify exactly which mental states should be succeeded by other mental states but it is presumably possible to describe the relations of succession, and identify the states in question, in non-normative ways. Of course, *if* such states are essentially normative, then this will not be possible. But I take it that we are supposed to be considering whether we have reason to believe this and not assuming that it is so at the outset. The non-normatively specified causal role may well be conceived of as the role which spells out the commitments of the belief but the role itself may have a fully adequate non-normative characterisation (as far as I can see).

Millar recognises that commitments can be flouted. We may participate in a practice and yet flout the rules (e.g. see Millar on soccer, Millar (2004), pp. 168-169). Thus, the inability to adopt the contemplative stance to certain attitudes is compatible with flouting their commitments.

Endorsing a particular mental state may be characterised in terms of ruling out flouting (if there are commitments to be flouted). Although Millar allows that flouting is possible and, indeed, in the case of soccer pretty easy, there are other occasions where he seems to take the existence of commitments to make flouting much more difficult. His discussion of a certain type of self-deception is a case in point.

A familiar example is that of a father who believes that his son is a fine painter because of the pleasure it brings him to believe this of his son, while, deep down, knowing that his son's artwork is poor (Millar 2004, pp. 151-152). Millar rejects this attribution of belief to the father in favour of either simply thinking that his son is a fine painter, or avowing that his son is a fine painter, or conjecturing his son is a fine painter, or being prepared to act as if he believes that his son is a fine painter. The strategy founders on its inability to capture the precise nature of the conflict and complacency that the father feels. You don't fill with pride, feel relaxed about your son's future career, feel unembarrassed about his work in most circumstances, and so forth, if you don't believe that he is a fine painter but just entertain the thought he is or conjecture that he is or act as if you believe he is. For all the rest of these attitudes, it is very much open whether or not the son is a fine painter and one gets little satisfaction from them. On the other hand, if the father is significantly challenged, certain matters are drawn to his attention, and so forth, you can be surprised about how quickly the points strike home and provoke a negative defensive reaction. You can be puzzled how the negative case seems, in a way, already familiar or how, in a defensive reaction, a further feature of the negative case is presupposed. There are going to be no open and shut examples of this kind but the attribution of two conflicting beliefs does make sense of all of this and differentiates these type of cases from ones involving the non-doxastic attitudes.

A constitutive principle of beliefs is supposed to be that they are aimed at the truth. It seems to be because Millar views this principle as pretty much unfloutable that he insists that the self-deceived don't have *beliefs* which, at the same time, they deep down take to be false. The assumption relates to his concern that, if such cases of self-deception were possible, they would show subjects could have a more detached attitude to their beliefs than, he presumes, is compatible with his theory (Millar 2004, pp. 151-152). However, this seems to be a mistake. If his theory allows flouting, then it allows for a more detached attitude to the principle alleged to be constitutive of belief. If flouting is ruled out here, we need an explanation why. Even if that is forthcoming, he does not have to deny the type of case of self-deception just described. A conflicted subject who believes *p* and believes that not-*p* will be equally conflicted in terms of his or her endorsements and detachment. When one belief is centre-stage, he or she will be in the grip of endorsements concerning it and detached from the other belief he or she is said to have, and vice versa. Millar is

quite at liberty to qualify his position to accommodate this without damaging its overall structure. Of course, he might think that such a qualification is unnecessary or implausible but, it seems to me, this follows from an antecedent view about a certain way of understanding self-deception and not because the structure of his position impels him in that way in the respects mentioned.

Millar's treatment of the paradoxical character of Moorean sentences such as

I believe that  $p$  and it is not the case that  $p$

also seems to rely upon flouting being much more difficult. According to Millar, the first conjunct claims that I'm in a state with certain commitments, for instance, assenting to  $p$ . The second conjunct has me failing to fulfil that commitment (Millar 2004, p. 125). If I can flout my commitments, then it seems that I should have no particular difficulty attributing to myself the belief that  $p$  while failing to endorse  $p$ . This would just be a case in which I am a conscious flouter. Yet, the state seems far more problematic than that. It is not inconsistent but it is not simply that we don't like thinking of ourselves as rule flouters.

It is open to Millar to appeal to the endorsement story I mentioned earlier (as he has noted to me in a personal communication). According to this, I cannot represent myself as having a belief that  $p$  while failing to endorse  $p$ . Hence a necessary condition of self-attribution would not be met in the situation that the Moorean sentence describes. However, as before, the appeal to the endorsement story renders apparently otiose mention of commitments.

Perhaps the thought is that I would have no reason to self-ascribe a belief that  $p$  if I felt no commitment to assent to  $p$ . However, this seems to get things after the fact (as it were). It is because my endorsement that  $p$  determines that I have a belief that  $p$  that, from the first person perspective, my reasons for  $p$  are reasons to self-ascribe the belief that  $p$ . Moorean sentences appear paradoxical because of this direction of endorsement rather than because of the subsequent commitments I have, if I am in these states. I see no grounds for thinking that the role of these reasons cannot be captured in causal terms. Equally, when Moran observes that our resolutions about what to do remain resolutions for which we are responsible even if we believe that we are prone to backsliding, he captures a certain causal profile: how we are apt to respond to what we have resolved, the way in which beliefs about their strength cannot influence our resolutions in certain ways, how our attitudes to these resolutions can, and so on.

Tacit appeal to different difficulties in flouting obscures the explanatory inadequacy of appeal to normativity in the explanation of our mental life. One challenge, arising from the previous discussion, is this. Given that there are cases of self-deception, and the like, in which the downstream causal role of belief comes apart from the upstream causal role, why are these cases so rare? In particular, why is it so difficult consciously to produce a belief at will? Here I don't mean consciously go about manipulating oneself so one gets in a position in which one ends up with a certain belief (for example, looking in the mirror every morning and repeating 'Every day and in every way I'm getting better and better'). Rather the thought is that we cannot consciously produce our beliefs directly without this kind of manipulation. Moreover, this inability is not like the fact that many of us are unable to waggle our ears but perhaps could learn to do so and that there are creatures who can. The inability consciously to produce beliefs at will seems to be stronger than that. Millar's floutable constitutive aims seem ill-suited to be the basis for an explanation. Talk of the constitutive principles governing the formation of belief provides no answer without a background theory concerning the ease to which these principles may be flouted.

My own favoured response begins with the thought that we find it strongly attractive to be disposed to act upon what we take to be true: that is, we find it very attractive to be in states with the downstream causal role of belief only if it also has the upstream causal role. So the aim of truth, such as it is, of a particular kind of state, belief, is derived from the desires we have regarding the basis of our action. However, now we simply have a relocation of the problem. The issue becomes

the prevalence of these desires.

One possible explanation is that evolutionary selection favours creatures with such desires. If that were right, then the aim of truth is not essentially associated with belief. It is just that we find it invariably associated in all evolutionarily evolved creatures. Given an appeal to evolution, an appeal to desires may seem inessential. All that we need is that, in evolutionarily evolved creatures, states with the downstream causal role of beliefs will have the upstream causal sensitivity to whether or not they are true (for instance, by being responsive to perceptual evidence). The problem with this position is that evolutionarily selected systems may malfunction both in some circumstances within a creature and for some creatures all the time: hearts providing a good example, the malfunctioning of which may lead to an early death. From the point of evolutionary selection, what could be more important than a functioning heart? Certainly not directly unmanipulable beliefs. Indeed, if systems may malfunction, there seems no reason why the following particular kind of malfunctioning might not occur: failing to have the appropriate desires. Yet, however irrational and misguided a subject is, it still seems as if they are seeking to respond to material which is guiding the subject towards the truth: evidence of various kinds.

I have argued elsewhere that an important feature of attentive consciousness is that it gives determinative weight to the norm of truth. It makes manifest the attractiveness of being disposed to act upon what one takes to be true. As a result, we make judgements and form beliefs - that is, go into states which dispose us to act on what we take to be true. When we are not consciously attending to the reasons for or against a particular belief, it is possible to be disposed to act upon what one takes to be true (since we take all our beliefs to be true) without the beliefs in question being sensitive to whether or not they are true. Classic examples involve self-deception. That's why a failure of consciousness at a crucial point is required for self-deception (Noordhof 2001, pp. 259-264, Noordhof 2003).

Of course, the same challenge might be turned upon me. Might there not be creatures which lack attentive consciousness as a result of evolutionary malfunction? If so, don't I have a similar problem with explaining why it is just not possible to have directly manipulable beliefs? The response is that the datum which we need to explain is why I cannot *consciously* believe something at will. The existence of self-deception and related phenomena demonstrates that I *can* unconsciously do this.

In the first part of this section, I argued that much (but not all) of the normative force which is attributed to meaning can be derived from the value of reliably believing the truth. In the second part of this section, I argued that the state-relative analogue of his account of the normativity of meaning cannot play the role he sought for it. Instead, appeal was needed to the idea of endorsement and/or floutability. If Millar's aim was to establish an explanatory role for norms derived from the idea that certain things - practices, concepts, states - have constitutive principles, then the case developed in this section puts pressure on it from two angles. The first part suggests that there is another source of normative force in central cases of meaning, the second part suggests that Millar's type of account does not earn its keep when applied to propositional attitudes.

An alternative would be to accept Millar's account of practice/state/concept relative principles (call them norms if you like) but argue that such an account has little to do with the explanatory appeal of normative considerations. The explanatory work is done, rather, by the way in which these principles reveal themselves to be attractive to follow, our endorsement of them, or various other explanations as to how they might be easier or more difficult to flout. If the normativity of meaning is taken to concern the latter, then no uniform account of it is likely. The principles may state, in general, what our obligations are but their existence does not constitute the obligations. Instead, obligations derive, sometimes, from the value of reliably believing the truth, sometimes from the value of successful communication and sometimes, for example in the case of children, from the attraction of avoiding being scolded by adults.

### 3. The High Conception of Propositional Attitudes

Part of Millar's justification for the high conception of propositional attitudes is that it is from our capacity for self-attribution that we recognise the constitutive aims of our propositional attitudes and, as a result of which, they have impact. Consciousness of their commitments is essential for propositional attitudes to play the role that they do. The latter point is one source of the claim that beliefs, desires and other such states may not be attributed to certain animals, those without the capacity for conscious thought about the commitments of these states. Indeed, the phenomena to which Millar adverts can receive an even thinner description in some cases. It matters little whether we describe what is required in terms of being conscious of a state's commitments. The question is simply whether consciousness of a state is required for them to play a certain kind of role. Thus, Millar argues that we carry out our intentions only if we know we have them (Millar 2004, pp. 146-148). Our failure to act as intended can be simply because we forgot - that is, were not aware of our intention - even though we had it.

Similarly, Millar argues, desires are only indirectly explanatory of action. Desires may be a cause of beliefs that I have them, and hence causes of my action, but they will not figure as the immediate cause of the action in question (Millar 2004, p. 12). According to Millar, this is because our desires do not lead us to act blindly. We act with a view to satisfying our desires. Beliefs which figure as assumptions in our reasoning require knowing that the assumptions are ones we believe and are relying upon (Millar 2004, p. 138).

In the case of intentions, Millar claims that a further justification for the high conception derives from the problem of deviant causal chains. In order to capture the difference, Millar argues that in a nondeviant case, we do something by way of intending to carry out the intention in question and being sensitive to its commitments.

If the activity of our mental states requires being able reflect about our propositional attitudes and form beliefs about them, then it seems that the capacity for such reflection is essential to having these propositional attitudes. The high conception of these states is true.

I have three worries about this defence of the high conception of propositional attitudes. The first is that it does not seem sensitive to the distinction between being conscious that one has a certain propositional attitude and knowing or believing that one does. The latter higher order state does not have to be conscious and, if it is not, we are presumably no better off. For example, if I fail to act because I forget my intention, then action is not secured by the equally forgettable belief that I have the intention. A consequence of this point is that Millar has no particular need to endorse a non-intentional characterisation of self-deception. I don't deny that it is very plausible to suppose that there are cases of purposive but non-intentional belief-production as Mark Johnston envisages (Johnston 1988). Nevertheless, as we can know what intention we have without being conscious of it, Millar could combine the high conception of intention - if he so wished - with the possibility of intentional self-deception without threat of paradox (Millar 2004, pp. 156-157). Perhaps Millar is committed to a higher order account of consciousness according to which, beliefs about lower order states make them conscious without the beliefs themselves being conscious (see e.g. Rosenthal 1990). Nevertheless, this is not a matter he discusses.

Second, and relatedly, even if beliefs about states are required for them to play their role, on pain of regress, there will be some states that play their role in the absence of beliefs about them, for example, the higher order beliefs themselves. The question is, if this is allowed, then what's so special (or lacking) about the lower order states which makes beliefs about them required for their role to be played? Perhaps Millar's thought is that, when a higher order state enables a lower order state to play a role, the higher order state is not, itself playing a role. When it is, then an even higher order state concerning it is required in turn. If something along these lines is correct, more detail is desirable about when a state requires another state to enable it to play a role because, certainly, the higher order state plays *some* role by enabling lower order states to play theirs.

Third, and finally, it doesn't seem obvious that Millar's appeal to beliefs about the commitments

of states avoids the difficulty arising from deviant causal chains. The essential role of considerations from deviant causal chains arises in the following way. Millar accepts that agents may follow the constitutive plan of an intention they possess without conceiving of themselves as implementing the intention in the plan. For example, Davidson's climber wants to let go the rope so that the climber below (who has lost grip) fails to drag him down the mountain. He becomes so nervous that this causes his hand to jerk so letting the rope go. Nevertheless, the climber was not acting upon the intention and did not conceive of himself as so doing. Here we have a deviant causal chain which implements the planned results.

Millar makes two moves to deal with such cases. First, following Searle, Millar talks about the *intention-in-action* - described from the first person perspective - as *that I  $\Phi$  by way of carrying out this intention* (Millar 2004, pp. 145-147, Searle 1983, p. 85). Second, he appeals to the idea that, in normal, that is nondeviant, cases, we are sensitive to the commitments of our propositional attitudes.

Regarding the first move, it is not obvious that, if I have an intention with the content that I  $\Phi$  by way of carrying out this very intention, it follows that the resultant behaviour is non-deviantly caused. Searle claims that what is lacking in the deviant cases, and present in the nondeviant cases, is an experience of the movement (corresponding to the action) as a result of the intention in action. For instance, if the action is raising one's arm, the corresponding movement is one's arm raising. The action is the combination of the experience and the movement caused by it. This is peculiar. It suggests that something may be a perceiving of something else - the causing of movement - while being prior to it because it is doing the causing. It is usually thought that a necessary condition of perceiving something is that the latter is a cause of the former. Perhaps Searle supposes that the experience has the content of a causing of a movement without being a perception of the causing (Searle 1983, pp. 91-95). If that's right, then it is quite unclear why one shouldn't have a case which isn't action and yet in which one has the relevant experience. Indeed, the climber may experience his or her intention causing the action because it is having the intention that makes him or her nervous.

In considering this case, Searle says that there is no moment at which the climber might say I am now letting go of the rope as a result of the intention (Searle 1983, p. 108). However, that is not the correct test. The test is whether the agent is experiencing the rope being let go as a result of the intention. The answer to that is surely 'yes' unless 'as a result of' is intended in a non-deviant causal sense. In which case, we need a specification of what this sense is. Once we have it, there would be no need to appeal to this component of Searle's analysis because we could go straight to 'go' and use it to characterise non-deviant causation directly.

Making intentions reflexive not only seems ineffective, arguably it gets the phenomenology wrong. It is certainly correct that, when an agent intends to  $\Phi$ , he or she does not suppose that the action will be caused by other means. It is also true that, in those circumstances, he or she will not have acted if the action is caused by other means. Nevertheless, I don't see why it follows from these two things that the intention in question is reflexive. When we act we don't typically intend to produce an action which has a certain aetiology. We just aim to produce something with a certain outcome (and our beliefs and desires are geared towards this). It seems to me that we have here, at best, the illicit infiltration of the requirements of an analysis - regarding the difference between deviant and non-deviant causal chains - into the proper specification of the content of a state.

This brings me to the second component of Millar's position - indeed, the one on which he places emphasis - namely that we should be aware of the commitments of the intention. The problem here is that one may be so aware and yet have an action which is deviantly caused. Consider Davidson's climber. He thinks that the only way he will survive is by letting go the rope. He intends to do this and is aware of the commitments of his intention. As a result of this, he becomes so nervous, that he does it anyway. If you consider the original case plausible, it is hard to see why one should not consider this case plausible. Indeed, it is plausible that this way of

describing the case is central. It is no special variant. The whole reason why the climber became nervous is precisely because he is aware of the commitments of the intention he has formed.

The same goes for Searle's variant on this position. He suggests that we should appeal to whether or not we have a reliable case of causation by the intentional aspects of the relevant states (Searle 1983, pp. 135-140). But such intentions may be quite reliable in getting our climber nervous as a result of their intentional features. Still the result would not be an action because, intuitively, he is neither in control of his nervousness nor is this part of the standard causal route to action.

The message of deviant causal chain cases is not that you should identify more mental states that can capture the difference - for instance, knowledge of one's intention - or that there should be a particular content or quality that the action should have. The message is rather that, when all these are taken into account, we need to talk of a particular way in which *this lot* causes the action in question. For example, in the case of belief, Ralph Wedgwood appeals to a rational capacity or disposition to respond to the commitments of a belief to form any of a range of other beliefs, when we are considering basic steps of reasoning (as opposed to those involving further mental states as intermediate steps in reasoning) (Wedgwood 2006, p. 672). As an illuminating response to the problem of deviant causal chains, I have my doubts about this. Extending the approach to the case we have been considering, I presume that the intention to let go the rope causing the action via nervousness would fail to count as a *rational* capacity or disposition even though it may count as a mental disposition of the subject, even if this disposition were highly sensitive to a range of different cases, and even though, as we saw with regard to the other proposals, it is a response to the commitments to the intention. That suggests that the deviant-nondeviant distinction is, at least to some extent, embedded in the classification of a capacity as rational or otherwise. Nevertheless, as an articulation of exactly the kind of account we need - appealing to a certain way of causing rather than intermediary states - Wedgwood's proposal seems bang on the money.

#### **4. Explanatory Relevance of the Normative (and the Character of the Explanations in which it figures)**

In his initial characterisation of the problem of explanatory relevance, Millar writes that the issue is

If rationalizing explanations are causal, then the attitudes cited in such an explanation explain *qua* causes of the action...The challenge ... is to show how the fact that the agent's attitudes rationalize what they explain can be relevant to the explanation of what is explained (Millar 2004, p. 21).

This suggests that there are three defensive steps needed to vindicate an appeal to the explanatory relevance of the normative. First, identify a potential causal role for the commitment-inducing properties of the contents of beliefs, and other propositional attitudes, for instance, their conceptual structure or logical form. Second, explain how the commitment-inducing properties of the contents of beliefs (say), and not simply their realisations, are a cause of the transition from some beliefs to other beliefs. This would involve Millar in discussing what Jaegwon Kim has dubbed the exclusion problem (e.g. Kim 1998). If the world of physics is causally complete and mental properties are not identical to, but realised by, the properties identified by physics, then there seems to be no causal role for mental properties. Their realisers will cause each other and the mental properties will just come along for the ride. Third, argue that the causal role which the commitment-inducing properties play is one which makes their normative nature explanatorily relevant. That is, the explanatory force is genuinely normative and not simply causation by the normative. The last step is suggested by Millar's distinction, in the quoted passage, between rationalising attitudes being causes and it being the fact that they rationalize which explains the action (see also Millar 2004, p. 192). We might dub the overall challenge that these steps answer, the integration challenge. How do you integrate normative explanatory material with a causal framework?

Unfortunately, Millar's main response to the integration challenge seems to be to explain how the normative may be efficacious and, even there, as he acknowledges, the defence is incomplete. This doesn't quite measure up to the initial problem set. He begins by rejecting the idea that the role of commitment-inducing properties is simply that the commitments identify how things will go in idealised appropriate circumstances. He notes that if we have dispositions whose manifestation corresponds to, and thereby display an implicit grasp (perhaps) of, logical relations, it doesn't follow that the commitment-inducing features have an explanatory role (Millar 2004, pp. 196-197). Although this seems right, it is slightly surprising that Millar makes this point and yet accepts that the failure of such dispositions to be manifested captures what goes wrong in deviant cases (Millar 2004, p. 198). As we saw in the previous section, sculptured nervous dispositions would still seem to be deviant even if they display the envisaged structure. If dispositions may correspond to, and thereby display an implicit grasp of logical relations, without commitment-inducing features have a suitable explanatory role, then these very same dispositions cannot constitute the essential difference between deviance and non-deviance. Failure to have a suitable explanatory role will correspond to a particular class of deviant cases of causation. On the other hand, if the dispositions are only manifested if they are responsive to commitments in a non-deviant way, then likewise it seems that they will not be manifested if there is simply correspondence with a content's logical relations.

The second point he makes against the appeal to the disposition to respond to commitments is that it would not be able to account for cases in which, while the reasons failed to support the action, rational explanation is still possible. Things which seemed a good idea but weren't couldn't be the manifestation of a rational disposition (Millar 2004, pp. 203-206). Yet surely, Millar argues, we would not want to say that the commitments of propositional attitudes failed to have an explanatory role.

The general form of Millar's proposal is that, when rational explanation is operative, we take ourselves to have an adequate reason to believe that *p* (say) *as a result of* which we believe that *p* (Millar 2004, p. 204). In good reason cases, we take ourselves to have adequate reason because there is such a relationship. In bad reason cases, when we don't stand in this relation, there should be an explanation of why we take ourselves to do so (Millar 2004, p. 207). There is some unity but also some important differences in the cases regarding how the explanation works (Millar 2004, pp. 209-210).

The problem with this position is that the difficulties he has identified, if genuine, seem to apply as much to *taking oneself to have an adequate reason* as they do to the beliefs mediated by it, on the one hand, and the rejected appeal to dispositions, on the other. If all that Millar means by the italicised phrase is that the subject is disposed to respond in the appropriate way, then it is not clear how his proposal avoids the first problem he raises for dispositional accounts. If *taking oneself to have adequate reason* is a mental state on a par with beliefs and desires, then we need a story of how its commitment-inducing properties have an explanatory impact. At a certain level, it seems we are going to be appealing simply to causality. We will have causation by the normative and not an explanation with genuine normative force.

It is not obvious that the integration challenge - understood in the way that Millar seems to understand it - can be met. Once it is accepted that the normative is, and must be, efficacious in order to have explanatory force, it seems that the best we can expect is that such explanatory force rests upon a particular kind of causation by the normative. In that case, though, the substantial challenge to the explanatory force of the normative stems from those who endorse a variant of Jaegwon Kim's exclusion argument. If there is a complete causal story characterised in non-normative terms, then, on pain of systematic overdetermination, there is no role for the normative to play. Yet Millar does not discuss this argument and, he has confirmed to me in personal communication that, while he assumes that there is some answer to it, he does not seek to defend a particular approach in his book. In the light of my discussion above, this strikes me as a lacuna.

Another development of his approach to the explanatory role of the normative is revealed in his discussion of theory theory versus simulation accounts of explanations citing propositional attitudes. Millar rejects the idea that, if we appeal to reasons or commitments, we must take the resulting explanations to be those favoured by the simulationist. Here it is helpful to distinguish between two issues. First, is the material we use to explain the mental life of others something that can be represented as a theory? It is not obvious that talk of commitments is a natural part of scientific theorising. Second, is the right description of the cognitive processes that enable us to understand others properly described as theorising or simulating? This distinction is slightly obscured by Millar's talk of the *basis* for ascribing attitudes which can be thought of in either way (Millar 2004, p. 215). Millar takes simulation theorists to deny that a theory can be formulated but this is much too strong if one is a simulationist regarding the second issue. Contrary to what Millar suggests, even if a theory can be formulated, simulation would scarcely be useless. First, the theory may not be available to the subject. Second, simulation may help us deal with the complexity of the subject matter and, also, provide a cognitive saving in processing power. By the same token, we may agree with Millar that the resolution of these issues do not proclaim one way or another upon whether it is appropriate to appeal to the normative to understand and explain others.

## 5. Concluding Remarks

The starting point in Millar's work is that there are certain things which have constitutive principles concerning them: practices, states and concepts. Two additional claims are then made: first, that these constitutive principles are normative (although not necessarily justificatory) and, second, that they play a certain explanatory role. I have tried to argue that these two claims are in tension. The explanatory role identified - in particular, to explain when a contemplative attitude is not possible - cannot be played by appealing to constitutive principles alone. This is fundamentally because there will always be a further question of whether the constitutive principles are ones we should respect even if, in some sense, we have committed ourselves by the terms of the practice, by instantiating the state, or by utilising the concept. Contrary to what Millar urges, there is a further question of what people have resolved to do or whether they have endorsed the principles which is crucial to the proper understanding of how they respond (see Millar, *Precis*, p. 15). The proper understanding of this is a matter of psychology (and philosophy) and cannot be exhausted by appeal to the normative.

Even acknowledging this, there is a question over whether we should conclude that there is a genuinely stronger notion of normativity as yet uncaptured by talk of constitutive principles or whether that's the only notion we have and the rest is to be dealt with under the heading of endorsement or resolution. My suggestion is that we keep the option of supposing that there is a genuine normativity as yet uncaptured and, as a working hypothesis, at least, take the constitutive principles to be setting out a standard: a standard to which, at least in principle, we may adopt the contemplative attitude and question whether it works for us.

At the beginning of section 1, I claimed that, in many cases, reasons themselves are non-normative states of affairs. Their bearing on propositional attitudes, which I characterised in terms of their *oughtiness*, stemmed from the constitutive principles that governed the formation of these attitudes. We are tempted to ascribe oughtiness to the world where the relevant states and constitutive principles are not forthcoming. If constitutive principles simply lay down a standard, then, the oughtiness of reasons cannot stem from the constitutive principles alone. They may characterise how states of affairs bear upon the world - which states of affairs may count as reasons for which propositional attitudes - but they will not by themselves provide an explanation of why we *ought* to be in these states in such and such circumstances. The obligatory character depends upon another feature as well, that touched upon under the heading of endorsement or resolution. The constitutive principles lay down distinctive standards for various kinds of things but our endorsement of, or resolution to adopt, these standards is required to turn the alloy of principles or

regulations into the gold of obligation. At the end of section two, I noted that endorsement may rest upon a number of different factors so that there is no unitary account of all of the obligations attached to meaning, concepts and propositional attitudes and not even a unitary account for each of them.

In developing a theory of endorsement, we may hope to arrive at answers to the explanatory issues raised so eloquently in Millar's work and discussed in sections three and four. In particular, when we provide normative explanations, must the application of endorsed principles be mediated by higher order states and, if so, how? How will these endorsed principles be explanatorily relevant when we also explain things in terms of brute causality? What kind of causation must be involved? A theory of endorsement will need to explain how a subject is sensitive to the exact content of the principles and resolves to be governed by them. In spelling this out, we will need to explain what is involved in government and, presumably thereby, when it is relevant to cite the fact that we were so governed<sup>1</sup>.

## References

- Anscombe, G. E. M. (1957), *Intention*. Oxford: Basil Blackwell.
- Broome, J. (1999), "[Normative Requirements](#)" *Ratio*, XII, no. 4, pp. 398-420.
- Broome, J. (2001), "[Are intentions reasons? And how should we cope with incommensurable values?](#)" in Christopher Morris and Arthur Ripstein eds., (2001), *Practical Rationality and Preference: Essays for David Gauthier*. Cambridge: Cambridge University Press, pp. 98-120.
- Broome, J. (2007), "[Wide or Narrow Scope?](#)" *Mind*, 116, no. 462, pp. 359-370.
- Collingwood, R. G. (1938), *Principles of Art*. Oxford: Oxford University Press.
- Dancy, J. (2000), *Practical Rationality*. Oxford: Oxford University Press.
- Davidson, D. (1974, 1980), "[Psychology as Philosophy](#)", in S. C. Brown (ed.), *Philosophy of Psychology* (London, Macmillan) reprinted in his (1980), *Essays on Actions and Events*. Oxford: Oxford University Press, pp. 229-244.
- Goldie, P. (2000), *The Emotions*. Oxford: Oxford University Press.
- Hare, R. M. (1964), "The Promising Game", *Revue internationale de philosophie*, 70, and in his (1989), *Essays in Ethical Theory*. Oxford: Oxford University Press, pp. 131-144 [page references in text to latter].
- Horwich, P. (2005), *Reflections on Meaning*. Oxford: Oxford University Press.
- Horwich, P. (2006), "[The Value of Truth](#)", *Noûs*, 40, no. 2, pp. 347-360.
- Hursthouse, R. (1991), "[Arational Action](#)", *The Journal of Philosophy*, 87, no. 2, pp. 57-68.
- Johnston, M. (1988), "[Self-deception and the Nature of Mind](#)", in B. McLaughlin and A. Rorty (eds.), *Perspectives on Self-Deception*. Berkeley: University of California Press, pp. 63-91.
- Kim, J. (1998), *Mind in a Physical World*. Cambridge: Massachusetts, The MIT Press.

---

<sup>1</sup> I am very grateful to Alan Millar for his comments on the previous draft of this paper which saved me from significant errors, Patrizia Pedrini for bearing with me (her patience was sorely tested) and, finally, to the Leverhulme Trust part of whose generous support enabled me to conduct the research upon which the paper was based.

- Kolodny, N. (2005), "[Why Be Rational?](#)", *Mind*, 114, no. 455, pp. 509-563.
- Mackie, J. L. (1977), *Ethics*. Harmondsworth, Middlesex: Penguin Books Ltd.
- Millar, A. and Bermúdez, J. L. (2002), "Introduction", José Luis Bermúdez and Alan Millar (eds.), [Reason and Nature](#). Oxford: Oxford University Press, pp. 1-11.
- Millar, A. (2002), "Reasons for Action and Instrumental Rationality", José Luis Bermúdez and Alan Millar (eds.), [Reason and Nature](#). Oxford: Oxford University Press, pp. 113-132.
- Millar, A. (2004), *Understanding People*. Oxford: Oxford University Press.
- Moran, R. (2001), *Authority and Estrangement*. Princeton and Oxford: Princeton University Press.
- Noordhof, P. (2001), "Believe What You Want", *Proceedings of the Aristotelian Society*, 101, no. 3, pp. 247-265.
- Noordhof, P. (2003), "[Self-Deception, Interpretation and Consciousness](#)", *Philosophy and Phenomenological Research*, 57, no. 1, July, pp. 75-100.
- Rosenthal, D. M. (1990), "A Theory of Consciousness", *ZiF Technical Report*, Bielefeld Germany, reprinted in Ned Block, Owen Flanagan and Güven Güzeldere (eds., 1997), *The Nature of Consciousness*. Cambridge: Massachusetts, The MIT Press, pp. 729-753.
- Searle, J. (1969), *Speech Acts*. Cambridge: Cambridge University Press.
- Searle, J. (1983), *Intentionality*. Cambridge: Cambridge University Press.
- Michael Smith (1998), "The Possibility of Philosophy of Action", J. Bransen (ed.), [Human Action, Deliberation and Causation](#). Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 17-41, reprinted in his (2004), [Ethics and the A Priori](#). Cambridge: Cambridge University Press, pp. 155-177 [page references in text to latter].
- Stocker, M. (1979), "[Desiring the Bad](#)", *The Journal of Philosophy*, 76, pp. 738-753.
- Velleman, J. D. (1992), "[The Guise of the Good](#)", *Noûs*, 26, 1, pp. 3-26, reprinted in his (2000), *The Possibility of Practical Reason*. Oxford: Oxford University Press, pp. 99-122.
- Velleman, J. D. (2000), "On the aim of belief", in his (2000), [The Possibility of Practical Reason](#). Oxford: Oxford University Press, pp. 244-281.
- Watson, G. (1975), "Free Agency", *The Journal of Philosophy*, 72, no. 8, pp. 205-220, reprinted in his (2004), [Agency and Answerability](#). Oxford: Oxford University Press, pp. 13-32 [page references in text to former].
- Wedgwood, R. (2006), "[The Normative Force of Reasoning](#)", *Noûs*, 40, no. 4, pp. 660-686.
- Williamson, T. (2000), [Knowledge and its Limits](#). Oxford: Oxford University Press.

# Replies to Authors

ALAN MILLAR

I am grateful to my critics for their comments and for the work that they have put into producing them. Thinking about the criticisms, and how best to reply to them, has been a valuable exercise for me. I hope my responses will be of some use. I have focused on what I take to be the main points. Many of the issues raised merit further discussion.

## Reply to De Caro

I am encouraged by Mario De Caro's appreciative comments. In the book I developed a broadly Davidsonian line of thought on which rationality is taken to be essential to the possession of propositional attitudes. I suggested that rationality is not just a matter of internal coherence and the legitimacy of transitions in reasoning, but has to do with being in touch with reality. De Caro asks, 'how we, the interpreters, can be sure that the interpreted speaker is *actually* in touch with reality—even if this is what it seems like to us' (p. 25). Perhaps radical translation, as opposed to immersion in the language once it is learned, does not enable us to be sure. The best we could achieve in this enterprise would be plausible hypotheses that are subject to refinement or rejection in the light of further evidence. The key point is that in arriving at these hypothesis we are guided by our knowledge of the environment and by considerations about what it make sense for people to think and do, given the state of the environment and what is going on within it. People standing in a café drinking coffee, reading newspapers, chatting with each other, and interacting with those who are serving, are doing just what we would expect them to do in those surroundings. It would be difficult to explain their behaviour except on the assumption that they are in cognitive contact with their surroundings. In a calm setting such as this, in a location that is relatively safe, a person who displayed intense fear might well not be in adequate cognitive contact with the surroundings. It would depend on whether there is some reason for him or her to be fearful. Of course, we view such a person against a cultural background in which there is shared knowledge of customary modes of behaviour, the function of cafés, and so on. If we are in an environment where the culture is less familiar and we do not know the language, it will sometimes be harder to make sense of what people are doing. But even then we are not baffled when, at what is evidently a market, people are buying and selling goods, or when people quicken their pace to avoid an approaching vehicle, or adjust their movements while walking so as to avoid bumping into others.

These common sense observations do not by themselves suffice to meet philosophical scepticism about knowledge of whether people are ever in touch with reality. Philosophical sceptics, and those who think sceptical challenges are hard to meet, tend to work with theories of knowledge that make it hard to see how the requirements for acquiring some type of knowledge are met. In recent philosophical tradition, Quine stands out as a theorist who limits the basis for knowledge and fruitful theory to stimulations—activations of nerve-endings.<sup>1</sup> Aware that it is impractical to represent our theorising as if we literally take account of what is going on at our nerve-ending, he settles for the next best thing—an evidential basis that is maximally (though not completely) free of what, from his perspective, have to count as theoretical presuppositions. From this perspective the 'middle-sized' objects that we take ourselves to encounter through perception

---

1 See Quine 1960, especially chapter 1 and 2.

are theoretical posits. Applied to radical translation the upshot of this view is that our evidential basis for understanding others is limited to verbal dispositions conceived at a level that does not rely on detailed assumptions about what those we seek to understand mean by the words they use. To read into the evidential base any particular ontology would be to presume what needs to be established, but in the end cannot be established because of the inscrutability of reference. The upshot is, in effect, a kind of scepticism not just about knowledge of meaning and reference, but about whether the very idea of meaning and reference has a place in a respectable, truly scientific, theory of the world. If the view is to be challenged it must be by showing that the underlying theory of knowledge is open to question. This, of course, raises the question of what it is to which our theories of knowledge are answerable. This is not something I explored in the book. My view on the matter is in some respects fairly traditional. I hold that when doing epistemology we should seek to make sense as much as possible of what, prior to theorising, we would on reflection count as knowledge. Our data will include real or imagined cases in which it is not a matter of real dispute that people know this or that, and will also include our pre-theoretical grounds for thinking of these cases as cases of knowledge. In relation De Caro's challenge it is therefore relevant to point to what would in practice assure us that others are or are not in touch with reality: our seeing those we seek to understand respond to their environment in ways that make sense. It is also relevant to take note of how remote from our actual thinking is Quine's radical translation thought-experiment. There remains room for further discussion on the matter beyond anything I tackled in the book. In particular, we need a theory of our knowledge of others that is not confined to the framework of the traditional epistemological problem of other minds, but makes sense of the ease with which we really do know a very great deal about what others think and know, and about they are doing.<sup>2</sup>

A central theme of the book is that in understanding others we must view them as rational agents, operating, as they say, within 'the space of reasons', but I also hold that psychological explanations are causally explanatory. Much attention has been paid to a problem concerning how to make sense of causation at the level of intentional psychology. The problem, in essentials, is that if, as many suppose, physical events admit of complete causal explanation at the level of physical science, there seems to be no room for autonomous psychological explanation. We have a problem of *explanatory exclusion*. I agree with De Caro that this is an issue of importance, though I did not tackle it in the book. I give some indication below of what I take to be a promising line of thought, though clearly it requires development.<sup>3</sup>

Suppose that I receive an e-mail from a colleague asking me for an opinion on some disputed matter. I read the message. The upshot is that I understand the colleague to be asking me for an opinion on the disputed matter. There is, let us suppose, a scientific account of the chain of events leading from the brain of my colleague, via his keyboard and computer, through the university network to my computer and terminating in events in my brain. The account is couched in scientific terms and does not draw upon the categories of intentional psychology. Why should we assume that this explanation is complete? It may be complete in the sense that each change in the chain of physical events leading to the change in my brain-state as I understand the message occurs in accordance with the laws that cover such changes. But it would not be complete in the sense of fully explaining my reading and understanding my colleague's message. An account of this that did not advert to the fact that I read and understand the message, and to my capacity to do so, is not a complete account of my reading and understanding the message. So there is a sense in which the scientific explanation is incomplete. It is not incomplete in the sense of having gaps or jumps that are not in accordance with the general workings of my colleague's brain and body, of the computer network, and of my brain and body. But it is incomplete in the sense that it does not relate the

---

2 I may add that it seems to me that the epistemological wrings of John McDowell are crucial for this enterprise. See in particular, McDowell 1981 and 1982. I pursue these matters further in Millar (forthcoming 2008).

3 For some remarks on what, in the book, I called the problem of explanatory relevance, see my reply to Paul Noordhof.

events in my brain to my coming to understand what my colleague wants. So perhaps the problem of explanatory exclusion arises in part because the notion of a scientific account in non-intentional terms, that is without gaps in the sense explained, has been conflated with the notion of a complete explanatory account.

The line of thought I have just sketched is couched in terms of explanatory completeness. Some will think that it does not address the real issue, which is posed by the thought that if the scientific story deals in causal chains without gaps then, barring causal overdetermination, there simply is no room for a causal-explanatory story at the level of intentional psychology. The picture in the background here is that, since the real causation is going on at the scientific level, either there is no causation at higher levels or, if there is, it is causation in some etiolated sense. But I do not see why we should not suppose that there are layers or levels of causation, just as there are layers or levels of explanation.

Imagine that I am pushing a wheelbarrow while working in my garden in order to transport compost from one part of the garden to another. A scientist might be interested in explaining why the barrow moves in just the way that it does in terms of the forces acting upon it—forces that enable the barrow to overcome the friction exerted by its contact with the ground. Why the barrow is being moved by me would be irrelevant to such an explanation. That is because the phenomenon to be explained is a physicalistically specified motion and the explanation sought is in scientific terms. If the scientist wished to follow the explanatory chain from the motion of my arms to the contraction of my muscles and beyond then it would be necessary to account for the events that took place within my body that lead to the barrow moving. Such an account would not be relevant to an explanation of why the barrow is being moved by me. An explanation of that must be in terms of my intention to transport compost from one part of the garden to another. Here the explanandum is an intentional action of mine. It is clear here that the phenomena to be explained are different and it is not surprising that the explanations should differ. Plausibly, we have here causation at two levels: my taking steps to move the barrow in order to transport compost is a cause of the barrow's being moved by me; the causal factors picked out by the scientific account cause the barrow to move in such-and-such a physicalistically specified manner.

Against this view of the matter it might be said that there is still one event—the barrow's moving—that on my account has two sets of causes. After all, if the barrow is moved by me, then it moves and my intentions are therefore causally relevant to the occurrence of this movement. But why can't we just accept that this is so with good grace? There is the one motion of the barrow, but that motion can be considered from different points of view with different explanatory interests in mind. There remains a difference between the causal factors relevant to an explanation of the physical characteristics of the barrow's moving in terms of the physical forces exerted upon it and an explanation of the movement's being one that is initiated by me in this context.

### **Reply to Gozzano**

Simone Gozzano says that '[o]ur epistemic conduct, the way we proceed from one thought to another, may follow very different strategies' (p. 32) With this in mind he claims that '[i]t seems quite difficult, given a set of beliefs, to say in advance what is entailed by this set and what this set entails without specifying what kind of epistemic conduct we are to follow and what is the aim of our reasoning' (p. 32). Gozzano appears to work with a very broad notion of entailment. In the book I spoke of what is implied by what one believes. By 'implies' I meant what some would express by the phrase 'analytically entails'. Gozzano's notion of entailment is not apt for expressing the views I expressed by my use of 'implies'.

Gozzano pursues his theme by noting that there are different ways one might go faced with what looks like a counterexample to a generalisation one accepts. If I understand him, he thinks that merely taking account of the fact that believing something commits one to believing what this entails does not tell us what move we should make—what belief if any we should adopt or abandon.

I agree; knowing what a belief or intention commits you to doing leaves open a great deal about how you will and how you should think and act in consequence. But my aim was not to explore strategies for adjusting beliefs or acting. Because discussions of normativity have been hampered by lack of specificity about the sorts of normative consideration that govern our thinking, I focused principally on the implication commitments of believing and the means-end commitments of intending. I took it that many would agree that we incur such commitments even if they disagreed with my view of the form that these commitments take. There are issues concerning the further ramifications of having incurred a commitment, which I did not raise. That was because my principal concern was (i) with the link between beliefs and intentions and the commitments that they incur and (ii) with the importance of our reflective stance on our beliefs and intentions and on the commitments they incur. Properly to address these matters it was not necessary to discuss all the many different ways in which it is possible to react to a given commitment.

Gozzano observes that there are different ways of conceiving of entailments, for instance, in terms of classical or intuitionistic logic. He asks whether I think that ‘it does not matter which method you follow as long as, once you have taken one, you follow it thoroughly’ ( p. 33). The answer is, ‘No’ because I do not think that classical and intuitionist logic correspond to different methods that we might or might not adopt, depending on our aims. It suffices for my purposes that we are often in a position such that it is not in dispute that certain things one believes entail other things. If there are cases in which some think that P entails Q and others deny this then there will be, correspondingly, disagreement over what believing P commits one to. I do not see that this poses a fundamental problem for my view since it was never part of the view that it should always be evident to us whether a given commitment is incurred. I think a similar response should be given to Gozzano’s claim that we can never be sure that folk psychological talk is paradox free (p. 33). That may well be so, but again it does not seem to affect what matters for my purposes. So far as belief is concerned what matters is that it can sometimes be evident that some things we believe commit us to believing other specified things. If there are indeterminacies with respect to what is entailed by what we believe then, on my view, it follows that there are indeterminacies in the content of the belief and thus also in what we ascribe when we ascribe that belief in a content-specifying way. I acknowledge that that there are deep issues here, which the book does not tackle.

When it comes to intention Gozzano suspects that the means-end ideal is truistic. He says, ‘if [it] amounts to saying that if you want to  $\Phi$  then do whatever you think can bring about  $\Phi$ , then it is equivalent to giving the *grammar* of realizing an action, but it cannot provide any substantial help for giving a correctness condition for an intention’ (p. 33).

I need to make two points of clarification. First, the ideal and the corresponding commitment are tied to intention, not to wanting. (See *Understanding People*, pp. 120-121). Second, it is important that both the ideal and the corresponding basic means-end commitments deal in what is actually necessary to carry out an intention, rather than what one thinks is necessary. If I intend to  $\Phi$  and I believe that  $\Phi$ -ing is necessary if I am to  $\Phi$  then I incur a belief-relative commitment to  $\Phi$ -ing. (See *Understanding People*, pp. 75-76). But the belief-relative commitment is grounded in the basic means-end commitment. Rationality calls upon us to aspire to discharge the basic commitment. Doing that requires me to find out what are the necessary means and in trying to do so I may acquire a false belief on the matter. In virtue of such a belief I would incur a belief-relative commitment that may be at odds with the basic commitment. Even so, it is the means-end ideal that gives rise to the basic commitment, which in turns explains why I incur the belief-relative commitment. From the stance in which I have a false belief about means, to be in breach of the belief-relative commitment would be to be in breach of the basic means-end commitment. The latter is what I must avoid.

I am happy to accept that both the means-end ideal is truistic and wish that other theorists would agree. Gozzano doubts that my apparatus delivers a ‘correctness condition for intending to  $\Phi$ ’ (p. 33). I did not present it as doing so. My aim was to be explicit about the shape of

commitments incurred by intentions and to raise the question of how intentions are linked to the commitments they incur. This is the project that I call *explaining normative import*, which I pursued in chapter 4.

Gozzano takes me to task for saying that normativity is primarily a feature of judgements, beliefs, statements, claims, etc. I am not sure exactly what his worry is, though it is connected with a desire to avoid prejudging the issue of whether belief depends on language. I do not think that the offending statement does pre-judge the issue. Perhaps I should have made it explicit that normativity attaches to statements and the rest *in virtue of their content*. Thus a statement can have a content such that it is normative, and so can a belief. For instance, a belief to the effect that one ought to avoid believing contradictions, would be normative. So would a statement with the same content.

In the book I claimed that our beliefs and intentions differ in kind from any states possessed by animals that lack our reflective capacities. Gozzano's perspective on how I treat this matter is guided by his assumption that I think that there is 'some sort of irreducible gap between sub-doxastic states, those information-bearing states that manage our motor behavior in driving the car on a busy road or helping us to catch a thrown ball, and beliefs and intentions, those states whose content is consciously elaborated in a linguistic way' (p. 35). I certainly take human beliefs to be different from any sub-doxastic information-bearing states in the sense that no human belief is identical with any state that is purely sub-doxastic. Similarly, I take it that human intentional action is not identical with any sub-intentional yet goal-directed activity. Gozzano suggests that the two levels 'are not completely apart, at least not in us' (p. 36). I do not think I made any claims in the book that commit me to thinking otherwise. In connection with intention I acknowledge that sub-intentional activity may be implicated by intentional action (pp. 52-3). It seems evident that there must be an interplay between beliefs under my high conception and sub-doxastic states. Gozzano says that when driving along, unreflectively, engaged in discussion with a passenger, one may halt the discussion and (consciously) attend more to what one needs to do to drive safely in the prevailing conditions. He describes this as letting one's sub-doxastic states resurface in the conscious dimension. I prefer to view the change as a transition from sub-intentional activity, for instance, adjustments to the steering wheel, shaped by features of the situation registered at the sub-doxastic level, to noticing and thinking about what needs to be done. It is compatible with my view of belief that some beliefs are necessarily tied to sub-doxastic states. For instance, it is plausible that if I have a demonstrative perceptual belief that I would express by saying, '*That* ball is coming straight at me', then it is constitutive of my having that very belief that I am tracking the trajectory of the ball at a sub-doxastic level.<sup>4</sup> Whether or not this is right it is compatible with preserving a distinction between human beliefs and any belief-like states on the part of animals that lack our reflective capacities.

Gozzano asks whether we need to invoke a cognitive difference between him and his dog Fido in order to account for their both mistakenly believing that the cat is up a tree which they are looking at. Perhaps not. Both, after all, will be in a state in which they are primed, for instance, to search that tree if trying to find the cat. I do not suggest that to account merely for the mistake we need to invoke a cognitive difference between the two. When I defend a high conception of human believing it is to do justice to a characteristic role of beliefs in our thinking—that of supplying us with assumptions that form or contribute to our reasons for action. (See *Understanding People*, pp. 138-39, and chapter 7.) When we exploit our beliefs in this way we take stock of what we believe. Dogs presumably do not, and it is an empirical matter whether higher primates do.

Gozzano ascribes to me the view that 'to have the belief that  $p$  ( $Bp$ ) entails (in general) knowing that one has the belief ( $KBp$ )' (p. 37). I do not work with a conception of entailment that holds sometimes but not at others, where contents are held constant. But, even with Gozzano's qualification 'in general', the view ascribed to me looks stronger than any that I hold. In the book

4 I touch on the interplay between perceptual belief and sub-doxastic states in Millar 2007.

(pp. 138-39) I was careful to say that *when beliefs perform a certain characteristic role—that of supplying us with assumptions that form or contribute to reasons for other beliefs*—we know that the assumptions with which they supply us are ones we believe and that we are relying on them. The idea is that if, believing that *p*, I exploit that belief in this way, treating the consideration that *p* as forming or contributing to a reason to believe that *p*, and believe accordingly, then I shall know that I am doing so. Suppose, for instance, that I believe that Tom is going to the party and that if Tom is going then Mary will go. I then exploit those beliefs, and form the belief that Mary will go to the party on the grounds that Tom is going and that if he is going then Mary will go. The claim is that when I form this belief in that way I will know what I am doing. Perhaps there are circumstances in which believing that Tom is going to the party and that if Tom is going then Mary will go will cause me to believe that Mary will go, without my knowing why I formed that belief. In that case the beliefs that constitute the cause are not performing the role under consideration and their contents do not figure as my reasons for believing that Mary will go. We have many beliefs that are dormant standing states and that we never exploit at all or never exploit in the way under consideration.

Gozzano (p. 37) also ascribes to me the view that if I have to give a verdict on whether *q*, and I believe that *p*, and know that the proposition that *p* entails that *q*, then I ought to believe that *q*. This view is actually contrary to a central idea in the book, which is that while I would, in the circumstances envisaged, be committed to believing that *q* it does not follow that I ought to believe that *q*.

I claim in the book (Ch. 4) that under dispositionalism beliefs have a standing that is akin to that of a character trait. That is to say, believing something is considered as a complex dispositional property that is manifested in a variety of ways depending on the nature of the factors that trigger the disposition. It is a problem for this view that if it were true then it would be possible to adopt a contemplative stance to our current beliefs. Such a stance would be one on which I do not take my having the belief to have any normative import for other things I might believe or do. It does not seem to me to be possible to adopt such a stance.

Addressing my critique of dispositionalism, Gozzano says,

Dispositionalism, as specified, can be taken as a thesis concerning belief's formation: I see a red car and I come to believe that there is a red car in front of me. My perceptual apparatus is structured in such a way that if there is something red in front of me I come to believe that something red is in my visual field. Millar resists calling these beliefs; rather he takes these as purely information bearing states. If so, the problem lies in the concept of *belief* rather than in the dispositionalist view ... (p. 39).

I do not deny that belief-formation implicates dispositions. Nor do I deny that genuine beliefs can be formed in direct response to sensory experiences. My target was a conception of beliefs and intentions on which each such state implicates a cluster of non-normatively specifiable dispositions. I argued (*Understanding People*, Ch. 4) that it is not at all clear that dispositionalists can adequately account for the fact that beliefs and intentions incur normative commitments. I also tried to make it plausible that we have little general conception of which dispositions are characteristic of which beliefs and intentions (*Understanding People*, pp. 125-31). Our conceptions of what a person is liable to think or do given this or that belief or intention relate to real or imagined circumstances. They and rely on assumptions and sometimes knowledge about much else that the person believes, intends, hopes, and so on, in those circumstances. Roughly speaking, we think of the sorts of thing that it would *make sense* for the person to think or do, given the beliefs or intentions, and whatever else we know about the person and his or her circumstances. Thinking about what would make sense includes thinking about what there is reason for the person to think or do, and about what the person is committed to thinking or doing. I think that this has a bearing on Gozzano's concerns about styles of reasoning. Focusing on a particular belief tells us very little about what a person is

liable to think or do even granted that we take the implication commitments of the belief into account. In trying to make out what the person might think or do we need to know much else besides.

### Reply to Haddock

The Implication Commitment Principle is a regimentation of something that might be expressed in more everyday parlance by saying that if you believe some things you have to believe others. I take it that believers are committed to thinking this true, and could be brought to appreciate that it is true with a few examples. The regimentation makes explicit how it needs to be understood if it is to help to make sense of our thinking. First, you don't have to believe everything entailed by what you believe. But if you believe *P*, and *P* entails *Q*, then *if you give any verdict on Q* it has to be belief. Second, it interprets the notion of *what you have to believe* in terms of the idea that you have to avoid a certain combination of stances. (Analogously, saying that *Q* must be true, having noted that it is entailed by *P*, in a circumstance in which the truth of *P* is taken for granted, amounts to saying, '*P*, and necessarily, if *P* then *Q*', not 'Since *P*, necessarily *Q*.) If I believe *P*, and *Q* is entailed by *P*, then I ought to avoid continuing to believe *P* while giving a verdict on *Q* other than belief. So the sense in which I have to believe *Q* is not such that I ought to believe *Q*, since I could do what is required by no longer believing *Q*.

In calling the principle a conceptual truth I meant only to mark the fact that it is a truth and something to which one is committed if one has mastery of the concept of belief. Notwithstanding our post-Quineian felicity, I know of no good reason for doubting that there are conceptual or analytic truths in the sense explained. As Adrian Haddock suggests (footnote 3) there is an oddity in supposing that an unmarried male in a civic partnership with another unmarried male is a bachelor. That seems to me to show, not that there is some doubt as to whether it is analytic that a bachelor is an adult male who has never married, but rather that the concept of a bachelor has little use in current social circumstances. At a time at which there was a prevailing expectation that most males would end up married in the conventional sense, and that those that did not, would not enter into anything which, like a civic partnership, is akin to marriage in the conventional sense, there was some point in dividing adult males into those who are or have been married and those who are not and have never been. Given the proliferation of statuses that exist now and are widely acknowledged, the concept has little use, for want of anything interesting in common among those to whom it applies. (It's ages since I have heard anyone described as a bachelor.)

Haddock suggests that the Implication Principle might be conceived as a necessary truth that is not analytic. It is widely held that the claim that water is (identical to) H<sub>2</sub>O is necessary but not analytic. Its necessity reflects the fact that it specifies the nature of water. I take Haddock to be suggesting that we might similarly think of the Implication Principle as partially specifying the nature of belief even though it is not analytic. The position taken in the book is that the concept of belief is not the concept of a natural kind or in the relevant way akin to such a concept. I assumed that the nature of belief is, as I put it, captured or specified by the concept. (See *Understanding People*, pp. 191-92.) Haddock wonders exactly what I mean by this (p. 44). He conjectures that the underlying thought might be that 'if *S* grasps the concept of *X*, then she believes all the truths there are to believe about the nature of *X*' (p. 44). It is just as well that I am not committed to such a strong and obviously implausible thesis. The idea is rather that the concept of *X* captures the nature of *X* if there is no more to being *X* than would be ascribed to something by saying of it that it is *X*. I take it that the statement that a substance is water does not meet this condition. There is more to a substance being water (its being H<sub>2</sub>O) than is stated by saying that it is water. I claim that there is no more to someone's believing that the stock market is volatile than is stated by saying that that person believes that the stock market is volatile. That the nature of belief is captured by the concept explains why our epistemic route to the nature of belief is via reflection on the concept. What other route could there be? Some claim that the concept of belief picks out a physicalistically specifiable

state, the character of which can be investigated empirically. But so far as I can see there is really very little reason to believe that this is so. (That is compatible with supposing that there would be no beliefs if our neurophysiological systems did not function in appropriate ways.) Given my stance on the status of the Implication Principle, the issue that preoccupies Haddock—whether calling the principle constitutive adds anything to calling it necessary but not analytic—was not an issue for me. I acknowledge that it is something worth pursuing just because it takes us into interesting methodological issues about the metaphysics of belief and about whether philosophising about belief should be viewed as a form of conceptual enquiry.

The stance I have just taken has an impact on at least part of Haddock's attack on my case for thinking that we can best account for the fact that believing incurs an implication commitment in terms of the idea that it is part of the nature of belief that beliefs incur such commitments. He suggests that we could as well account for the commitments incurred by beliefs by invoking the Implication Principle understood as necessary but not analytic. While I think that this is an interesting suggestion, it has problems of its own. If belief is not in a category akin to a natural kind then the question arises as to how we determine that the principle is a metaphysically necessary truth. Be that as it may, I resist dropping the suggestion that the principle is a conceptual truth because it seems to me to be something that our actual thinking about belief commits us to. For all that Haddock has said metaphysically necessary truths about beliefs need have no impact on our actual thinking. Even if that is not objectionable in itself, we should not wish to place the Implication Principle, or at least its implications for particular cases of belief, where it lacks such impact.

More pressing, I think, is Haddock's suggestion (p. 50) that we may think of the Implication Principle, as having to do with the nature of *rationality*. As Haddock acknowledges I raise the question of why the Implication Principle should be thought hold. The issue concerns what licenses the transition from, say,

(1) Sally believes  $P$

(2)  $P$  entails  $Q$

to

(3) Sally committed to believing  $Q$ , if she gives any verdict to  $Q$ .

A very natural thought is that all we need to explain this move is a conception of rationality on which rationally demands that Sally believe  $Q$ , in circumstances in which (1) and (2) are true and Sally gives a verdict on  $Q$ . If this is right then there is no need to assume that it is partially constitutive of Sally's believing  $P$  that she incurs commitments to believing entailments of  $P$ . In the book (pp. 117-18) I reject such a move in the context of a discussion of dispositionalism. The reason is that if it is not in the nature of believing that it is commitment-incurring in the manner described then it becomes an issue why we should suppose that the Implication Commitment Principle holds. The problem emerges when we consider that, on the dispositionalist view, believing  $P$  is akin to having a character trait—it is simply a matter of having a cluster of dispositions. If I am an aggressive person then I am unusually prone to being aggressive. Thus in situations that would not provoke most others I am liable to be aggressive. If this is how I am then it might well be rational for me to do something about it. Yet there is nothing about the trait of being aggressive that in and of itself explains this. To explain it we need to take account of the impact of my being aggressive in the light of what is of value. It is of value that I should maintain good relationships with others. Since my being aggressive will tend to undermine, or prevent me from having, such relationships, there is good prudential reason for me to try to be less aggressive. Now, if Sally's believing  $P$  is a dispositional state in the way that my being aggressive would be, similar considerations would apply. Under dispositionalist assumptions, appeal to rationality alone will no

more explain why Sally should avoid continuing to believe *P* while giving a verdict on *Q* other than belief, than appeal to rationality alone will explain why I should do something about being aggressive. It is easy to lose sight of this because it seems so obvious that Sally should avoid the condition specified. But, even though it is obvious, this needs an explanation. I claim that the best explanation is that belief is essentially commitment-incurring in the way I have described. What rationality demands does not drop out of the sky; it is grounded in the nature of the states (for instance, believing and intending) on which it imposes constraints.

Haddock draws attention to difficult issues concerning the extent to which our beliefs are under our control. He imagines a theorist who believes that our nature, being determined by evolution, leads us to have a tendency to have beliefs that are racially discriminatory, while seeing no reason to believe such things. Haddock takes the example to be suggestive of two points that pose problems for my views. One is that if I have a discriminative belief, which my nature determines that I should have, then it becomes questionable that I ought to avoid continuing to have the belief while giving verdicts on what it entails other than belief. For even if I come to know that some things entailed by the belief are false it is not in my power to abandon the belief. The other point is that while I *should* adopt a contemplative stance towards the discriminatory belief, my theory has it that such a stance is not open to me. These are important matters and they certainly require discussion beyond what I have provided so far. In the book (pp. 76-78) I acknowledge that failure to discharge the commitments I describe need not be failure in rationality. Sally may reject *Q* because ignorant of the fact that it is entailed by *P*, which she believes. Her ignorance need not be due to irrationality. To address this, and a parallel point concerning means-end commitments, I distinguished between what I called *ideals of reason* and *requirements of rationality*. It is an ideal of rationality that one avoids believing things while giving verdicts on what they entail other than belief. On the picture presented in the book, particular implication commitments are grounded in this ideal. Limitations on our logical acumen mean that it is possible to be in breach of the ideal through ignorance that simply reflects our cognitive limitations, rather than irrationality. Yet it is part of what it is to be rational that we are required to aspire to satisfy the ideals so far as it is in our power to do so. An implication of this approach is that if implication commitments are expressed in terms of 'oughts' then in the relevant sense 'ought' does not entail 'can'. But it is far from clear that 'ought' always does entail 'can'. In particular, it seems to me that we can make some sense of 'oughts' linked to ideals, for which 'ought' does not entail 'can', but does entail something like 'can aspire to conform to'. Haddock is especially concerned with beliefs that it is not in our power to abandon. It is tempting to address his concerns simply by noting the point about 'ought' that I have just made. While I am sure that this should be part of an adequate response, it is less than satisfying taken on its own. It is central to my perspective that believing something is of its very nature sensitive to the commitments which that belief incurs. If it is not open to me to abandon a belief, then a severe limitation is imposed on my ability to discharge the commitments that it incurs. In view of this it might be wondered whether I may count as having the sensitivity to the commitments that, on the story I tell, is constitutive of what it is to believe something. I shall make a couple of remarks on this. First, suppose that there are beliefs that their subjects are powerless to abandon. It is a further question whether we can make sense of people maintaining such a belief, while also acknowledging either that there is no reason to think it true or that it entails things known to be false. A more plausible scenario would be one in which there are psychological mechanisms in place that do one of two things. Either they make the subject blind to the fact that there is no reason to think the belief true or evidence for thinking it false, or they lead the subject to pretend that there is reason to think the belief true and to explain away what is alleged to be evidence of falsehood. In cases of pathological believing, sensitivity to relevant commitments may be manifested in such ways. Second, if a subject appears to believe something while fully aware that there is no reason to think it true, or even aware that there is strong evidence that it is false, there is a real question as to whether the situation is at it appears. That is a Davidsonian consideration but it is far weaker than

any interpretation of the constitutive role of rationality on which a belief is taken to be problematic if it is not rationally held. Here we are not simply imagining an irrational belief, but rather a belief held by a subject who in effect recognises that it is irrational. Are people not often in this situation? Are not those who suffer from depression often powerless to abandon beliefs about their various inadequacies, even though they know that there is no reason to think these beliefs true? Again, it seems to me that we must be careful how we describe cases. One possibility is that such subjects have fluctuating beliefs, alternating between acknowledging that they are not really so inadequate and, under the dark clouds of depression, believing that they are inadequate. These fluctuations pose no problem for my view and do not implicate the kind of condition that Haddock envisages his theorist with the racist beliefs to be in: they do not require us to ascribe to the subject a belief at the same time as acknowledging that there is no reason to think it true. There might be cases in which a depressed subject acknowledges that it is false that he or she is inadequate but continues to *feel* inadequate. Again that is not the condition that Haddock takes his theorist to be in. I think we would find it hard to know what to make of a person who appeared to be in this condition. That would reflect the fact there is a real question then as to whether the subject who is supposed to have the unavoidable racist belief counts as having that belief or, if he does, whether he is truly in a situation in which it is not open to him to change his mind. An alternative understanding is that such a person cannot avoid having racist *thoughts* spring to mind. That would fall short of having racist *beliefs*. And even granted that the subject has the belief, this would not settle that he should adopt a contemplative attitude towards that belief. I remain sceptical that it is open to us to adopt such a stance to our current beliefs.

### **Reply to Noordhof**

Paul Noordhof's lengthy and detailed scrutiny of the book raises many issues that merit fuller treatment. I shall do my best to try to defend or at least clarify some of the points with which he engages.

#### *Normative reasons for action and justification for action*

The first part of Noordhof's discussion is concerned with my view that not all normative reasons for action are justificatory reasons. Noordhof thinks that, on the contrary, '[i]f it seems that justification is not available for a certain kind of attitude, then there are no reasons and the attitude, and actions which follow from it, fall, at least partly, outside the ambit of normative explanation' (p. 50)

My view about normative reasons for action, outlined in Chapter 2 of the book, is based on considerations about what we would naturally count as reasons for action and also on more theoretical considerations about the nature of intentional action. For instance, it seems to me quite natural to suppose that my intending to buy fresh fish today, along with certain facts about the availability of fish near where I live, and about the time and expense I am prepared to put into finding fish, provide a reason for me to go to our local fish shop. The reason here gives an instrumental point to my performing a certain action: it is a step towards buying fresh fish. It strikes me as odd to regard my reason as providing a justification for my going to the fish shop. The sense of oddness has, I think, two sources. The first is that the fact that an action would contribute to doing something I intend does not by itself justify my performing that action. My intention to buy fish is innocent enough, but many intentions, and the means to carrying them out, are not. The second is that, in the absence of further description of the scenario, talk of justification seems out of place. Not all actions call for justification. I take Noordhof's point (p. 50) that there can be justification for an action even when there is no uniquely right thing to do in the circumstances. Even so, a justification for an action must surely show that there was at least nothing wrong with performing that action. The reason in the example I have just given does not do that, but nor does it have to do that in order to be a reason for the action.

Noordhof might be moved by a sense that to count reasons for action as normative when they do not provide justification is to empty the notion of the normative of any content. As I see it, my reason for going to shops is normative because, though it would not be right to say that it justifies my going to the shops, it provides a rationale for the action in that it confers a point on that action. To that extent it recommends or favours my going to the shops. To that extent it is, therefore, a normative reason. In this sense even perverse or foolish intentions can generate a reason for an action—a reason in the light of which the action can be seen to have a rationale and to be thus far intelligible. But no action would on that account alone be justified.

This takes us to the more theoretical considerations. What counts as a reason for a belief, or as a reason for an action, depends, respectively, on the constitutive aim of belief and the constitutive aim of intentional action. The constitutive aim of belief is truth. Consequently, a normative reason for a belief that  $p$  cannot do less than justify thinking it true that  $p$ . I claim that the constitutive aim of intentional action is simply that the action should have a point and that an action can have a point merely by being a means to carrying out an end that the agent actually has, whether innocent, noble, foolish or perverse. Some actions call for justification. They call for justification in the face of considerations that suggest that performing the action might, for instance, lead to some harmful result, or be in breach of, or make it impossible to fulfil, some obligation. Whereas the intrinsic character of belief determines that a reason for belief should justify that belief, the intrinsic character of intentional action does not determine that a reason for action should justify that action. Whether an action calls for justification and whether it is justified depends on how it stands, or would stand, in relation to what is of value, and what is obliged. There might be an issue as to whether I may use some time for leisure activity, arising from obligations I have to perform certain tasks. In that event spending the time on leisure might call for justification. Whether it is justified will depend on whether the leisure activity is compatible with performing those tasks or perhaps on whether, in any case, I need the leisure time to conserve my own energy and health. For an action to be justified more is required than that it should have a point (not be pointless).

I work with a conception of the commitments incurred by beliefs and intentions that I acknowledge to have affinities with John Broome's conception of normative requirements. This being so, Noordhof (p. 53) wonders why I do not stick with the idea that having an intention incurs a commitment to taking the necessary means and drop the claim about reasons? Part of the answer is that it restores common sense to acknowledge that having intentions can generate reasons to take the necessary means. Suppose I ask my wife whether we have any reason to go into town today. She answers, 'You said you were going to cook fish this evening'. She is reminding me of my intention to cook fish and implying that so long as that remains so we have a reason to go into town. (We like to check out the fish together.) If my intention to cook fish does not provide a reason to go into town, then it looks as if my wife's answer is beside the point and that she has failed to grasp what it takes to be a reason for action. But her reaction clearly makes sense. I want to do justice to that. I do not, however, rest content with intuitions. There are the wider considerations that contrast belief and intentional action, with respect to their constitutive aims and the wider considerations about what genuine justification is.

#### *Arational, yet intentional, actions?*

Noordhof tackles my discussion of supposedly arational actions, in the sense of Rosalind Hursthouse (1991). Such actions are supposed to be done for no reason. Examples include kicking a car when it does not start or tearing up the photograph of a rival. Noordhof suggests that

... the challenge of the examples is not removed by allowing that the reasons need not justify the agent. The point was rather that the things that might flash before the agent's mind resulting in the action couldn't possibly be taken as considerations in favour of the action in any sense. (pp. 51).

Noordhof's idea is that, for instance, kicking the car won't make it start; nor will it teach the car a lesson. So the reason for the action cannot be that it will bring about those ends. I agree but I do not look for the rationale for the action in these terms. By my account the reason for the kicking the car is to satisfy an intention to give way to the urge to do so. Noordhof says

Urges may causally explain the actions but the actions aren't done to express those urges. They are not the consideration why the agents acted in the way that they did. Indeed, if urges were the reasons for the actions, then the agent would appear to have a justification after all. (p. 51)

Obviously urges are not considerations. They are intense inclinations. The urge to kick the car is borne of frustration and anger. In the absence of some pathological mental condition, kicking the car is likely to be intentional. If intentional it is informed by an intention to kick the car, but that intention, by itself, does not bring out the character of the action because it does not reveal the point of the action. Nor, by itself, does the urge to kick. What brings out the character of the action, and reveals its point, is that the agent intentionally gave way to the urge. He kicked the car because, as we might say, he felt like it and gave in to that feeling. But he was not under any compulsion to kick. He meant to satisfy the urge to kick the car and that confers a point on his doing so. It might be that agents are sometimes prompted to violent acts in the grip of passions so strong that they are quite out of control. It seems to me that if an agent were to kick a car in such a state then what he does would be too far removed from anything he means to do to count as an action of intentionally kicking the car. If car-kicking really is intentional, as opposed to the out-of-control manifestation of a rage (and it would have to be some rage), it has to have some point beyond itself. My proposal is that the point is to give way to, and thus to satisfy, the urge to kick the car. We need something like this to make sense of how the act can be something that the agent meant to do—something of which he is, therefore, the author. But it also seems to me to be very clear that neither the urges, nor intentions to satisfy them, provide justification for kicking the car. If we allow ourselves to talk as if they did we deprive ourselves of convenient means of distinguishing, as we should, between what I would regard as genuine justification and the kind of point-conferring rationale that figures in my treatment of these and other examples.

Noordhof thinks that my view distorts the phenomenology of the actions in question. At one point, speaking of jumping for joy, or punching the air from joy, he says, 'We don't have a desire to jump up and down, or punch the air, and then satisfy it, yet the point of the action is to express a certain emotion' (p. 51). He adds that '[p]hilosophers should think long and hard before they commit themselves to understanding such cases in terms of a desire to express an emotion and a belief that (say), if it is joy that one is seeking to express then jumping up and down would be a good way of doing it' (p. 51). The aim here is to embrace the idea that actions springing from emotions can have a point, and on that account be intentional, and yet not be done with a view to expressing the emotion and so not be done for such a reason. We agree that there are genuine actions that in some sense express, or are borne of, emotions, or longstanding attitudes like love and affection. Examples would be kissing one's children, hugging someone who has suffered bereavement, warmly shaking someone's hand, and jumping for joy. Some such actions may be done in order to express the emotion or attitude. Here the expression might be a matter of communicating to the other what one feels. But the action need not be done with this kind of aim. Hugging the bereaved person out of love or concern might simply be aimed at comforting that person. All of these are cases of intentional action. It seems to me that they are all cases in which the agent acts for a reason—to communicate love, affection, or concern, to comfort, and so on. Noordhof's worry about jumping for joy might be that if we think of such actions as being guided by a belief that jumping is a means to the expression of joy then we over-intellectualise what is going on and underplay the spontaneity of the act. But to think of them as being both intentional and done for reasons does not commit us to over-intellectualisation and it is important not to

overlay the spontaneity. It is not in dispute that such acts do not spring from deliberation and need not involve a step from consideration of an aim to choosing the means to satisfy that aim. Yet greater or lesser degrees of know-how and social skill lie behind them. Not any kind of hugging will serve the purpose of comforting or expressing concern. We don't hug just anybody and we don't jump for joy just anywhere no matter who is in our company. Or if we do we are liable to incur censure or ridicule. Such actions can have a measure of spontaneity while also being informed by understanding. The agent's reasons for doing them—giving comfort, showing concern, expressing joy—shapes how they are done and bring into play the agent's understanding of how they are to be done. It is not to the point that one does not think to oneself, 'Doing this is a way to achieve this end'. An understanding to this effect is nonetheless manifested by the way in which the action is modulated by the aim and knowledge of the means.

It is worth adding that there are many actions that are indisputably done for a reason though the agent does not consciously think, 'This action will achieve that end'. Suppose that following a routine of heading home from work around a certain time, I switch off my office lights, and leave my office, and walk in a certain direction. I need not consider, in any ordinary sense of 'consider', why I am doing these things. Yet I am doing them for a reason—I am taking steps towards going home, knowing what I am doing, and why.

### *Commitments and practices*

When he turns to consider my view of commitments incurred through participation in practices, Noordhof charges me with taking practices too seriously (p. 59). My account of practices was intended to treat them with no more seriousness than they deserve.

A practice is an essentially rule-governed activity or cluster of such activities. Some practices are deliberately instituted with explicitly formulated rules. Others grow without design. Certain ways of proceeding come to be understood as the *thing to do*. These understandings are manifested, not only through mutual expectations of conformity, but also through critical reactions to breaches of the rules, and the recognition that failure to conform demands explanation, justification or excuse. Recall that the commitments incurred by beliefs and intentions are quite independent of whether the belief or intention is reasonable. The commitments are, so to speak, what you sign up to by forming the belief or forming the intention, for good or ill. Similarly, the commitments incurred through participation in a practice are independent of the reasonableness of participating in the practice. They are what you sign up to by participating in the practice, for good or ill. To take practices too seriously would be to suppose that participating in them obliges one to follow their rules. I explicitly set myself against any such view. It seems to me that we do not take them seriously enough if we deny that it is constitutive of practices that participating in them incurs commitments to following their rules.

Some might be attracted by the idea that the rules of a practice do not generate any normative commitments beyond those incurred by aiming to participate. If we ask what reason is there for players of a game of football to follow the rules, the answer by this account would be, 'Because they intend to play the game and could not do so without following the rules at least to some extent'. On this way of thinking the only normativity we need to acknowledge in connection with rule-governed activities is instrumental—following the rules to some extent is what you have to do if you are to play. The rules have no normative force; it's just that you won't count as playing unless by-and-large you conform to them. If this is right there need be nothing wrong with a football player's being in breach of a rule. Flouting a rule evidently need not be at odds with being a player and so does not put one in breach of the commitment incurred by intending to play the game—the commitment to doing what is necessary to that end. All we have is a clash between what the player has done and what some rule prescribes. I do not think that this is satisfactory. There is something wrong about being in breach of a rule and this is reflected in the fact that breaches of the rule are subject to legitimate criticism. Of course, in the professional game, players try to get away

with as much as they can. But criticism is still appropriate, based on what is called for if one plays this game. I suggest that to explain this we need the notion that it is in the nature of rules governing a practice that those who participate in the practice incur a (normative) commitment to following the rules. Behaviour that flouts a rule is open to being criticised because it is in breach of that commitment. Incurring this commitment does not entail that you ought to follow the rules. There can be games so appalling that no one ought to play them and therefore no one ought to follow their rules. By my account it remains true that if you play such a game you incur a commitment to following the rules. The commitment amounts to this: you ought to avoid becoming a player, or continuing as a player, and not follow the rules, unless there are countervailing considerations. (On the latter, see *Understanding People*, p. 86.)

Noordhof asks, '[W]hat is wrong with playing along, and hence participating in the practice, and yet eschewing the practice-relative commitments as not applying to you?' (p. 59). I should say first that the commitments incurred by participating in a practice are practice-relative only in the sense that one incurs them through participation, not in the sense that the commitments only count as such from the perspective of the practice. Whether one is a participant or not one is obliged to avoid being (or becoming) a participant while not following the rules, unless there are countervailing reasons. This obligation is unconditional and arises from the nature of a practice in general. One may be indifferent to the obligation since one might not be a participant and there might be no live issue as to whether to become one. The obligation would come into play if one were a participant since, absent countervailing reasons, it would in that case be incumbent upon one either to follow the rules or withdraw from the practice. That is what incurring the commitment to following the rules amounts to. There might be nothing wrong—indeed, there might be something right—about playing along within a practice, while thinking that there ought not to be this practice with those rules. Maybe this ought to be the attitude of the police officer in a police force with corrupt practices. But it is incoherent to suppose that one can both view oneself as a participant in a practice and not view the rules as applying to one. And to view the rules as applying just is to view oneself as *having to follow them*, as one might put it in ordinary speech. If in the midst of a board game I said, 'I am playing this game alright but this carries with it no obligation relating to how I treat the rules', my fellow players would find this astonishing. No doubt a component of their astonishment would be directed at how I could expect them to continue playing with me. But it seems to me that part of their astonishment would be an inchoate recognition that it makes no sense to suppose that one can continue playing while remaining indifferent to what the rules require. Noordhof's example of the hunters of humans does not force a different treatment any more than the commitments incurred by crazy beliefs and intentions compels us to suppose that these attitudes are not genuinely commitment-incurring. Part of what you sign up to by participating in a practice is *following its rules*. You might not follow some of the rules while remaining a player because you can cheat covertly, or cheat openly and take the consequences. Or you might flout the rules, as whistleblowers do, to subvert the practice from within, while remaining a participant for some time. But you have still signed up to following the rules. The metaphor of signing up might suggest that if you are at fault at all for flouting the rules it is because you have broken an implicit or explicit contract requiring you to follow the rules. That is not what I mean to suggest. We need the idea of a genuine commitment in order to make sense of the criticisms that would be appropriate on breaches of the rules. The commitment need have nothing to do with morality. Morality might dictate that either you withdraw from the practice or subvert it from within. It could be that morally you ought not to carry out this or that commitment. Nonetheless, it is your having incurred the commitment that demands that there should be a good reason for neither carrying it out nor discharging it by withdrawing from the practice. Somewhat similarly, the fact that there is overwhelming evidence against some implication of what you believe dictates that you should not accept that implication. But it is the commitment to accepting it that requires you, in that case, to give up the belief that incurred it. (In this case there is no room for the option of sticking with the commitment-incurring

belief yet not accepting the implication despite acknowledging it to be an implication.)

*Dispositionalism about belief and intention*

A key part of my case against dispositionalism about the attitudes is that we cannot adopt a purely contemplative attitude to our own current beliefs and intentions. In recognising that I intend such-and-such or believe such-and-such, I endorse that intention or belief, as the case may be. It is, therefore, not an open question whether having the belief or intention has normative import for me (commits me to further beliefs or to a certain course of action). Noordhof attributes to me the view that ‘if we currently intend to do *A* but do nothing to ensure the implementation of *A*, we don’t really have an intention at all’ (p. 61)’. He also attributes to me the corresponding view concerning belief. On this basis he thinks that my view of the normative dimension of belief and intentions—that believing and intending are essentially commitment-incurring—does not help with understanding why we cannot adopt contemplative attitudes to our current beliefs and intentions. Noting that I compare believing and intending with promising, Noordhof suggests that on my conception of promising it would be possible to adopt an external attitude to the fact that I have made a promise, even though since I have made a promise I have incurred a commitment to doing what I promised. I would be adopting an external attitude if I were indifferent to doing what I promised or just curious to see whether I will end up keeping the promise or breaking it. This, he thinks, is suggestive of a problem for my conception of belief and intention.

... if the normative character of our beliefs and intentions involves commitments which are simply the result of constitutive aims and principles grounded in the nature of these attitudes—just as the practice of promise-keeping has certain constitutive rules concerning the practice-relative commitments of a promisee—then we should expect that it is possible to adopt the external perspective to our intentions and beliefs as well. (p. 61)

There are issues here that need separating out.

- (i) There is no problem with the idea that it is possible to take a contemplative (external) attitude to one’s own promises. That is because, recognising that one has made a promise, does not implicate an endorsement of the promise. Belief and intention are unlike promising in this respect. You might exploit the practice of promise-keeping in order to deceive others. In doing so you would distance yourself from the commitments you have incurred. You can distance yourself from beliefs or from intentions you *have had* by wondering whether you should retain those beliefs and intentions. But as I see it, when you do this, the beliefs and intentions are, so to speak, suspended. What you cannot do, I claim, is distance yourself from a belief and intention that you recognise that you now have. Accordingly, you cannot distance yourself from the commitments incurred by these beliefs and intentions, though you might, of course, overlook them. Noordhof, I suspect, takes me to press the analogy between beliefs and intentions, on the one hand, and promising, on the other, further than it will go. There is nothing in the story about beliefs and intentions that corresponds to the rules of a practice. The rules of a practice exist only because those participating treat them as rules. Part of treating them as rules is recognising that so long as one participates one is normatively committed to following them. (Whether one is committed in the psychological sense is a different matter.) The implication-commitments of beliefs are the resultants of the contents of those beliefs and the attitude of believing. The means-end commitments of intentions are the resultants of facts about what is necessary to carry out those intentions and the attitude of intending.
- (ii) I do not hold, as Noordhof supposes, that it is a necessary condition of my intending to do something that I do something to carry it out. Such a view would make it impossible to make sense of *lapsed intentions*—intentions we do not carry out but do not give up. We clearly can

have lapsed intentions. I might, for instance, forget a dentist's appointment though I intended to keep the appointment and at no point changed my mind about doing so. The contemplative attitude of which I speak would be one in which I recognise that I believe this or intend that and yet do not view this as having any normative import for what else I believe or what I do.

Noordhof at one point suggests that the component of my story that does the work is endorsement not the account of commitments. It is true that someone might be attracted by the idea that recognising that one believes this or intends that involves endorsement of what one believes and intends, while not yet viewing this as having any implications concerning the normative dimension of believing and intending. However, I envisage opponents agreeing that when you believe or intend something you incur commitments, while denying that this establishes anything about the (intrinsic) normative dimension to believing and intending. I take it that if these opponents acknowledge the point about endorsement, then they should see that endorsement involves taking on board the commitments incurred. The question for them is how to explain why this should be so on a purely dispositionalist account on which incurring the commitment is not intrinsic to the attitude. In the book I point to a difficulty about attempting to deal with this by invoking covering general principles.

A further problem for dispositionalism relates to the role of beliefs and intentions in rationalizing explanation (*Understanding People*, pp. 125-31). Dispositionalism is suggestive of the idea that rationalizing explanations are dispositional in that they explain belief-formation and action in terms of the triggering of the dispositions that characterise the attitudes invoked in the explanation. I argue that this is unhelpful since we have very little idea of what dispositions characterise believing and intending.

My conception of the normative dimension of believing and intending is meant to supply an alternative picture that addresses these problems. According to the account, there is no 'logical gap' between believing or intending and incurring the appropriate commitments. In viewing someone as believing this or intending that we are thereby committed to viewing them as incurring the relevant commitments. This is why endorsing our own beliefs and intentions and signing up to whatever commitments they incur go together. We have no clear grasp of how an agent will be disposed in virtue of having this or that belief or intention which is expressible in terms of non-normatively characterised dispositions. Our grasp of how an agent is likely to behave in virtue of having this or that belief or intention, and being in this or that circumstance, are thoroughly imbued with normative considerations. Roughly speaking, we think of what it would make sense to think or do in the specified circumstances, given the belief or intention in question. That involves thinking about what there is reason to think or do, and what one would be committed to thinking or doing. There is a disposition in the picture: we treat people as being disposed to do what would make sense, or what would seem to make sense to them, given what they believe or intend and the circumstances in which they are placed.

Noordhof suggests that, for all I have established, it could be that the causal role of a belief admits of a 'fully adequate non-normative characterisation' (p. 62). I do not claim to have established that no such characterisation is possible. However, (i) the history of gestural descriptions, and sketchy Ramsifications, bearing on how one would be disposed if, say, one believed that it was raining, do not encourage hope, and (ii) there is no evident reason to suppose that such characterisations would have, or do have, any role in our actual thinking about belief and intention—the thinking that, in my picture, pins down what the states of believing and intending are.

### *The high conception of belief and intention*

In the book (chapter 5) I defended the view that there is a certain reflexivity to believing and intending. I suggested that a characteristic role of intention is to get us to do things by way of carrying out those intentions. I claimed that when we carry out an intention we know that we have

that intention and know that doing the thing in question is by way of carrying it out. Similarly, a if I fail to act because I forgot my intention characteristic role for belief is to provide us with assumptions that form or contribute to our reasons for belief and action. I claimed that when what we believe serves this reason-constituting role we know that the assumptions in question are things we believe and that we are relying on them. (See *Understanding People*, p. 138.) In the present state of philosophy of mind, these are controversial claims. They call for closer attention than I gave them in the book or can give here.

Noordhof's first concern about my view is that it 'does not seem sensitive to the distinction between being conscious that one has a certain propositional attitude and knowing or believing that one does' (p. 65). I recognise that knowing or believing that  $p$  need not be conscious, if by that is meant being in a state that involves thinking to oneself that  $p$  (so that it is before one's mind that  $p$ ). Noordhof claims that with this concession I am 'no better off' (p. 65). Expanding, he says, 'if I fail to act because I forgot my intention, then the action is not secured by the equally forgettable belief that I have the intention' (p. 65). Indeed so, but the importance of the reflexivity of intentions lies not in securing that our intentions will be carried out—nothing is going to secure that. It lies in the fact that when we do carry out our intentions, as opposed to doing things caused by our intending, we do not act blindly—we know what we are aiming at and that means knowing what we intend.

Noordhof's second worry about reflexivity is a regress problem: 'there will be some states that play their role in the absence of beliefs about them, for example, the higher order beliefs themselves' (p. 65). I think it important to recall here that where beliefs are concerned the role on which my discussion focused was that of supplying us with assumptions that form or contribute to our reasons for action. As I write, I believe that I intend to go to a seminar in Edinburgh next week. This belief will, I hope, be brought to mind in the meantime, making me aware of a reason I have to plan for going and to avoid taking on any projects incompatible with going. Here a higher-order belief has the potential to play the reason-providing role of belief. It plays this role by making its content available for reasoning. That role does not require me to have a further belief to the effect that I believe that I intend to visit Edinburgh next week. But, of course, should the question arise whether I have this further belief, I would answer in the affirmative, though with some puzzlement that the question was raised.

Noordhof's third problem for reflexivity is whether it solves the problem of deviant causal chains. I did not take myself to have solved the problem. The point of the example of the nervous climber was to give some (tentative) support for Searle's view that the content of intention is self-reflexive. I am rather doubtful that the problem of deviant causal chains can be solved. But my scepticism is linked to a suspicion that it need not be solved. If one is in the business of providing a reductive analysis of *intending to  $\Phi$*  then it becomes pressing to provide conditions that will screen out deviant cases come what may. It is becoming increasingly doubtful that this is the business we should be in. The absence of such an analysis is not an obstacle to recognising that the climber did not carry out his intention to let go of the rope: if he had done he would have intentionally let go of the rope, but he didn't intentionally let go of the rope; it was his nervousness that made him drop it.

### *The explanatory relevance of normative*

The problem of explanatory relevance that I address arises within the realm of the mental. The problem is that 'if ...the explanatory insight provided by rationalizing explanations is causal, and tied to appropriate generalizations, then it is not clear what the provision of the rationalization adds' (*Understanding People*, p. 192). I do not claim to have tackled the problem of explanatory exclusion, which concerns how the mental can be causally explanatory. (For some tentative remarks in that direction see, my response to Mario de Caro). I take for granted that propositional attitudes are causally relevant to the formation of other attitudes and the performance of actions. The task is to explain how the fact that attitudes rationalize a belief or action can be relevant to the explanation of why the agent forms that belief or performs that action. This can be conceived as an issue about

integration, as Noordhof suggests, but I do not view it as an issue about integrating the mental with the physical, or integrating mentalistic explanation with physical explanation. Noordhof thinks this is a lacuna, but unless one thinks there is a real doubt as to whether the mental is causally efficacious, it's not clear to me that it is.

I observed in the book that the dispositionalist approaches to the problem of explanatory relevance that I consider are unsatisfactory because they attempt to solve the problem 'without reference to the stance that agents have on their own reasons' (*Understanding People*, p. 199). Let us suppose that the consideration that  $q$  is a reason for me to believe that  $p$ . For all these accounts tell us, it might be the case that I believe that  $p$ , for the reason that  $q$ , though the fact that the consideration that  $q$  is a reason to believe that  $p$  has nothing to do with why I believe that  $p$ . Echoing a point about acting from duty made by Kant, we should distinguish between believing that  $p$  in conformity to reason (to what there is reason for me to believe), and believing that  $p$  from reason (on account of there being a reason for me to believe). The accounts I am criticising do not explain what it is for me to believe that  $p$  for the reason that  $q$ .

My approach to the problem was to identify a type of case in which it is tolerably clear that the agent believes something for a certain reason and to suggest that the farther removed from this type of case we go, the less clear it will be that we have a case admitting of genuine rationalizing explanation. The feature of the case that mattered was that the agent takes the consideration that constitutes the relevant reason to be a reason for the belief in question. Noordhof (p. 68). responds by saying, in effect, that if taking the consideration that  $q$  to be a reason to believe that  $p$  just means that the subject is disposed to respond in the appropriate way, then we are no further forward. That, he thinks, is because we lack a story of how the commitment-inducing properties of taking the consideration that  $q$  to be a reason to believe that  $p$  'have an explanatory impact' (p. 68).

My account was an attempt to do justice to a plausible—I think irresistible—idea that would hardly be in doubt but for the mistaken view that an adequate conception of belief and intention must not take these attitudes to be inextricable from reflective capacities. The idea is that believing something, or doing something, for a reason involves having an understanding of one's reasons as reasons. If I believe that my wife has already arrived home for the reason that, approaching our house, I see her car in the driveway, I not only understand what it is for her car to be in the driveway and what it is for her to have already arrived at home, I also understand that the fact that her car is in the driveway is a sign, and thus a reason to think that, she has already arrived at home. There may be variations in how such an understanding is articulated. Not everyone wields the vocabulary of reasons and justification with the ease of philosophers. The understanding would be manifested, for instance, if I were to explain or justify thinking that my wife is already at home. How does this help with the problem of explanatory relevance? Because in believing that my wife has already arrived home, I am responding to the fact that her car is in the driveway *qua* reason for thinking that she has arrived home.

Noordhof focuses attention on my taking it that the fact that my wife's car is in the driveway is a reason to think she has already arrived home. How, he asks, can *this* belief have explanatory impact in virtue of its commitment-incurring and reason-giving properties. Perhaps I have not fully understood Noordhof's point. It seems to me that as so far stated the answer is straightforward and unthreatening. The answer is, 'In the same way as any other belief'. If the higher-order belief comes into play in such a way that its content becomes my reason to believe something, or directs me to some commitment of so believing, then I shall be responding to this content *as* constituting a reason or as responsible for a commitment.

## References

McDowell, J. (1981), "[Anti-Realism and the Epistemology of Understanding](#)" in Herman Parret and

Jacques Bouveresse (eds.) *Meaning and Understanding*. Berlin: Walter de Gruyter. Reprinted in McDowell 1998.

McDowell, J. (1982), “[Criteria, Defeasibility, and Knowledge](#)”, *Proceedings of the British Academy*, 68, 455-79. Reprinted in McDowell 1998.

McDowell, J. (1998), [Meaning, Knowledge, and Reality](#). Cambridge, MA.: Harvard University Press.

Millar, A. (2007), “[What the Disjunctivist is Right About](#)”, *Philosophy and Phenomenological Research*, 74, 176-98.

Millar, A. (forthcoming 2008), ‘[Perceptual-Recognitional Abilities and Perceptual Knowledge](#)’ in Adrian Haddock and Fiona Macpherson (eds.) [Disjunctivism: Perception, Action, Knowledge](#). Oxford: Clarendon Press.

Quine, W. V. (1960), *Word and Object*. Cambridge, MA.: MIT Press.