



PERGAMON

Language & Communication 22 (2002) 281–299

LANGUAGE
&
COMMUNICATION

www.elsevier.com/locate/langcom

The self-regulating mind

Victoria McGeer, Philip Pettit*

*Research School of Social Sciences, Australian National University,
Canberra, ACT 0200, Australia*

Abstract

A system which is to count as minded must represent its environment and must act in pursuit of its goals as those representations make it right to act; or at least it must do this in favourable conditions. But a system may comply with representation-related constraints, simply because that is how it is designed at the subpersonal level to function. Or it may comply for that reason together with the additional fact that it has a special, personal-level ability to represent certain constraints as constraints and to regulate its performance so as to make their satisfaction more likely. This paper argues that the human mind belongs in the second category and then goes on to explore the implications of its being a self-regulating system. © 2002 Published by Elsevier Science Ltd.

Keywords: Language; Mind; Self-regulation; Judgement; Self-knowledge; Normative; Hope; Trust; Self-deception; Development

1. Introduction

We agree with orthodox opinion that there is a sense in which human beings, non-human animals and some artificial systems count equally as minded. But we argue, nevertheless, that human beings are minded in a distinctively self-regulating way. The self-regulation of the human mind shows up in a broad range of activities, driven by a variety of goals. It even appears, so we think, in the process of belief-formation. This is particularly striking, since belief has a mind-to-world direction of fit and is the very paradigm of a state that we might expect to be regulated by the world rather than by the mind.

We try to make a case in this paper for the self-regulating character of the human mind, focussing in particular on belief. The paper is in two parts. In the first we try to show that the fact of employing sentences to express beliefs makes it plausible

* Corresponding author.

E-mail addresses: mcgeer@coombs.anu.edu.au (V. McGeer), philip.pettit@anu.edu.au (P. Pettit).

that human beings should regulate belief-formation, and not form their beliefs in a wholly routinised manner. This part of the paper effectively establishes the existence of two kinds of mind, one routinised, the other partly self-regulating. In the second part of the paper, we go on to look at a variety of phenomena that are unsurprising under the assumption that the human mind is self-regulating in this way but that are quite problematic if the self-regulating capacity is not explicitly recognised. This part of the paper does something, though still only a very little, to illustrate the range and implications of self-regulation in areas beyond belief-formation.

2. Part I. Two kinds of minds

2.1. *The routinised mind*

To be an intentional system, and therefore qualify as ‘minded’ in some minimal sense, is, on standard approaches, to be a system that is well-behaved in representational and related respects. The well-behaved system represents things as they appear within the constraints of its perceptual and cognitive organisation. And it acts in ways that further its desires—presumptively, desires that reflect its overall needs and purposes—in the light of those representations or beliefs (Dennett, 1987). An intentional system may not be perfectly behaved in these action-related and evidence-related ways but it will have to attain a certain threshold of rational performance—and perhaps do so as a result of a certain history or organisation—if it is to seem like it is minded at all.

The approach that is sketched in these remarks allows of many variations but in its essence it is common to most contemporary images of mind, philosophical and psychological. The key idea is that there is no stuff and no substance that is essentially mental—(pace Searle, 1983)—and that whether an organism or artifice is intentionally minded is fixed by whether it conforms to evidence-related and action-related constraints in a satisfactory measure and manner. Conforming to the constraints in question is taken to be consistent with the rule of natural laws—it does not require a breach of the natural, causal order—and one reason for the popularity of the constraint-conforming theory is that it shows us how to be naturalistic about the mental.

We shall be taking the constraint-conforming approach to mindedness as our starting-point in this paper. We do so in a way that abstracts away from detailed commitments, however, and we hope not to beg very many questions. Thus, we do not commit ourselves on how the cognitive architecture of a system’s internal states must be causally organized in order to account for its performance as an intentional system (Jackson, 1993). Nor on whether the states or types of state involved must be recruited to their joint role, evidence-related and action-related, by natural selection or by individual training (Millikan, 1984; Papineau, 1987; Dretske, 1988). Nor on whether it may be sufficient that beliefs are identified by their action-guiding role, and have their represented contents fixed by the causally distinctive way in which they covary with relevant states of affairs (Fodor, 1990). Nor, finally, do we assume anything about what are the feasible limits beyond which we cannot expect

constraint-conformity (Cherniak, 1986), and what are the favourable circumstances under which a lack of conformity would tend to suggest an absence of mindedness (Pettit, 1999).

The most striking feature of the constraint-conforming approach to mind is that it supports the possibility that even a relatively simple, routinised system can be minded. We can readily imagine that, without itself having any sort of control over the process, an animal or robot might be so constituted as to adjust to incoming information in a relatively faithful manner, forming beliefs that determine, in the presence of certain standing or situationally variable desires, what the system does. Its beliefs update in the light of perceptual input, according to the routine of a suitably pre-set design, and they combine with the creature's desires to issue according to a similar routine in this or that action. Or at least they do this within intuitively feasible limits and under intuitively favourable circumstances.

Such a routinised mind, be it animal or robot, will produce actions in an intentional or voluntary way, so far as its desires combine in a suitable manner with its beliefs to produce them (Davidson, 1980). But the system will not act intentionally with regard to how far it conforms to evidence-related and action-related constraints as such; it will not even have beliefs as to what those constraints are or require. Its conformity with the constraints will happen by courtesy of nature or nurture; it will not be intentionally achieved or intentionally reinforced.

With that point in the background, we can outline the plan of discussion in the remaining sections of this first part of our paper. We want to argue that the constraint-conforming theory of mind, no matter in what version, countenances the possibility of wholly routinised intentional systems; but that it also points us towards the possibility of systems whose conformity to some of the required constraints is supported, not just by brute routine, but also by activities of intentional self-regulation. We think that this possibility is important, because it marks the difference between being minded in the human way and being minded in the manner characteristic of animal and artificial systems.

We will begin the discussion by drawing attention to the fact that not only do human beings have beliefs, they also use sentences to express those beliefs. We will then argue that this capacity for linguistically expressing beliefs engenders a capacity to attend to what we believe as such: a capacity, as we call it, for content-attention. After that we will argue that a capacity for content-attention underpins in turn a capacity to identify constraints on rational belief-formation. And then we will see that this capacity for constraint-identification underpins a capacity for constraint-implementation: in effect, a capacity for regulating the process of belief-formation. In distinguishing these capacities for belief-expression, content-attention, constraint-identification and constraint-implementation, we do not suggest that they are separable from one another in practice. The distinctions are meant to serve analytical, clarificatory purposes, not to mark any joints in the reality of things.

We conduct our discussion with two groups of readers in mind. First, those who endorse the constraint-conforming approach to mind but who say little or nothing—and may notice little or nothing—about the distinctively self-regulating way in which human beings come to be minded. And second, those who are emphatic about

the self-regulating aspect of the human mind but who think that this undermines the constraint-conforming approach. We want to stake out a position that is naturalistic enough to appeal to the first group, yet ecumenical enough to appeal to the second.

2.2. *Belief-expression*

When we think of a well-behaved intentional system, we see that the beliefs it reveals in behaviour fulfill a dual role. They are responsive to evidence in a suitable measure: they co-vary with certain ways things are or appear to be, at least under normal conditions, and within feasible limits. And when they are present they are directive of relevant action—they generally dispose agents to act as if things were that way: to act in a manner that would satisfy their desires, other things being equal, if things were that way (Stalnaker, 1984). We would not ascribe a belief that *p* to intentional subjects if they failed to display such dispositions, or at least failed in a way that we could not independently explain. The subject who is quite insensitive to evidence that bears on whether *p* or not *p* can hardly be said to believe that *p*. And any state which comes and goes with evidence that *p* or that not *p*—assuming we could detect such covariation—can hardly be said to constitute a belief that *p* if it has no tendency to cause the subject to act as if it were the case that *p*.

But if beliefs can be said to represent how things are for any intentional system so far as they play this dual role, then we must recognise that creatures like us don't represent how things are just in our beliefs. We also represent them in the sentences that come to our lips and that are issued over our names. Assertoric sentences, when uttered seriously and sincerely, satisfy the evidence-related and action-related constraints that we associate with beliefs. We make such assertions, when we choose to make them, in a way that co-varies with the ways things are presented, at least under favourable conditions of performance. And we generally act as if things are the way we sincerely report them to be. It is because of this co-variational pattern that we can take a sentence-type to be broadly assertoric, signalling that things are a certain way—namely, the way to which the sentence is evidentially responsive. And it is because of the action-connection that we can take the utterance of the sentence to be serious and sincere: so far as the sentence is acted upon, we take the utterance to be a representation that guides what the agent does.

Of course, even though assertions count as equally representational by these criteria, they are very different from beliefs. Used to assertoric purpose, sentences are voluntary signs in Locke's sense: they are employed according to certain conventions and they are uttered, not autonomically, but only when the subject chooses to utter them (Locke, 1975). Beliefs, by contrast, are non-voluntary in so far as they are not in general formed just when the subject wishes. One cannot decide whether or not to believe that *p*—or at least not in any straightforward sense—but one can decide whether or not to say that *p*.

How then do assertions connect with beliefs in the linguistically competent intentional subject? The obvious connection is that people's sincere and serious assertions must generally answer to their beliefs—beliefs that may only be determinately formed under pressure of articulation (Dennett, 1987)—if they are to count as *bona fide*

representations (Shoemaker, 1996, p. 76). It is a person's beliefs that dictate how things are taken as being by that subject, and it is a person's beliefs that dictate how he or she will act linguistically and non-linguistically. Thus sentences will represent how things are by the subject's lights, and they will give voice to assumptions on which the subject can be expected to act in pursuit of any goal, only so far as they are used to say how things are according to the person's beliefs: only so far as they are used to express the beliefs or, more specifically, to express the contents of the beliefs.

How to spell out the difference between a subject that does not or cannot express the contents of certain beliefs in words and the subject that can and does? Both subjects have beliefs *in* those contents, we can say. But whereas the first sort of subject is blind to the contents believed—those contents may be as invisible to it as the air it breathes—the second sort of subject is not blind in this way. The contents of the beliefs are visible to that subject—they exist *for* that subject—so far as they assume material form in the sentences that express them. This means, as we shall now see, that the subject is capable of paying attention to the contents in question.

2.3. *Content-attention*

While any intentional system will be capable of intentional action, at least in principle, the range of intentional activity that is open to it will be limited to the domain in which it is capable of exercising discrimination and attention, and forming beliefs and desires. The extent of that domain may differ dramatically from creature to creature. A dog, quite plausibly, can form beliefs and desires in regard to food, comfort, walks and the like, as well as beliefs and desires involving places, people and other animals; as evidenced through its behaviour, such items are available in perception as matters that it can discriminate and pay attention to. But a dog does not form beliefs about mathematical problems, for example, because there is no way for it to be exposed to them. It is unable to discriminate and attend to mathematical scenarios, unable to form beliefs and desires about them, and so unable to act with a view to affecting them.

As it is quite implausible to think that a dog might be able to discriminate and attend to mathematical problems, so it is equally implausible to think that it might be able to discriminate and attend to the contents of its own beliefs. The dog may have a belief that there is food at the back door. But it is extremely hard to see how it could ever form any beliefs about the content of that belief as such: about the proposition, as we say, that there is food at the back door. There are various things that hold of that proposition: that it is believed by the dog, that it is true or false, that it is unsurprising, and so forth. But we cannot imagine the dog believing those things.

The fact that human beings can use sentences to express their beliefs, however, means that they can overcome this limitation (Pettit, 1993, Chapter 2). They are in a position where they can discriminate and attend to, not just the worldly objects in respect of which dogs may hold beliefs, but also the very contents that they actually or potentially believe. Suppose that people use a sentence 'p' in ordinary discourse to assert that p, and thereby to express the belief that p. To the extent that they can

use the sentence in such a representational way, they can also use it as an exemplar of a representational item. A piece of cloth may be used to exemplify the sort of material of which it is made, enabling people to use a demonstrative ‘that’ to refer to the material in question rather than just to the particular piece of cloth (Goodman, 1969). In parallel fashion, the exemplar usage of an assertion would enable people to talk about the sort of thing asserted, not the particular token displayed: it would enable them to use the demonstrative ‘that’ to refer to the content expressed by the sentence, viz., *p* (Davidson, 1984, Essay 7).

The idea, then, is that people can rely on the exemplar usage of certain sentences to hold up for their inspection the contents expressed in assertoric usage by those sentences. They can use the sentences, not to represent how things are in the manner of direct speech, but to exemplify in the manner of indirect speech the ways things are thereby represented as being. People can make contents or propositions demonstratively salient in the sentences that express them. They can draw them within the scope of their attentive capacity and make it possible to form new beliefs and desires in respect of those contents.

The position adopted in these remarks does not involve any substantive commitment on the relation between assertions and beliefs, language and thought. Like everyone else, we find it philosophically challenging to explain how human beings come to be speakers as well as believers, voluntary as well as non-voluntary representers of their environment. We have nothing to say on that issue in this paper and so we abstract away from the many debates in the literature. We are content to take the phenomenon of speech as given and to try and explore the significance that we think it has for the mode of human mindedness.

Just to illustrate the openness of our position on thought and language for the purposes of this paper, it may be worth mentioning that we need not commit ourselves to any particular developmental story about how the capacity to speak and the capacity to understand intentional states as having content emerge. It may be, as a certain psychologicist theory would imply, that before becoming language-users, intentional systems have to be able to discriminate certain of their own intentional states and have to be able to form beliefs about them, in a way that dogs, presumably, are unable to do (Bennett, 1976). Or it may be, as we ourselves are inclined to think, that people first learned to use sentences as more or less primitive signals, matching them unconsciously to the ways they believed things to be, and that it was only later that they learned to form beliefs concerning their own intentional states (McDowell, 1980). Whether the psychologicist or non-psychologicist story of emergence is sound, the important point is that the availability of language makes it unproblematic to claim that people can discriminate and attend to the contents of their intentional states and so form beliefs and desires about them.

2.4. *Constraint-identification*

The upshot of the argument in the last section is that people’s ability to use sentences explains how they can discriminate contents or propositions as objects of attention. But nothing is ever going to be the object of attention *tout court*; it is

always going to present itself as the bearer of certain properties. So what features are human beings going to identify as proper to the contents that they discriminate? And how are they going to be able to identify those properties?

The property most widely ascribed to contents or propositions is truth or falsity. And there is a straightforward explanation of how human beings might come to attribute this property to a given proposition. People will be in a position to attribute truth to a proposition just so far as they find themselves disposed to assent to it. They will be in a position to attribute falsity to a proposition just so far as they find themselves disposed to dissent from it. And they will be in a position to attribute neither—they will have to reserve judgment—just so far as they find themselves lacking such dispositions (Wright, 1992, Chapter 1). No mystery in any of this. It reflects the disquotational character of the truth-predicate: the fact that for arbitrary sentence 'p', 'p' is true if and only if p.

As it is with truth, so it can be with other, related properties of contents. If I can learn to say and believe that 'p' is true so far as I find myself disposed to say that p, so I can learn on similar bases to say and believe that contents are evidentially well-supported, that they are supportive of this or that action, that they are inconsistent, or that one entails another. I can learn to say and believe that a perception or belief with such and such a content gives support of degree X to the truth of a content 'p', so far as I find myself disposed to that degree, in the presence of the perceptual or belief representation, to say that p. I can learn to say and believe that the truth of 'p' gives support of degree X to action A, in the presence of a certain desire, so far as I am disposed in that degree to A by the fact of believing that p in the presence of that desire. I can learn to say and believe that the truth of 'p' is inconsistent with the truth of 'q' so far as I find myself disposed to say, in no matter what imagined circumstances, that not (p and q). And I can learn to say and believe that the truth of 'p' entails the truth of 'q' so far as I find myself disposed to say, in no matter what circumstances, that not (p and not q).

The fact that people can discern truth and truth-related properties in propositions means that they can identify constraints on what beliefs they should form and act on, as well-behaved intentional systems. For when people form beliefs about which propositions are true or false, which propositions are consistent with one another, which propositions are inductively or deductively supported, and so on, they thereby identify constraints that the well-behaved system must generally be disposed to satisfy. Special explanations aside, the intentionally well-behaved system must tend to believe the true; must tend to believe only consistent contents; must tend to believe contents that are inductively or deductively supported; and so on. In having the ability to attend to contents or propositions, therefore, human beings will also have the ability to identify constraints on how they should perform, if they are to be well-behaved intentional systems.

2.5. *Constraint-implementation*

Some will think it possible that certain creatures might identify such constraints, in the sense of having the corresponding beliefs, without identifying them as constraints.

They might fail to see that the constraints are specifications that they must fulfill on pain of losing or diminishing their title to be regarded as intentional subjects (Dennett, 1987). Could such creatures still be affected in this case by the presence of the beliefs that identify the constraints? The orthodox constraint-conforming view of intentionality would say ‘yes’.

Those in the orthodox camp who expect any intentionally well-behaved system to satisfy certain constraints in the beliefs it forms will expect the system whose beliefs include truth-related beliefs to satisfy such constraints in its expanded belief-set. If it believes that *p*, that the truth of ‘*p*’ entails the truth of ‘*q*’, and that not *q*, for example, they will expect it to tend, at least under favourable circumstances, to resolve this sort of inconsistency by giving up one or other of these beliefs. More generally, they will expect its truth-related beliefs to interact with its other beliefs and to be equally subject to the regime—the wholly routinised regime, as it might be—whereby a reasonable level of intentional performance is achieved. They will not expect its truth-related beliefs to have a marginalised, epiphenomenal status.

Whatever holds with the unselfconscious creatures imagined—and they may not even be coherently imaginable—it is clear that human beings are not of their ilk. Human beings ascribe beliefs and desires to themselves—they recognise what propositions they believe and desire—and in doing this they see themselves as intentional systems. Thus they are in a position to identify the constraints associated with the truth-related properties of propositions as constraints on their performance. And so the question which we must ask is whether there is anything they can do about implementing the constraints they identify, over and above responding passively to them? Are there opportunities for human beings to take intentional initiatives that promise to promote their satisfaction of truth-related constraints?

There are opportunities for such initiatives in two broad categories. First of all, there are various things that human beings can intentionally do in order to give themselves a richer exposure to constraints, and presumably therefore a better chance of having true beliefs and securing their desired ends. And secondly, there are various things that they can do to enhance their own responsiveness to those constraints: to ensure that they apply the constraints more faithfully in the formation of their beliefs.

People can give themselves a richer exposure to constraints in any of a number of ways. They can think about what follows from their existing beliefs, taking time to work out what those beliefs commit them to. They can take care about determining what those beliefs support in the presence of this or that body of putative evidence. Or, of course, they can go and seek after new evidence: they can go into the field, or into the library, in pursuit of new data. They can make intentional efforts to ensure that their beliefs are formed on the basis of an exposure to all the constraints that become visible in the light of patient reflection and informed research. This activity is distinctive of human beings and it represents a familiar way in which they transcend the passive, routinised operation of representational constraints.

But not only can people make intentional efforts to secure a richer exposure to relevant constraints. They can also make efforts designed to ensure that they respond adequately to those constraints. These efforts come in two different forms,

depending on whether the agent is spontaneously and reliably disposed to form appropriate beliefs in response to the constraints in question.

In many cases, people will be suitably disposed to form an appropriate belief in response to the recognition of what a given constraint requires. I see that I am committed to believing that *p* by the truth of beliefs I do not currently question and so I effortlessly come to form that belief. I see that the inductive evidence supports the truth of a certain content, ‘*q*’, and I immediately come to form a belief in that content. I rely on such automatic responsiveness in most forms of reasoning, *a priori* or empirical. It may be intentional on my part that I reason—that I seek out relevant considerations and pay attention to what they require in the way of belief—but I rely on myself to make the required inference, and draw the appropriate conclusion, in a completely automatic way (Burge, 1996, 1998).

Even when belief-forming habits are spontaneous and generally reliable, however, there are various intentional efforts I can make to increase the chance that they do not lead me astray: that is, to enhance my responsiveness to the constraints in question. I can rehearse carefully the inference that comes naturally to me, studying each step in the transition and guarding against malfunction. I can try to check the habit of inference that takes me spontaneously to the conclusion by seeing where it would lead in other, parallel cases. More generally, I can intentionally monitor and regulate the inferential transitions that come spontaneously to me.

There are other cases, however, where there is room for a more dramatic form of intentional monitoring and regulation. While reasoning leads me to a certain conclusion in these cases, that conclusion does not automatically come to reign in my responses; I do not spontaneously come to act as if it were true. On the contrary, I find myself stuck with a pattern of action that evinces the continuing presence in the architecture of my representations of a different, inconsistent belief. In these cases, people can adopt intentional initiatives to counteract such a resistance to reason and to promote in quite a vivid fashion their satisfaction of relevant constraints.

Consider the principle that identifies the mistake in the so-called gambler’s fallacy: the fallacy of thinking that with independent events such as a coin’s coming up heads or tails, a string of heads increases the likelihood that it will next come up tails, and vice versa. It may be that I sustain that belief, in good part, only because I take the principle to be true and make intentional efforts to stay faithful to it: to stop backsliding, for example, into the sort of belief that may mesmerise me at the gambling table, to the effect that tails is bound to come up next, given the run of heads. I may have an independent inclination to fall prey to the gambler’s fallacy, in other words, and it may only be by dint of sustained effort that I manage to stay faithful to the principle I take to be true. I may continue to display a belief in the principle only so far as I exercise intentional self-control in relation to my belief-forming inclinations (Mele, 1995).

Or, to go to another example, consider the representations of their orientation that pilots come to believe when they try to fly under instrument conditions, i.e. out of visual contact with the horizon. It is impossible for pilots to form and unform reliable beliefs on the usual basis of proprioceptive cues. The reason is that gravity and acceleration are fundamentally equivalent from the point of view of a force

sensor such as the human body, so that once pilots lose sight of the horizon, trying to fly 'by the seat of their pants' will cause them to lose control. Thus pilots have to learn to form and unform their beliefs about pitch, altitude, even up-down orientation, in an intentionally sustained manner. They have to recognise that with any representation of such details, the basis on which to determine whether it is believable will be given by the instrument panel, not by proprioceptive cues, and that they had better follow that evidence and suppress the independent inclination to go by proprioceptive cues. Recognising this, they may come to fly by the instrument panel in a more or less habitual or effortless manner; or they may succeed in flying by the instrument panel only by dint of a continuing effort: this, presumably, is how it will be with beginner pilots.

In both of these examples, the subjects are resistant to the dictate of truth-related considerations—if you like, the dictate of reason. In the one case, that is because of attachment to a fallacy; in the other, because of a natural inclination to go by proprioceptive cues. The subjects succeed in being governed by reason, and in forming their beliefs after an appropriate pattern, only by grace of an intentional effort to school themselves into going by the book. They intentionally keep themselves mindful of what the book teaches, and this serves to channel their responses into suitable grooves.

2.6. *Conclusion*

The story told in the preceding sections documents the distinctive capacity of human minds to conform to the constraints associated with being an intentional system. Such a system must be intentionally well-behaved in a suitable measure and manner, according to orthodox theory, but the routinised, more or less mechanical basis on which this is achieved for simple systems is not the only basis available. Given that human beings can express some beliefs in words, it becomes overwhelmingly plausible that they can pay attention to contents as such, identify truth-related constraints on representational performance, and do things intentionally with a view to better implementing such constraints.

Embracing the view that human beings can do this amounts, in our terms, to concluding that they are self-regulating and not just routinised minds. But stated like that, the conclusion is in one way too downbeat and in another too upbeat. Some corrective comments, therefore, are required.

The conclusion is too downbeat in the sense that it may suggest, wrongly, that the self-regulation of the human mind is limited to regulation that is designed to promote satisfaction of truth-related constraints. Consider the constraints on theory-selection that are often taken to govern the pursuit of science: constraints to do with economy and simplicity, generality, precision and refutability. As human beings can be expected to be able to regulate themselves for the satisfaction of truth-related constraints, so they can be expected to be able to regulate themselves for other epistemic constraints of these kinds. And as they can regulate themselves for the satisfaction of epistemic constraints, so there is every reason to think that they will be able to regulate themselves for the satisfaction of non-epistemic constraints or

desiderata as well; we will be discussing an example of this sort of regulation in the next part.

The conclusion is too upbeat in a number of respects, for it fails to register various limitations on the self-regulating capacity of the human mind. A first limitation is that human subjects can only regulate themselves in respect of beliefs that they can express in language and consciously acknowledge as their own: as beliefs that bear their first-person signature (Burge, 1996, 1998). We say a little more on the appearance of the first person in the next part. Here we simply note that there will always be contingent limitations on the linguistic expression of contents, and so a limit on how far individuals can achieve regulative control of their own minds. It is possible, of course, for human beings to expand their expressive resources and to articulate the contents of beliefs that have hitherto remained unspoken. A formal example of this occurs when logicians articulate rules of inference to which they were already committed as rational beings. And other examples will be provided by people's attempts to clarify the commitments implicit in the less formal ways in which they reason (Brandt, 1994; Pettit, 1998).

A second, less plastic limitation is that while human beings may exercise their self-regulating capacity in relation to a belief in any expressible content, they cannot do so in relation to all their expressible beliefs taken at once. The intentional attempt to promote the satisfaction of truth-related constraints in respect of 'p' presupposes the presence of other beliefs, say beliefs to the effect that 'p' is well-supported, or entails 'q', or whatever. And so it will not be possible to try to satisfy such constraints in respect of the presupposed beliefs at the same time that one is trying to be rational in respect of the original one. Not all beliefs can be taken at once to the tribunal of active examination; in the examination of any particular beliefs—for example in checking what they entail or are entailed by (Carroll, 1895)—the soundness of others must be taken for granted.

A third limitation is that the self-regulating mind may not pursue its regulatory activity to its fullest potential. That human beings are self-regulating means only that they have the capacity to regulate their representational performance, not that they regulate it as a matter of unavoidable necessity. Thus people may be too lazy to exercise the capacity for self-regulation as fully as they might. Or they may be motivated to avoid regulation, if the consequences—say, the consequences of imposing consistency—are too painful to face.

3. Part II. The utility of the distinction

We hold that distinguishing between self-regulating and routinised forms of mind throws light on a variety of issues and we would like to illustrate that claim in this second part of our paper. With the distinction in hand between two kinds of mind, we can see the place that judgment has in human psychology; we can give a good account of self-knowledge; we can see the sense in which belief among humans involves a normative commitment, as certain minority views emphasise; we can make sense of people's capacity to liberate themselves from theoretical rationality,

as in self-deception or in hope or trust; and we can understand the developmental role of folk psychology. It is not possible to vindicate these claims in full and proper measure here but we can at least point to the considerations that support them.

3.1. *The judgmental mind*

Many philosophers, following ordinary usage, want to mark a distinction between two forms of belief-like phenomena: judgment or opinion, on the one hand, and credence, on the other (de Sousa, 1971; Dennett, 1979; Rey, 1988; Engel, 1995). But there is no settled view as to how these relate to one another and as to why both are necessary. Decision theory is often taken as a regimentation of folk psychology, for example, but it only countenances credences or degrees of confidence; it leaves no room for distinctively judgmental episodes.

Judgment, as commonly conceptualised, has three salient features that distinguish it from credence. First, it is an event, not a state. Second, it doesn't come in degrees of confidence: a person may judge that it is probable to this or that degree that *p*, but the judging itself does not come in degrees; it is an on–off matter. And third, the fact that someone judges that *p* or that it is probable that *p* requires them to form the corresponding credence, on pain of their sincerity coming into question, but not the other way around; there is no judgment without credence—at least in intuitively favourable circumstances—but there is certainly credence without judgment.

If the human mind is wholly routinised or mechanical, then it is hard to see what role there is for judgment to play. Credence will exhaust the representational roles available and there is no other role that appears to fit judgment. A judgment cannot be identified, for example, with the formation of a credence, since it won't then require the formation of credence in the appropriate way. And neither can judgment be identified with anything that precedes credence and serves as input to it, since this role is already played by perception, proprioception and the like.

Once we see that the human mind is self-regulating, however, then there is an obvious part for judgment to play. Judgment can be identified with a sort of event that is distinct both from perception and from the formation of credence. The judgment that *p* can be taken to be the person's act of registering relevant truth-related constraints as requiring that he or she believe that *p*, where the registering involves not just seeing that the constraints require that belief but feeling their force: being disposed to adjust or act in response to them (Burge, 1998).

If judgment is this active registering of constraints or reasons as requiring a certain belief, then we can see why it has the distinguishing features mentioned. It will be an event, like an act of perception, not a state. It will be an event in which uncertainties are marked in the content, not in the degree of confidence with which it occurs: judgment will typically register the degree of support enjoyed by a given representation or a prospective action. And, finally, it will have one-way implications for the credences—and, where relevant, the actions—that the subject is required to form.

Will judgment still deserve to be as closely associated with belief as it is in normal discourse? Yes, because just as seeing and perceiving is typically believing, so judging

will typically be believing too (Dennett, 1984, p.40). It will give rise under favourable conditions to the sort of credence that is naturally identified with belief; a failure to do so will require special explanation. Judgment may not constitute a form of belief under this story, but it is intimately related to belief. It is the sort of event that puts belief in place when the human mind exercises its self-regulative capacity.

3.2. *The self-knowing mind*

One of the recurrent puzzles in the philosophy of mind bears on how people can tell what they believe. Not how they can tell infallibly what they believe, since they are not infallible on the matter, but how they can tell what they believe with ease and assurance. If believing something typically involves a disposition to act as if it were true, and so behave in an open-ended variety of ways, how can I learn that I believe the matter in question? How can I tell straight off, as it often seems I can, that I instantiate the appropriate pattern of disposition?

The problems raised by self-knowledge of one's beliefs become relatively untroubling on the view adopted here. For, under the regime of self-regulating intentionality, a person who wonders whether or not they believe that *p*—wonders if they can rightly avow the belief that *p*—need only ask after the evidence that they see for and against the truth of '*p*'. The regime ensures the reliability of a judgmental heuristic for determining whether or not one believes that *p*: viz., see if the evidence available supports the truth of '*p*'.

Why does the self-regulating regime make this heuristic reliable? In a word, because the regime supposes that in any case where a person judges that *p*, they will come at least under intuitively favourable circumstances to believe that *p*. The presence of that judgment may ensure the presence of the belief that *p* in a more or less effortless, though still checkable way. Or, failing that, it may ensure the presence of the belief with the help of the self-monitoring and self-discipline illustrated earlier in the gambler's fallacy case. For any case of judging that *p*—for any case of seeing the evidence as supporting the truth of '*p*'—the person can rely on their own self-regulative capacity to be sure that they prove to believe that *p*: to ensure that they display the disposition to act as if it were true.

The emerging picture of one's knowledge of what one believes is essentially the agency model of self-knowledge developed elsewhere by one of the authors (McGeer, 1996; Moran, 1997; Bilgrami, 1998). The idea is that self-knowledge of one's beliefs comes about, not as a result of having a special insight into whether one has the dispositions that make one a believer that *p* or that *q*, but as a result of having a special ability to develop the dispositions that make one a believer that *p* or that *q*. This is not the ability to make oneself, depending on what one's will dictates, a believer that *p* or that *q*; otherwise it would be just a capacity for wishful thinking. Rather it is the ability to make oneself, depending on what one's judgment dictates, a believer that *p* or that *q*. It is the ability to believe along the pattern that judgment sets out, whether by grace of one's being automatically attuned to respond in that way or, failing such attunement, by grace of a capacity to keep oneself in line through intentional self-monitoring.

3.3. *The committed mind*

Not only does the distinction between self-regulating and routinised forms of mentality enable us to make a place for judgment, and to make good sense of self-knowledge. It also enables us to give countenance to a theme that is much emphasised in the heterodox tradition that rejects the constraint-conforming theory of mind. The theme is that holding by a belief is not a matter of having a disposition but rather something closer to adopting a normative commitment (Levi, 1980; Davidson, 1984; Brandom, 1994; McDowell, 1996; Bilgrami, 1998). Holding by a belief, so the line goes, involves adopting a normative stance and the presence of such a stance will not be guaranteed just by the fact that a subject meets certain naturalistic constraints. In response to this argument, what we want to point out is that, by a set of observations rehearsed below, believing something will amount to a sort of normative commitment when the believer is a self-regulating mind of the characteristically human kind. We hope that once this is recognised, it may become plausible to countenance the two sorts of mind distinguished earlier and to see the normative stance as a feature only of self-regulating believers.

First observation. The fact that self-regulating minds can try to ensure that their beliefs meet the constraints of rationality means that they can be held responsible—can be praised or blamed—for what they believe in a way in which it would make no sense to hold passively minded believers responsible. If a self-regulating person believes that *p* and that *q*, for example, and their doing so is ruled out by their having previously registered that the truth of '*p*' is inconsistent with the truth of '*q*', then assuming that no special excuses apply, we can sensibly blame them for believing that *p* and *q* (Pettit and Smith, 1996). The intentional effort whereby the inconsistency of their believing such things would have become salient to them, and would have led them to revise their belief, will be generally just as accessible as it is in the case of actions for which we routinely hold ourselves and others to be responsible (Wallace, 1996). It will be the effort involved in trying to form a conscientious judgment on the matter.

Second observation. But if self-regulating believers can be held responsible in this way, then that is going to be itself a matter of common belief: everyone will be in a position to recognise that it is so, to recognise that everyone recognises that it is so, and so on (Lewis, 1969). Thus, not only will these subjects be held responsible for what they believe, they will expect to be held responsible for what they believe; others will expect them to expect to be held responsible; they will expect others to expect this; and so on.

Third observation. Let it be granted, then, that self-regulating believers can be held responsible, and this as a matter of common knowledge, for what they believe. That means that they must be taken to assume responsibility for various of the things they hold and that holding by those things will have a characteristically normative aspect. Suppose that a self-regulating believer is given the chance to consider whether or not to maintain a certain belief, and does then maintain it. Given that this is done in full knowledge, shared in common with others, that they will be held responsible for the belief, the person can be said to assume responsibility for what they believe. They

assume responsibility for the belief in the sense that they must take themselves to be ready and able to meet whatever critiques or challenges are made against it.

Fourth observation. But there is also a further point to add. Consider the things that a self-regulating believer holds which have been taken for granted from their background or training and have never been explicitly examined. Consider in particular those things that they believe in this way that can be expressed in the publicly shared language and that can be challenged by others. Are we to take it that the believer assumes responsibility for such beliefs also? Well, it is certainly plausible that the conventions of responsibility work along lines of strict liability, so that even for beliefs that a person has not actively examined, it is perfectly proper for others to apply praise or blame; here as elsewhere, a regime of strict liability would have the effect of enhancing the self-regulatory performance of those who live under it. Under such a regime it will follow that every expressible belief that a person holds is held in such a way that they must be said to assume responsibility for it. Maintaining any such belief will amount to more than just seeing things a certain way: being disposed to act as if things were that way. It will have the profile of taking up a position, and being ready to defend it against attack. It will present itself, precisely, as a normative commitment.

3.4. *The liberated mind*

It is a commonplace of human psychology that we frequently deceive ourselves in a manner that parallels the way we deceive others and that we sometimes adopt attitudes of hope and trust that go beyond the belief-related reasons available to us. In these sorts of cases, though in very different ways, we transcend our rationality in the sense in which rationality involves satisfying truth-related constraints. We persuade ourselves to believe contrary to the palpable evidence, as in self-deception. Or we get ourselves to ignore the evidence, as in certain cases of hope and trust, and to act as if something that is not well supported could be the case.

This sort of liberation is almost impossible to make sense of on the part of a passive, routinised mind. Such a mind is the slave of its beliefs and cannot help but act in accord with them. Thus there is no room for the possibility of its setting aside its beliefs—suspending the rational pressures that produce those beliefs—in favour of acting on some other assumptions. But that is not so when the mind in question is self-regulating in character. The mind that is capable of trying to be theoretically rational—trying to satisfy truth-related constraints—is a mind that is also going to be capable of suspending rationality.

Consider self-deception. As the self-regulating mind can intentionally seek to discover what truth-related constraints require, so a desire to maintain certain beliefs can lead it intentionally not to pursue such questions. And as the self-regulating mind can take the deliverances of those demands at face value, moving to form the required beliefs, so a desire to maintain certain beliefs can lead it to try to explain them away and to silence the effect that they would otherwise have. Thus the self-regulating mind can do things intentionally with a view to maintaining certain beliefs, if not with a view to maintaining them *qua* false beliefs. And so it can mimic

in regard to itself something close to the deception—the intentional deception (Mele, 1987)—that it can practise in regard to others. This deception will set aside theoretical rationality but it may be motivated in a practically rational way by the greater degree of desire-satisfaction that it procures for the subject.

Consider also hope or trust. The hopeful or trusting person has a desire to act as if it were the case that a certain hoped-for prospect is possible or even likely, or to act as if it were the case that a particular person is trustworthy in regard to a certain action. And typically, they have a desire to do this that is robust enough to prevail even in the admittedly extreme case where belief fails them. The hope-related desire may be grounded in the fact that unless the person treats the hoped-for prospect appropriately, they will give up all effort and fall into a despairing malaise; the trust-related one in the fact that unless they treat the trustee as trustworthy then they can never expect to form a respectful relationship with that person. The desires may be rational in practical terms—in terms of the ends sought—even though they motivate a neglect of theoretical rationality: a neglect of truth-related constraints.

The reason self-regulating subjects can act out of hope or trust in these cases is that, just as they have the capacity to regulate themselves according to the demands of theoretical rationality—as in the gambler's fallacy case considered earlier—so they have the capacity to regulate themselves according to other, practical demands, and contrary to the demands of theoretical rationality. Recognising the importance of what is at stake—the avoidance of a despairing malaise or the enjoyment of a respectful relationship with the trustee—these subjects quite overtly decide to act against what they are inclined to believe: to act as if the hopeful or trusting assumption is sound. Everything is above board here, by contrast with the case of self-deception. But in both sorts of cases people practise a suspension of theoretical rationality of a kind that only a self-regulating mind could display.

3.5. *The trained mind*

We are happy to go along with common wisdom in acknowledging that there is a sort of folk psychology that we all share. This consists, at its core, in our shared set of assumptions as to what beliefs and desires and other such states or frames of mind are and how they relate to behaviour. We the folk do not endorse those assumptions, of course, in the fashion of theoretical tenets but rather in the manner in which we countenance rules of inference—say, *modus ponens*—that we may not be able to spell out in formulae (Pettit, 1998).

The standard account of folk psychology depicts it as a purely descriptive theory that is used, first, to identify intentional or minded systems and, second, to explain and predict the responses of systems that are assumed to be intentional or minded. That account is the only one available, indeed, under the assumption that minds are all simple, routinised systems. For what will there be for folk psychology to do under that assumption, beyond serving such explanatory and predictive functions?

Once we recognise that human minds are self-regulating, however, then a very different role comes into view. We can imagine that folk psychology serves not just to map the way minds function, but to train young minds in the functioning that

makes such mapping possible. Self-regulating minds can be trained in the constraints by which they are to regulate themselves. And folk psychology will be uniquely equipped to provide the catechism on which to base such training (McGeer, 1996, In preparation).

The prediction supported by this observation is that folk psychology will actually be used in a variety of ways that have a training, but not a simple explanatory or predictive, rationale. And the prediction is borne out. We ordinary folk use our shared psychological theory to teach our children what is expected of them as intentional subjects and to train them to live up to those expectations. We use it to point children towards the patterns of response that are expected of them, and to reinforce them in those patterns.

Consider the well-known phenomenon of ‘parental scaffolding’ or the over-interpretation of children’s ‘intentional’ behaviour (Kaye, 1982; Bruner, 1983; Dennett, 1996). Little Susie happens to put her toy rabbit in bed and we tell her that that’s right, she should look after the rabbit and keep it warm, because it’s her friend. Or she unthinkingly gives her baby brother a sloppy kiss on the forehead and we tell her what a good girl she is to love him like that and to want to show him that she loves him. Or she parrots a teacher’s claim that New York is a big city and, perhaps correcting other claims she makes about New York, we tell her how clever she is to know that and try to get her to understand the claim by treating her as if she had made it with full understanding: for example, by adding that she is really clever to know that New York has so many more people living there than in her home town.

Under the standard account, this use of folk psychology is going to seem anomalous and slightly preposterous; it will look like the sort of game we play when we talk to our pets or even our plants. But the employment of folk psychology after such a pattern becomes quite intelligible, once it is recognised that human beings have self-regulating minds that require training. It is no surprise that adults use folk psychology to hold out patterns to which they wish to attract their children. What better way to train self-regulating minds than by rewarding and reinforcing their approximation to the patterns in which they ought to regulate themselves?

4. Conclusion

We saw in the first part of this paper that a series of plausible observations gives us reason to acknowledge that intentionality or mindedness comes in two modes, self-regulating and routinised. We have seen in the second part of the paper that if we do acknowledge this distinction, then a number of phenomena fall naturally into place. We can see the role of judgment as distinct from credence. We can make sense of the ready way in which we often know our own intentional states. We can understand why belief in humans often has the aspect of a normative commitment. We can make good sense of the suspension of rationality involved in self-deception, hope and trust. And we can explain why we use our folk psychology to a developmental or training purpose. This is not an insignificant list of benefits and we hope that it may convince people who hold otherwise quite distinct philosophies of mind

that they should agree in giving countenance to the distinction between the self-regulating and the routinised mind.

Acknowledgements

We were greatly helped in preparing this paper by comments on earlier drafts by Dan Dennett, Jennifer Hornsby, Cynthia Macdonald, Al Mele, Mike Martin, Bojana Mladenovic, Denis Robinson and, in particularly helpful detail, Martin Davies. We are also grateful for comments received when the paper was presented at the Italian Academy, New York University and Trinity College, Dublin and for guidance on the complexities of flying from Dave MacKeen and Tad McGeer.

References

- Bennett, J., 1976. *Linguistic Behaviour*. Cambridge University Press, Cambridge.
- Bilgrami, A., 1998. Self-knowledge and resentment. In: Wright, B.S.C., Macdonald, C. (Eds.), *Knowing Our Own Minds*. Oxford University Press, Oxford, pp. 207–241.
- Brandom, R., 1994. *Making it Explicit*. Harvard University Press, Cambridge, MA.
- Bruner, J., 1983. *Child's Talk: Learning to Use Language*. Norton, New York.
- Burge, T., 1996. Our entitlement to self-knowledge. *Proceedings of the Aristotelian Society Suppl.* Vol. 96, 91–116.
- Burge, T., 1998. Reason and the first person. In: Wright, B.S.C., Macdonald, C. (Eds.), *Knowing Our Own Minds*. Oxford University Press, Oxford.
- Carroll, L., 1895. What the tortoise said to achilles. *Mind* 4, 278–280.
- Cherniak, C., 1986. *Minimal Rationality*. MIT Press, Cambridge, MA.
- Davidson, D., 1980. *Essays on Actions and Events*. Oxford University Press, Oxford.
- Davidson, D., 1984. *Inquiries into Truth & Interpretation*. Oxford University Press, Oxford.
- de Sousa, R., 1971. How to give a piece of your mind; or, a logic of belief and assent. *Review of Metaphysics* 25, 52–79.
- Dennett, D., 1979. *Brainstorms*. Harvester Press, Brighton.
- Dennett, D., 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. MIT Press, Cambridge, MA.
- Dennett, D., 1987. *The Intentional Stance*. MIT Press, Cambridge, MA.
- Dennett, D., 1996. *Kinds of Minds: Towards an Understanding of Consciousness*. Weidenfel and Nicolson, London.
- Dretske, F., 1988. *Explaining Behavior*. MIT Press, Cambridge, MA.
- Engel, P., 1995. *Believing, Holding True, and Accepting*. CREA, Paris.
- Fodor, J., 1990. *A Theory of Content*. MIT Press, Cambridge, MA.
- Goodman, N., 1969. *Languages of Art*. Oxford University Press, London.
- Jackson, F., 1993. Block's challenge. In: Bacon, J., Campbell, K., Rhinehart, L. (Eds.), *Ontology, Causality, and Mind: Essays in Honour of David Armstrong*. Cambridge University Press, Cambridge, pp. 235–248.
- Kaye, K., 1982. *The Mental and Social Life of Babies: How Parents Create Persons*. Chicago University Press, Chicago.
- Levi, I., 1980. *The Enterprise of Knowledge*. MIT Press, Cambridge, MA.
- Lewis, D., 1969. *Convention*. Harvard University Press, Cambridge, MA.
- Locke, J., 1975. *An Essay Concerning Human Understanding*. Oxford University Press, Oxford.
- McDowell, J., 1980. Meaning, Communication, and Knowledge. In: Straaten, Z. (Ed.), *Philosophical Subjects*. Oxford University Press, Oxford, pp. 117–139.

- McDowell, J., 1996. *Mind and World*. Harvard University Press, Cambridge, MA.
- McGeer, V., 1996. Is self-knowledge an empirical problem? Renegotiating the space of philosophical explanation. *Journal of Philosophy* 93, 483–515.
- McGeer, V., *Mind in Time: the Developmental Character of Human Intentionality*. MIT Press, Cambridge, MA. In preparation.
- Mele, A., 1987. *Irrationality*. Oxford University Press, Oxford.
- Mele, A., 1995. *Autonomous Agents*. Oxford University Press, Oxford.
- Millikan, R., 1984. *Language, Thought and Other Biological Categories*. MIT Press, Cambridge, MA.
- Moran, R., 1997. Self-knowledge: discovery, resolution, and undoing. *European Journal of Philosophy* 5, 141–161.
- Papineau, D., 1987. *Reality and Representation*. Blackwell, Oxford.
- Pettit, P., 1993. *The Common Mind: An Essay on Psychology, Society and Politics*, paperback edition 1996. Oxford University Press, New York.
- Pettit, P., 1998. Practical belief and philosophical theory. *Australasian Journal of Philosophy* 76.
- Pettit, P., 1999. A theory of normal and ideal conditions. *Philosophical Studies* 96, 21–44.
- Pettit, P., Smith, M., 1996. Freedom in belief and desire. *Journal of Philosophy* 93, 429–449.
- Rey, G., 1988. Towards a computational account of akrasia and self-deception. In: McLaughlin, B., Rorty, A. (Eds.), *Perspectives on Self-deception*. University of California Press, Berkeley, CA, pp. 264–296.
- Searle, J.R., 1983. *Intentionality*. Cambridge University Press, Cambridge.
- Shoemaker, S., 1996. *The First-Person Perspective and Other Essays*. Cambridge University Press, Cambridge.
- Stalnaker, R.C., 1984. *Inquiry*. MIT Press, Cambridge MA.
- Wallace, R. J., 1996. *Responsibility and the Moral Sentiments*. Harvard University Press, Cambridge, MA.
- Wright, C., 1992. *Truth and Objectivity*. Harvard University Press, Cambridge, MA.