

Direction of Fit

I. L. HUMBERSTONE

1. Three quotations, by way of introduction

In her seminal presentation of the distinction between what have since come widely to be called two “directions of fit”, Anscombe described a man going shopping with a shopping list while being tailed by a private detective listing the man’s purchases, and asked what distinguishes the shopping list from the detective’s list. She answered the question thus:

It is precisely this: if the list and the things that the man actually buys do not agree, and if this and this alone constitutes a *mistake*, then the mistake is not in the list but in the man’s performance (...) whereas if the detective’s record and what the man actually buys do not agree, then the mistake is in the record. (Anscombe, 1957, p. 56)

The “direction of fit” terminology actually antedates Anscombe’s monograph, as we’ll see in a moment, though not for marking quite the distinction to which she drew attention; for a nice example of its use in this capacity, we may quote from Mark Platts

The distinction is in terms of the *direction of fit* of mental states to the world. Beliefs aim at being true, and their being true is their fitting the world; falsity is a decisive failing in a belief, and false beliefs should be discarded; beliefs should be changed to fit with the world, not vice versa. Desires aim at realization, and their realization is the world fitting with them; the fact that the indicative content of a desire is not realised in the world is not yet a failing *in the desire*, and not yet any reason to discard the desire; the world, crudely, should be changed to fit with our desires, not vice versa. (Platts, 1979, p. 257)

Having himself quoted the above passage from Platts, Michael Smith adds the following further reflections:

For the difference between beliefs and desires in terms of direction of fit comes down to a difference between the counterfactual dependence of a belief and a desire that *p*, on a perception that *not p*: roughly, a belief that *p* is a state that tends to go out of existence in the presence of a perception that *not p*, whereas a desire that *p* is a state that tends to endure, disposing a subject in that state to bring it about that *p*. Thus, we may say, attributions of beliefs and desires require that different *kinds* of counterfactuals are true of the subjects to whom they are attributed. We may say that this is what a difference in their direction of fit *is*. (Smith, 1987, p. 54)

These quotations help us in their different ways to light upon a single and apparently significant distinction, which for the moment we may take to be understood and indeed familiar. In the following section, its significance will be illustrated by examples of its application. After that, we will return to the passages quoted above and assess their characterizations of the distinction (§3), as well as commenting on some other suggestions. In §4 we present tentatively a positive proposal, distilled out of criticisms of these suggestions, with the aid of an idea taken from J. O. Urmson. The interest throughout is on saying, clearly and non-metaphorically, what direction of fit is, and in what difference in respect of direction of fit consists.

It will be useful to have some terminology to mark the distinction. Searle (1979) applies the distinction to effect a taxonomy of speech acts. He distinguishes the two directions of fit, for this application, as the *words-to-world* direction (statements, predictions, etc.) and the *world-to-words* direction (commands, promises, etc.).¹ Since we will be interested mainly in the distinction as it applies to propositional attitudes, or mental states, rather than to (putative) expressions thereof, this terminology is too specifically linguistic. Obviously one could reach for some familiar dichotomy and press it into service—for example, *theoretical* vs. *practical*, or *cognitive* vs. *conative*—but each such contrast comes with its own potentially distracting associations, and it seems preferable to start afresh. Accordingly my proposed terminology will distinguish the *thetic* and *telic* directions of fit, as generalizations of the linguistic words-to-world and world-to-words directions, respectively.²

It was, incidentally, *à propos* of the linguistic arena that the phrase “direction of fit” was first used: namely in Austin’s study (1953) of speech acts involving predication. His use of a contrast in respect of direction of fit was quite different from those described as linguistic applications above, since it pertained to putatively fact-stating speech acts, all of them, in our terms, having the thetic direction of fit.³ The use of the direction-of-fit terminology to mark the present, telic/

¹ Compare also the direction-of-fit based explication of the subjunctive/indicative contrast in James (1986); subjunctivity will make a brief appearance below in §4.

² In Searle (1983) the appellations “world-to-mind” and “mind-to-world” are used to mark the present, non-linguistic, version of the distinction. “Thetic” and “telic” are preferred here partly on grounds of brevity; these adjectives may be found, incidentally, in the OED, with senses not rendering too inappropriate their proposed technical usage here (though my own preference would be for a pronunciation, in both cases, with “e” as in “be” rather than—the OED’s recommendation—as in “bed”). Actually, brevity aside, I have another reason for not following Searle’s terminological proposals, namely a difficulty in remembering always that “mind-to-world” abbreviates “mind-to-fit-world”, rather than alluding to the characteristic direction of causation “from-mind-to-world”, which is how I naturally interpret the phrase, but which picks out typically (though not invariably—see §3.3) precisely the reverse direction of fit—as noted in Searle (1979, pp. 97, 122).

³ Indeed, the present distinction between directions of fit is as close to Austin’s distinction in respect of what he calls “onus of match”, which distinction he contrasts with the distinction in respect of (what he calls) direction of fit. (In fact it is grasping the distinction between these distinctions which makes Austin’s paper so challenging.)

thetic distinction, as illustrated by our second and third opening quotations, seems to have gained currency in the first ten years after Anscombe's monograph appeared, sufficiently so that by 1966 Bernard Williams was writing of "the line between discourse which (to use a now familiar formula) has to fit the world, and discourse which the world has to fit" (Williams 1966). As already remarked, our interest will not be so much in types of discourse as in the typology of the attitudes such discourse expresses.

2. Applications

Let us consider—by way of reminder—some ways in which the thetic/telic distinction comes to our attention. The examples mentioned are drawn from these areas: the theory of motivation, the ethics of belief, the analysis of knowledge, and the doctrine of double effect.

2.1 Theory of Motivation.

Platts and Smith were addressing, in the discussions from which the quotations in §1 were drawn, the question of whether the mere possession of a belief could motivate its possessor to action. A negative answer to this question forms the core of Smith's "Humean" theory of motivation, and his main argument for that answer turns on the observation that to be motivated to procure this or that outcome is to have a certain *goal*, which is precisely, in our terminology, to be in a state with the telic direction of fit. Thus the only way the mere holding of a belief could motivate its holder to action would be if that state also had a thetic direction of fit. But Smith argues that no state could have both directions of fit, in view of the difference between the characteristic features of the two directions (as outlined in the quotation in §1).

It would not be appropriate here to enter the debate between those who, like Smith, support (what he calls) the Humean view and those who, like Platts, have opposed it. Our main concern is after all to draw attention to some of the roles the thetic/telic distinction has played in recent philosophical discussion, so that the importance of clarifying the distinction will be evident. However, it is germane to this more general project to register an observation made *à propos* of the motivation debate by Philip Pettit (1987), and developed by Huw Price (1989). This observation is that one may concede to Smith the claim that no propositional attitude can have opposite directions of fit in respect of its propositional object, without being forced to deny, for example, that desires are beliefs.⁴ For the concession means only, for this case, that a desire that *p* isn't a belief that *p*; not that, for example, a desire that *p* is not a belief that *q*, for some distinct proposition *q* (as

⁴ Whether this concession should be made, the present *ad hominem* context aside, is another matter. What about that attitude defined by saying that one ϕ 's that *p* iff one both believes and desires that *p* (compare: being glad that *p*).

it might be: the proposition that it is, in some suitably attenuated sense, desirable that p).⁵

2.2 *Ethics of Belief.*

This heading is intended to recall such discussions as that of Williams (1970) of the propriety of deciding—for whatever reasons (not bearing on its truth)—to adopt a belief—an obvious place to invoke the *thetic/telic* distinction. The sentiment that such a procedure is manifestly improper is articulated by drawing attention to the fact that a belief adopted because, say, of its comforting effect on the believer, is a belief adopted in flagrant disrespect for the idea that, in belief, the subject's state of mind is supposed to match the way the world is in the respect of the subject matter of the belief. (In §4 below, some effort will be made, however, to distance *direction-of-fit* considerations from those in play in much of the ethics of belief literature.)

2.3 *Analysis of Knowledge.*

One early response to Gettier's counterexamples to the Justified-True-Belief analysis of knowledge was the idea of explicitly excluding the case in which the justified believer comes by a belief that in the circumstances "just happens" to be true, the justification notwithstanding. While the recent history of epistemology abounds with analyses attempting to rule out such cases without ruling out in addition cases of the genuine article, our purposes here are served by citing the very direct response Peter Unger once suggested. Simply say, by way of analysis, that one knows that p iff it is not at all accidental that one is right about its being the case that p .⁶ Consider the case of a subject, S , whose beliefs about the future are monitored by a supernatural being who, taking (for whatever reason) a special interest in minimizing falsity amongst S 's beliefs, intervenes in the course of history so as to make these future-oriented beliefs of S true. Note that we do not suppose that S has the slightest inkling that this is what is going on. It does not seem correct to say that S , who believes, for example, that Islam will be the state religion of a United Europe by the year 2100, *knows* this to be the case, even though it is not at all accidental that S 's belief here is true. The trouble is that the non-accidentality pertains to a matching of the world to S 's mental state rather than in the converse direction that befits a *thetic* attitude.⁷

⁵ On the need for attenuation here, see Humberstone (1987); by ignoring it, it is possible to make spurious trouble for the present suggestion—see for example note 44 of Smith (1987), which is appended to a discussion anticipating the Pettit-Price objection.

⁶ See Unger (1968). By way of clarification, Unger adds (p. 159) "In my analysis of human factual knowledge, a complete absence of the accidental is claimed, not regarding the occurrence or existence of the fact known nor regarding the existence or abilities of the man who knows, but only as regards a certain relation concerning the man and the fact".

⁷ Of course by cashing out non-accidentality in causal terms, as in Goldman (1967), for example, attention to the direction of causation can eliminate the problem noted here.

2.4 Doctrine of Double Effect.

According to the proponents of this doctrine, the badness of intended consequences of actions (whether intended simpliciter, or intended as means to some further end) can make an action wrong in circumstances in which a similar badness of consequences merely foreseen and not intended would not. Opponents of the doctrine are sceptical about how this difference in respect of an agent's propositional attitudes could possibly have the repercussions claimed for the morality of that agent's conduct. The attitudes concerned differ in respect of direction of fit, since foresight is thetic directed and intention is telicly directed. From this perspective, we can put one aspect of the doctrine's appeal by saying that what is especially problematic about using evil means to achieve good results as against foreseeing that bad consequences will ensue either as side-effects or as after-effects of one's conduct, is that the telic direction of fit is inherently subject to moral constraints in a way that the thetic direction is not. The badness of a situation provides no reason whatever for not getting one's beliefs to fit the world in respect of its obtaining, whereas the badness of a situation provides every reason for not fitting the world to one's desire that it obtain.

3. Attempts to characterize the distinction

3.1 Smith's Characterization.

Recall that according to Smith a belief that *p* is (roughly) "a state that tends to go out of existence in the presence of a perception that not *p*, whereas a desire that *p* is a state that tends to endure, disposing a subject in that state to bring it about that *p*".⁸ There are two ways of interpreting the "perception that" locution here. I will argue that if this phrase is read (as it most naturally would be) as meaning "belief that", the characterization is vulnerable to a certain objection, to be called the *mutatis mutandis* objection, whereas if the phrase is read non-doxastically, the characterization does not meet a plausible requirement of universality.

On the first interpretation, the problem with the characterization, as an attempt to explicate the distinction between the two directions of fit, arises over its employment, in what is said about belief, of the concept of perception. Since it would be too restrictive to count the reference to a "perception that not *p*" as specifically to sensory perception, we should understand by this phrase: a coming to believe that not *p*. Or perhaps, since "perceive that" is a factive construction: a coming to know that not *p*. Whichever reading is chosen, it is clear that we are here explicating the thetic direction of fit by reference to states with the thetic direction. Indeed, even if we took the "perceive" in "perceive that" to allude to

In §3.3 we will express dissatisfaction with the idea of explicating direction of fit in terms of causal direction.

⁸ Compare E. S. Russell, as quoted in the final footnote of Braithwaite (1947): "If the goal is not reached, action usually persists".

genuine (i.e. specifically sensory) perception, the point would remain that this is a thetic propositional attitude.

It is true that, whereas the explicandum is a belief that *p*, the explication is in terms of the hypothetical adoption of a belief that *not p*, but this makes no difference to the present point. No asymmetry has been characterized by the observation that whereas a *belief* that *p* tends to disappear on the acquisition of a *belief* that *not p*, a *desire* that *p* does tend instead to persist on the acquisition of a *belief* that *not p*. To disclose an asymmetry, all the relevant *mutanda* must be mutated. The quoted passage itself conceals its vulnerability to this objection by the use of the phrase “perception that *not p*” which may convey the (deceptive) appearance of being neutral with respect to the distinction between directions of fit, and thus legitimately available as a fixed (“unmutated”) constituent in the explication of the asymmetry.

It is important not to misconstrue the dissatisfaction with Smith’s characterization here expressed. Obviously the point has nothing to do with the *truth* of (his form of) the claim that whereas beliefs..., desires---; nor, perhaps less obviously, is it a circularity objection that is being urged. The worry is not that some would-be analysis of the concept of belief fails in virtue of employing, in disguise, that very concept; for clearly no such analysis was being offered.⁹ The point is rather that you cannot informatively characterize a fundamental disanalogy between the ways in which beliefs and desires relate to their objects by contrasting them in a respect itself specified by reference to one of those two ways. It’s as if one were to suggest that there is the following deep asymmetry between men and women as regards sexuality: whereas a heterosexual man will *not* be sexually attracted to males, a heterosexual woman *will* be. A type of example raising similar objections will be familiar to many readers, *viz.*, alleged asymmetries between space and time. For instance, it is claimed that an object can be in the same place at two different times but not at two different places at the same time. It is then replied that there is a tacit “wholly” understood in the latter case, since an object can certainly be partly in one place and partly in another at a given time, and that when we look at the temporal analogue of this particular spatial “wholly”, we find the asymmetry disappears (see Garson 1971). Again, a premature disanalogy-claim resulted from an unnoticed *mutandum* (we say a little more about this *mutatis mutandis* style of objection in §3.2).

We turn to the second interpretation of Smith’s talk of perception, taking the subject’s perceiving that *not p* to amount to no more than its perceptually appearing to the subject that *not p*, where this is related to any belief that may or may not be formed as the *ground* of that belief.¹⁰ This corresponds to the narrow ver-

⁹ Contrast Mayo’s criticism of Schiller in Mayo (1967, p. 148).

¹⁰ I benefit, here and in some formulations to follow, from Michael Smith’s own explanations to me of what he actually had in mind in his (1987); thus this second interpretation is the authorially correct interpretation of the passage in question.

sion of the first reading. More broadly, the second reading would count as a perceiving that *not p*, any case in which it seems to the subject that *not p*, where this phrase is so understood as not to entail that a belief is formed on the basis of its so seeming. On this second reading, perceptions are merely appearances and seemings, and are not associated specifically with the thetic direction of fit in the unwanted way of actually themselves possessing that direction. Thus the *mutatis mutandis* objection lapses.

On the current interpretation, the need to talk of talk of states *tending* to go out of existence or to endure on a perception that *not p* arises from the feature of the present interpretation that makes the *mutatis mutandis* objection no longer applicable. That difference is that we are now reading talk of perception non-doxastically: how things seem or appear to *S*, on the current weak understanding of those verbs, need not be how *S* takes things to be. *S* may well take appearances at face value, and come to believe on the basis of those appearances, that that is how things are. But *S* may suspend belief, perhaps out of suspicion that the current circumstances are not conducive to reliable belief-formation, or indeed retain or form a belief that things are not how they seem in the respect in question.¹¹ The talk of a *tendency* for beliefs that *p* to disappear on perceptions that *not p* is a way of registering the default status of taking appearances at face value and the special or unusual nature of the circumstances prompting the other (suspension, disbelief) responses.

Can we, however, rest content with an account of that in which direction of fit consists which speaks in this way of mere tendencies? It is not just that belief, as a type of propositional attitude, has a certain characteristic direction of fit—one that we have baptized “thetic”—one which it might be deemed to have on the basis of how its instances typically behave, the existence of a minority of instances behaving otherwise notwithstanding. It seems rather that every individual case of believing is a case of attitudinizing which itself has the thetic direction of fit. (Likewise with wanting, and the telic direction.) If this is right, we should look for an account of direction of fit which is *universal* in the sense that it addresses something shared by all cases of belief (or all cases of desire) and locates the theticity of belief (the telicity of desire) in that shared feature.

3.2 Platts' and Anscombe's characterizations.

The objection in §3.1 to the first interpretation of Smith's formulation, as an articulation of the difference in respect of direction of fit between beliefs and desires, was that it pointed to a difference between the way thetic attitudes (beliefs that *p*) related to other thetic attitudes (“perceptions” that *not p*) and the way telic attitudes related to those other thetic (again) attitudes. An obviously cruder way of making this mistake would be to say that the big difference between beliefs and desires is that whereas a belief that *p* is a belief that *p*, a desire that *p* is *not* a belief that *p*. We would like the—as I put it—unmutated

¹¹ In the perceptual case, these matters are elegantly treated in Peacocke (1983) with the concept of representational content; see his example (p. 6) of the *trompe l'oeil* violin painting.

second reference to beliefs to be either mutated or else replaced by something neutral with respect to the terms of the disanalogy. If in Smith's case it is the constancy of this allusion to the thetic direction which thwarts the project, one's first reaction to Platts' discussion is that here the problem is in the constancy of the telic direction. According to the passage quoted in §1, "false beliefs should be discarded; beliefs should be changed to fit with the world", whereas, "the world, crudely, should be changed to fit with our desires".¹² The constant element across the contrast is the "should", and while this does not *make a reference* to a telic propositional attitude the upshot of the whole assertion is to (purport to) *express* such an attitude. (We can say this without commitment to a non-cognitivist—e.g., emotivist—metaethics.) Let us see if this particular kind of "failure to mutate" is objectionable.

As just introduced, the envisaged objection surely cannot be correct. Consider any account of the disanalogy we are interested in, of the form: whereas telic attitudes---, thetic attitudes..., the dashes and dots being filled purely descriptively (or non-evaluatively). It could hardly be an objection to such a proposal that both halves of the "whereas" contrast are suitable for the expression of *belief*. So there cannot be such an objection when in each case a normative (or evaluative) filling is provided, on the grounds that here both halves are suitable for the expression of a *desire*. Let us imagine a variation on Platts' account into which the concept of desire explicitly enters, and then return to his own account in terms of "ought".

One *wants*—so the suggestion would run—one's false beliefs to be abandoned, whereas one does not want one's unsatisfied desires to be abandoned: rather one wants them to be satisfied. Now *this* suggestion (not Platts', recall) may initially appear to be vulnerable to the "failure-to-mutate" objection considered above. The fully mutated analogue of *wanting* not to have false *beliefs* would be *believing* that one's *desires* are satisfied, and there is neither any kind of general tendency for this to be so, nor any condition of rationality which demands it. It is anything but clear, however, that this attempt to re-run the *mutatis mutandis* objection can work. For there is exposed here a difference, and one that deserves to be regarded as a difference in respect of direction of fit, between *what we want from* our desires and *what we want from* our beliefs. The "incomplete mutation" objection does not touch this formulation, the retained telic element notwithstanding, since the asymmetry is being claimed as one between the *contents* of our (higher-order) telic propositional attitudes rather than as one between the conditions for our *having* certain (lower-order) attitudes. (The positive proposal sketched in §4 below, in terms of controlling background intentions, will be of this general form.) In more detail, the contrast just drawn is as follows. The current proposal is that we distinguish an attitude *toward* believings from an attitude (having the same direction of fit) *toward* our desires. This is to be distinguished

¹² Similar formulations appear in Searle (1983, p. 8) and Smith (1988, p. 244).

from saying that having a belief requires meeting some condition, while for a desire some different condition is required. The *mutatis mutandis* objection sets in for an account of the latter form when this difference is itself just a by-product of the thetic/telic contrast it was supposed to illuminate.¹³

Similarly, returning to Platts' own formulation in terms of what *should* be changed to match what, we have a constant telic ingredient in the normative vocabulary. (Platts would not himself agree with this, since the line he is defending is that beliefs about what should be are as straightforwardly thetic in respect of direction of fit; but then we would have a constant thetic ingredient.) There is obviously a problem about all these "shoulds", since the most straightforward way of interpreting them is as moral vocabulary. Do we really agree, under this interpretation, that, as far as any desire is concerned, *e.g.*, Hitler's desire to reduce the Jewish population, the world *should* be changed to fit the desire? (Need even the individual whose desire is at issue hold that the world ought to be brought into line with it? Is there even a *prima facie* moral "should"?) We would prefer a normative but non-moral interpretation.

The non-moral normativity involved here is very clearly brought out by Anscombe's use, in the passage quoted in §1, of the concept of a *mistake*. For—concentrating on the thetic direction—we can see that the concept of a mistaken belief is not just the concept of a false belief. There is another element involved here. Suppose someone said: "I know the belief I held yesterday about the combination of the safe in the office was false, but what makes you say it was a *mistake*?—You are supposing I intended only to have such beliefs as were true". There is a kind of provisional intelligibility to this remark which should persuade us that one only makes a mistake when one thwarts one's own goals. The goal here is that of having only true beliefs; it is not one which every believer (possessing the concept of belief) must share. Its endorsement is again a substantive position in the ethics of belief. Notoriously, we must be careful in the way we make room for its non-endorsement, however. We need to avoid the incoherence, highlighted by Moore, of holding the conceded falsity of a *particular* belief of one's to be of no concern, since one never intended to have only beliefs that weren't false.¹⁴ That is why in the combination-lock case just given, the belief now conceded to be false was taken as one held on the previous day. Even without such a temporal shift, however, we see that the incoherence only arises for some *particular* belief cited as possessed in spite of its falsity. One might very well remain unconcerned at the thought that *some or other* of one's present beliefs were false, and indeed the prospect of devoting oneself to rooting out such false

¹³ A more precise formulation would certainly be desirable than that given here of what makes a particular account vulnerable to the kind of *mutatis mutandis* objection under consideration; I hope the remarks offered will be sufficiently suggestive for present purposes. (There is a similar difficulty in saying precisely what makes an analysis vulnerable to circularity objections.)

¹⁴ See Sorensen (1988), chapter 1, for a contemporary study; the above example is what Sorensen calls *commissively*, rather than *omissively* Moore-paradoxical. (This terminology marks a distinction noted in J. N. Williams (1979).)

beliefs at all costs rather than getting on with the rest of one's life seems itself to be more than faintly irrational.

These considerations, however, do not perhaps serve to rebut the charge that in holding a false belief one does make a mistake. Set aside the general aim—not, if the above is correct, obligatorily possessed—of believing only truths, and consider a believer's aims in respect of the single belief in question. Are these aims not that this state of mind should be responsive to the way the world is? Is this, at least, not required proposition-by-proposition, for belief? We have returned to the point from the end of our discussion of Platts, that the thetic direction of fit is specifically that we want (or "aim at") this responsiveness. But must this goal always take precedence over others? Suppose that last week I stole from the office safe and was due last night to be given a lie-detector examination. Accordingly, I underwent a few days ago a course of hypnosis to instill in me the belief that the combination was 10-24-39, as it had been in fact two years before, when I was employed in a capacity that made me legitimately privy to such information. Last night when the question came up in the test, I confidently and sincerely gave this incorrect answer to the crucial question, and removed myself from the list of suspects. And you are trying to tell me that in believing falsely on this occasion, I was making a *mistake*!¹⁵

You will of course reply—and with great plausibility—that we need to distinguish the question of whether it was a mistake to inculcate the false belief (which, let's agree, it wasn't) from the question of whether having inculcated it, I had deliberately got myself into the position of *mistakenly believing* that the combination was 10-24-39. I surely had. And we can say all this without retracting the point that there are no mistakes in performance other than those of performance based on mistaken belief. I made no mistake in inculcating the erroneous belief because I acted on the perfectly correct (and of course quite distinct) belief that this procedure would save me from detection. It is, then, not possible after all to override the "truth-tracking" aim of holding beliefs by an ulterior goal and avoid having the charge of "mistaken" apply to false beliefs. Such a conclusion is compatible with deeming it sometimes to be reasonable to hold mistaken beliefs when morality or prudence dictates this. The point is just that what might be called the "internal axiology" of belief continues to pass its own negative verdict ("mistake") even when its demands are reasonably overridden. As with Platts' characterization, we must conclude this treatment of Anscombe's suggestion with the observation that while it seems to be on the right track, the normative element we have just noted to be implicit in the suggestion

¹⁵ For the record: the experimental evidence on the success of hypnosis as a polygraph countermeasure is not particularly favourable. It should be noted, though, that effort is usually directed at reducing the anxiety associated with lying rather than with changing one's beliefs so that lying is not necessary. References to the relevant literature may be found in Barland and Raskin (1973, pp. 462-5).

remains somewhat mysterious.¹⁶ An attempt to dispel the mystery—to ground this normativity—will be offered in §4.

3.3 Naturalistic Characterizations.

One common response to the philosophically mysterious is to reach for the biologically explanatory. In this vein, Dennett (1971) directs our attention to the fact that “the capacity to believe would have no survival value unless it were a capacity to believe truths” (p. 101); thus “In general, normally, more often than not, if x believes p , p is true” (p. 102). A non-statistical interpretation of “normally” in this second remark best brings out the point of the first. It is a matter of what beliefs (or more accurately perhaps, belief-forming tendencies) are *for*. The function of beliefs is to represent the world the way it is: if they did not have this feature they would have no survival value for the creatures possessing them.¹⁷ That is their “proper function” in the terminology of Ruth Millikan, who has developed similar ideas at some length in her (1984): just as hearts would not (now) exist if they didn’t pump blood, which is why *pumping blood* counts as the function of the heart, so beliefs would not be here today if they didn’t (under “normal” circumstances) tend to represent the world correctly, which is why *being true* is the function of beliefs.¹⁸ Similarly, if desires had no tendency to be satisfied—if say, a desire for food tended by contrast to reduce the chances of consuming food—then desires would simply not have evolved.¹⁹ We should note in passing that this kind of “evolutionary functionalism” (and its analogue for intentionally created artifacts) is not, even when applied to the illumination of psychological phenomena, at all the same as what goes under the name of functionalism in the

¹⁶ We note that a further problem with Smith’s characterization, on either of the interpretations considered in §3.1, is that this normative element is not so much left mysterious as, rather, left out of account altogether. (Compare the objections in Kripke (1982, 23ff., 34ff.) to dispositional accounts of rule-following.) Similarly, though for reasons of space this is not undertaken here, one can assess Anscombe’s characterization with respect to the *mutatis mutandis* objection urged against one interpretation of Smith. One way of thinking of her deployment, in Anscombe (1957), of the distinction between mistakes in performance and mistakes in judgment is as protection against that objection; see, however, Houlgate (1966).

¹⁷ According to another strand of thought prominent in Dennett’s discussion, but not in those parts of it brought in here, we should refer, not, *à propos* of beliefs, to the creatures possessing them, but rather to the creatures to which we ascribe them. This other strand concentrates not on the utility to believers and desirers of their beliefs and desires, but of the utility to us of making sense of agents (animate or otherwise) by ascribing to them beliefs and desires.

¹⁸ I say “similar ideas” because although the view about desires which follows is Millikan’s, she does not make quite the present suggestion concerning beliefs. Beliefs being true is not what she would call their proper function, but rather a Normal condition for the proper function of desires. (Cf. Millikan 1984, p.99.) This point was clarified for me by James Hopkins.

¹⁹ As Millikan puts it (1984, p. 140): “It is the focused proper function of an explicit desire to produce a state of affairs onto which it maps in accordance with certain mapping rules. (This is not to say that explicit desires are usually fulfilled. Many desires, like sperm, emerge in a world that does not permit their proper functions to be performed. But surely desires proliferated in part because they were *sometimes* fulfilled. What use to have them otherwise?)”

philosophy of mind, since arbitrary black-box (input/output) descriptions of intentional states need make no special reference to this teleological dimension.²⁰

So far, the evolutionary functionalist perspective may appear to throw little light on the distinction between the thetic and telic directions of fit. It does represent one spelling out of the “aims at” terminology in descriptions (as in the quotation in §1 from Platts) of belief and desire as aiming at truth and realization, respectively; but this is by itself not to make a contrast in respect of direction of fit, since the “aims at” (however explicated) is common to both descriptions, and talk of a desire’s being realized (fulfilled, satisfied) is talk of the *same* relation of desire to the world as is involved in talk of a belief’s being true: namely, the relation given by: *a* attitudinizes that *p*, and, in fact, *p* (where “attitudinizes” holds a place for “believes” or “desires”). Indeed, the writers with whom we are here concerned were not addressing the direction-of-fit question. But under magnification, the proposal does offer some promise, since the successful evolution of the belief-forming apparatus requires a causal dependence of the beliefs formed upon the states of the world about which beliefs are held, whereas in the case of the mechanism for desire, what is required is that the way the world is affected by the organism be causally dependent on the desires possessed. So here we have an explication of difference in direction of fit in terms of evolutionarily advantageous difference in direction of causal dependence.

Having elicited this causal directionality difference, one may wonder how important the evolutionary part of the story is after all. Recall its role as grounding the “function of” talk, which was in turn suggested as the source of the normativity involved in saying what ought to fit what: we can say what beliefs ought to do in much the same way as we can say what hearts ought to do—subserve those functions their subserving which constitutes their *raison d’être*.²¹ But we should ask whether evolutionary considerations belong in this discussion at all. It is conceivable that a selective advantage should occasionally, or even frequently, be conferred on those prone to form false beliefs.²² But beliefs in the category in question would not differ from others in respect of how they are “supposed to” fit the world. (The desideratum of universality, again.) It is also conceivable that beliefs—and, for that matter, desires—should be possessed by a creature whose very existence is not to be explained evolutionarily (or “artificially”): for example, having emerged as a result of an explosion in an organic chemistry lab; or by a being (God, perhaps) who has always been in existence and

²⁰ This point is emphasized by Millikan (1984, p. 139), and it is also the main theme of Sober (1990).

²¹ The thetic part of the above proposal has been aired by Stalnaker (1984, p. 18) in terms he regards as naturalistic, though without explicit mention of evolutionary processes.

²² For example, it is reported that women who, having recovered from breast cancer, believe (falsely) that they were wrongly diagnosed and never had the disease, are less likely than those who acknowledge the fact, to suffer a recurrence.

always been capable of having beliefs. Such “merely logical” possibilities will be dismissed as *recherché* by those who want to “naturalize” epistemology and the philosophy of mind. But they cannot be ignored if what we want is an account of what the direction-of-fit contrast *consists in*, as opposed to a description of some of its contingent concomitants. Our present concern is not to explain the emergence of mental states exhibiting the telic and thetic directions of fit, but to explicate the distinction between those directions.

Even if we set aside as irrelevant the question of the evolution (or the design) of the capacities for believing and desiring, there remains, from the views canvassed here, the causal asymmetry. Beliefs get to be possessed because they are true, whereas desires get to be fulfilled because they are possessed. At least, that is how things go under ideal circumstances. Already the restriction to ideal circumstances involves a violation of the demand of universality from §3.1, which in the case of belief restricts attention to circumstances conducive to the formation of true beliefs. But the attempt to explicate direction of fit in terms of causal direction is in even more serious violation of the universality desideratum. Consider, for example, the case of beliefs of the “self-fulfilling prophecy” type, such as the belief that one will indeed be successful in a certain venture, success in which causally requires—and will in the circumstances be ensured by—such confidence. These cases, in which “faith in a fact can help create the fact”,²³ have the direction of causal influence going the wrong way—from mind to world—though they involve beliefs with the same direction of fit as any other beliefs, being appraised for correctness (more on which in §4) in terms of how well their content matches how things are with their subject matter. Our conclusion must be that in noting the typical causal asymmetries between telic and thetic states, we have not got to the bottom of what distinguishes their directions of fit.

4. A positive suggestion

We have noted that, from a logical point of view, the relation of a belief to its propositional object, which has to obtain for the belief to be true, is the same as the relation a desire must bear to its propositional object for the desire to be satisfied: if the object (or “content”) is the proposition that *p* then, in either case, for the attitude to have the property in question is for it to be the case that *p*. Yet while we find it perfectly natural to describe a belief with a true propositional object as a true belief, there is no tendency toward a similar transference in the case of desire. A somewhat more widely applicable term here is “correct”. Not only can beliefs be described as correct (as an alternative to “true”) when their propositional objects are true, but also expectations, answers, and various other things it would be somewhat strained to call *true*. Yet, for all its

²³ In the words of James (1891). (Compare also the example in §2.3 above of the person with the “Islamic Europe” belief.)

greater applicability in this respect, there is no lessening of our resistance to applying it to desires.²⁴ Isn't the reason for this simply that "correct" is a term of favourable evaluation for various propositional attitudes and speech acts, and while it is a merit in a belief (or an expectation, or an answer) to have a propositional object which is true, this is no kind of merit in desire? As Platts put it (in the passage quoted in §1) "falsity is a decisive failing in a belief" whereas "the fact that the indicative content of a desire is not realised in the world is not yet a *failing in the desire*"; correlatively, the fact that a desire *is* realised in the world is not a creditable feature of the desire. To summarize the intuitions in play here, we can say that although in the case of thetic and telic attitudes alike we have a sense of "things going right", when it comes to focusing this favourable evaluation more specifically, in the former case, the evaluation comes to settle on the attitude itself, in the latter case, it settles, if anywhere, on the state of affairs in virtue of which the content of the attitude is a true proposition. Of course, this is only another way of putting the favourable side of the coin whose unfavourable side was expressed by Anscombe in terms of where, if a mistake was to be located anywhere, the mistake should be located. As intimated in §3.3, to make a mistake is to jeopardize one's success in achieving some goal one has (at some level). Thus to trace the source of the normativity in talk of direction of fit for the thetic case, we need to see what makes believing *truly* believing *successfully*. With luck, what we shall see will suggest a suitable treatment for the telic direction.

What, then, makes truth the mark of success for beliefs? To answer this question, it will help to recall the troubles besetting traditional empiricist accounts of memory (Hume, Russell,...), as diagnosed in Urmson (1967). A recurrent theme in such accounts is a contrast between memory and imagination, with discussion tending to ask what grounds this contrast. Urmson points out that two such contrasts are confusedly in play: that between recollections which are genuine memories and recollections which are not, on the one hand, and that between recollections (veridical or otherwise) and mere free imaginings which do not even purport to represent the past as it was, on the other. It is this latter distinction which is relevant here. As Urmson puts it, what distinguishes the cases is not the

²⁴ Stalnaker notices that we appraise true beliefs, but not satisfied desires, as "correct" (1984, p. 80) and proposes to use this as a sign of membership in a broader class—of which belief is the paradigmatic member—of what he calls *acceptance* attitudes. This is conspicuously *not* to raise the question of *why* we describe beliefs but not desires as correct. Perhaps sensing that without an answer to this question, the division into acceptance attitudes and others seems founded on nothing very solid, he goes on to propose (p. 82) a formal characterization of attitudes of acceptance, as those which are (for rational agents) consistent and held towards any proposition which follows from other propositions toward which they are held. This suggestion doesn't work, because the conditions are satisfied by "The proposition that *p* follows from propositions *a* intends should be the case", which is not to ascribe an acceptance attitude to *a* in respect of *p*.

presence of any special features in such mental images as may accompany the recollecting or imagining. Rather (p. 87), “All we have to do is know what criteria of success are applicable, and that is a question which depends on our own intentions”. He has been considering a case in which one’s images are as of conducting the defence in a criminal trial, and the question is whether one is (correctly or otherwise) recollecting, or instead merely imagining, having conducted the defence. “We are recollecting”, Urmson continues, “not if we did conduct the defence in the trial but if it matters whether we did. We are imagining if some such criteria of success as general verisimilitude, or interestingness, are the relevant ones”.

Now, apparent memories, or recollections as Urmson calls them, are only a special case of beliefs. To the more general question of what distinguishes believing that something is the case from imagining it to be, the reply that it is a matter of which criteria of success one has chosen for one’s mental activity seems equally applicable. If the criterion of success is *truth*, then the propositional attitude is *belief*. This is not to say that for every propositional attitude there is some such associated criterion for success, and we need not even endorse Urmson’s suggestion that such criteria exist for imagining.²⁵ The upshot of our earlier observations on the absence of a notion of correctness for desires is that such criteria do not exist in their case. The present point is simply that unless one takes there to be a criterion of success in the case of an attitude towards the proposition that *p*, and, further, takes that criterion to be truth, then whatever else it may be, the attitude in question is not that of belief. So unless the attitude-holder has what we might call a controlling background intention that his or her attitudinizing is successful only if its propositional content is true, then the attitude taken is not that of belief.

This way of explicating the thetic direction of fit presents beliefs having that direction as a matter of a constitutive principle rather than a regulative principle. It’s not—as it might be in the case of some ethics-of-belief-inspired proposal—that one should regulate one’s cognitive life by the principle: make every effort to believe only truths; rather, it’s that unless one counts one’s (current) intentions in φ -ing that *p* as thwarted if it is not true that *p*, one’s φ -ing that *p* does not constitute believing that *p*. Thus the very concept of belief imports its own criterion of success, or, as it was put in §3.2, has its own “internal axiology”. We have here a contrast also with the evolutionary functionalist suggestion of an externally supplied criterion of success in terms of general adaptivity, or perhaps instead, case-by-case pragmatic utility. It is of course useful—to understate the point con-

²⁵ Such a non-endorsement accords well with the fact that imagination has no direction of fit in the sense of the present paper; there is nothing “wrong” with a subject’s mental state or with the world from that subject’s point of view if what is imagined to be the case is not the case (or if it is the case). J. D. Velleman has developed a distinction somewhat similar to that between telic and thetic directions of fit, in Velleman (1992), between propositional contents regarded as true and those regarded as to be made true. The former locution is not intended to suggest “believed”, but, more broadly, to include regarding something imaginatively or hypothetically or... as true. Thus imagination falls on the same side as belief for Velleman’s dichotomy, though it has no direction of fit in the sense of the present paper.

siderably—to have a faculty which is sensitive to the way the world is, since without such a faculty there would be no chance of acting so as to satisfy one's needs in a world which is that way, but the prior concept is the concept of sensitivity deployed in making this point. (One sometimes encounters the view—it is presented sympathetically, if not finally endorsed, for example, in §§1,2 of Chapter 3 in Adams (1975)—that the reason we want our beliefs to be true lies in the instrumental value of true belief to the project of satisfying our more mundane desires.²⁶ This wrongly suggests it to be a perfectly straightforward conceptual possibility—which the prudent agent will refrain from exploiting—for belief-formation to occur with complete indifference to any requirement that the belief in question be possessed only if it is true.²⁷

If we treat the thetic direction of fit in accordance with the above proposal, requiring for an attitude to count as thetic that it be subject to a background intention of only being possessed should the world be a certain way,²⁸ then the normative aspects of talk of correctness and of “what ought to fit what” are straightforwardly given by compliance with the intention. And since, on the proposed account, the intention in question plays a constitutive rather than a regulative role, it is not as though the normativity can be eluded by retaining the attitude while disowning the intention. (Hence the fact, noted in discussion of the lie-detector example in §3, that in believing that *p* when it is false that *p* one is *mistakenly* believing that *p*, whatever extraneous motives one may have had in getting into that state and however well those motives are served by being in it.²⁹) This leaves the telic attitudes, and here the natural suggestion is that we tell a similar story, with, this time, a certain background intention to the effect that the telic attitude, which may be a desire or may itself be an intention, should be fulfilled. In the case where the specific attitude is an intention, say, to cross the road, a question arises about the status of the intention that that intention be fulfilled. Is this the same intention over again, or is it a different—perhaps “higher-order”—intention? One might say that it's nothing but the same intention all

²⁶ Such a view is mentioned, though not endorsed, at p. 121 of Stich (1990). Stich's discussion in the chapter in question is of the truth of sentences in a language of thought whose interpretation is “up for grabs”, whereas we have been presuming a definite propositional content for a belief which has to be true for the belief to be true. This distinction is emphasized in §2 of Harman (1991), at p. 196 of which Harman also makes the remark “More generally, I often have the desire that I believe that *P* only if *P*”. (What worries me about this is of course the “often”.)

²⁷ This talk of a *requirement* is of course not meant to suggest any externally imposed restriction on one's mental life: there is nothing wrong with supposing, hoping, imagining, etc.—it's just that what you're doing won't count as *believing* unless this is a requirement you aim to respect in any given case.

²⁸ We don't want to say, specifically, “should their propositional content be true” since of course *disbelief* is also a thetic attitude.

²⁹ Compare the distinction—often encountered in discussions of the criminal law—between the intention with which, and the motive from which, an action is performed; see, for example, Fitzgerald (1962, p. 120).

over again.³⁰ On this supposition, the telic direction of fit requires little by way of comment: as the point was put above, the “ought” in “what ought to fit what” is given as a matter of compliance with this very intention. On the other hand, a more direct parallel with the thetic case is provided by taking the higher-order route, and the contrast between the two directions of fit is made visible in a formulation of what does the conditioning and what is conditioned, in certain conditional intentions, as we shall now see.

The controlling background intention in the case of belief is a conditional intention. Suppose that the propositional object of a belief is the proposition that p . Then this intention can be described as the intention not to believe that p , given that (or: in the circumstance that) *not* p . Let us represent this by: *Intend* ($\neg Bp/\neg p$). Recall the justification for postulating this intention (derived from Urmson): unless a piece of attitudinizing is thought of as controlled by such an intention, there is no reason to think of it as an instance of believing that p , as opposed to imagining that p , entertaining the proposition that p , supposing that p , desiring that p ,... and so on through the range of non-thetic attitudes. On the “higher-order” proposal for controlling intentions in the telic case, the intention is that it be the case that p , given the telic attitude toward p : intention, desire, or whatever. Actually, as we shall note in the final paragraph below, it may not be that intention is quite the right higher-order attitude to invoke here, but we gloss over these worries for the sake of presenting a definite proposal; a more cautious presentation of the current positive suggestion would treat our italicized “*Intends*” as schematic for some attitude akin to intention proper. Abbreviatively, let us write Wp in this case (for: the subject *wants* that p): *Intend* (p/Wp).

So far, all we have is some suggestive notation, contrasting (1) with (2):

(1) *Intend* ($\neg Bp/\neg p$)

(2) *Intend* (p/Wp)

The notation is meant to recall that of dyadic deontic logic, of course, and in particular to discourage—as in that context—the idea that such ascriptions of conditional intention can be contraposed. We can remain neutral on certain further features of their logical behaviour, as long as we hold firm to this one. It can be explained by either of two accounts of the construction in (1) and (2), between which it is not necessary here to choose. The first, taking its lead from the conditional obligation literature (*e.g.* Lewis 1974), would reach for a notion of preferability underlying claims of the form *Intend* (α/β), according to which this says that its being the case that α & β is preferable (from the intender’s point of view) to its being the case that $\neg\alpha$ & β , glossing this in possible worlds terms thus: some worlds in which the former is true are higher-ranked than any in which the latter is true. Give or take certain complications not to the point here, this amounts to saying that if we restrict attention to the β -verifying worlds, we find that those in which α is true are ranked higher (by the preference ordering, this

³⁰ Cf. Loar (1981, p. 198): “Isn’t desiring s and desiring that one’s desire that s should be T virtually the same thing?” (Think of “ T ” here as a predicate of truth or of fulfillment.)

time conceived as comparing worlds) than those in which α is false.³¹ It is clear that contraposition—the inference from *Intend* (α/β) to *Intend* ($\neg\beta/\neg\alpha$)—fails on this account, and illuminating to contrast (1) and (2) in the terms provided by the account. For (1), we look at the worlds in which the object of the belief is false, and declare our preference for those in which the belief is not possessed. For (2) we look at the worlds in which the desire is possessed, and declare our preference for those in which the object of the desire is true. So what is held fixed in the two cases is different: in the former case, facts about the object of the attitude, but in the latter, facts about its possession. This is a direct formal—or *structural*—rendering of the metaphor of difference in respect of direction of fit: the thing held fixed is that to which what is not left fixed is to be “fitted”. The thetic/telic difference is a difference in the structure of a controlling conditional intention, a difference over what does the conditioning and what is conditioned.

This general description—cashing direction of fit in terms of the logical structure of conditional intentions—also applies in the case of the second way, alluded to above, of interpreting (1) and (2); for this second way, we are to think of the dyadic construction they feature as really monadic after all, with *Intend* (α/β) being an alternative notation for

$$(3) \textit{Intend} (\beta \rightarrow S\alpha)$$

in which the arrow symbolizes (let’s say) material implication, and the “*S*” is a subjunctivizing operator.³² We can think of (3) as describing the attitude of a subject who would endorse the following: either it is not the case that β or else *let it be* the case that α . Such a subject, on learning that β , will (in the absence of a change of heart) come to intend tout court that it be the case that α (“*Intend* ($S\alpha$)” in the present notation). We need not concern ourselves with the semantics of this language here,³³ pausing only to note the failure of contraposition, even when *S* is taken to commute with negation (as in Humberstone (1982)). The passage from *Intend* (α/β) to *Intend* ($\neg\beta/\neg\alpha$), on the present approach, amounts to a passage from (4) to (5):

$$(4) \textit{Intend} (\beta \rightarrow S\alpha)$$

$$(5) \textit{Intend} (\neg\alpha \rightarrow S\neg\beta).$$

³¹ Reading the first “those” as “some of those” and the second as “all of those”.

³² This suggestion is adapted from some work of Castañeda’s, massaged into roughly the present form in Humberstone (1982). (I have some hesitation in referring the reader to this paper, in view of the damage done to intelligibility by the many typographical errors marring the published version: considerable reconstruction is needed at some points, though the general idea will be clear enough.)

³³ The semantic treatment in possible worlds terms of this apparatus in Humberstone (1982) is not quite right for the present application, and would need adjusting along lines suggested by Humberstone (1987); we need not go into the nature of this adjustment here.

But even with the above commutation property and also the assumption that equivalents may be substituted within the scope of “Intend”, the closest we come to (5) on the basis of (4) is

$$(5') \text{ Intend } (S \rightarrow \alpha \rightarrow \neg\beta)$$

which has the subjunctivity in the wrong place.

Alike in their resistance to contraposition, the two suggested understandings of (1) and (2) differ in respect of other inferential properties. Conspicuously, the inference-pattern sometimes called “Strengthening the Antecedent”, from *Intend* (α/β) to *Intend* ($\alpha/\beta \ \& \ \gamma$) is valid on the second interpretation but invalid on the first. Why is (non-)contraposition so important by contrast with these other features, on the basis of which we shall not even attempt to reach a comparative verdict as between the two interpretations?³⁴ As already noted, the failure of contraposition is crucial to seeing the structural contrast between (1) and (2) as formally registering the distinction between the thetic and telic directions of fit: what (1) [= *Intend* ($\neg Bp/\neg p$)] represents as (conditionally) intended is a matter of mental state, whereas material pertaining to mental states in (2) [= *Intend* (p/Wp)] is relegated to the role of a conditioning factor. This is the reason for the disparities observed earlier: fulfilled desires, unlike true beliefs, are not thought of as correct, and an unfulfilled desire, unlike a false belief, does not constitute a mistake. The normativity, either way, attaches to the belief rather than to the desire, because the controlling intention is the intention that one’s beliefs be a certain way in the thetic case (namely, as (1) says, that they not be false) but that the world be a certain way in the telic case. If such representations as (1) and (2) could be contraposed, precisely this much-needed asymmetry would be lost.

The presence of the two negations in (1) echoes (or is echoed by) the “only” in our earlier talk of a background intention in the thetic case to believe only that which is true; one would not with equal naturalness speak of intending to desire only what is in fact the case.³⁵ This difference would of course also be nullified by wanton contraposition. Such formulations call for an additional comment to

³⁴ The existence of telic attitudes which are not, in the terms of Parfit (1984, p. 151), “conditional on their own persistence” would appear to tell against the first interpretation. More accurately, what is relevant is the modal version of this temporal concept: the point is not so much to set aside desires (for example) which one only wants to be satisfied at later times so long as one has those desires at the later times, as to set aside desires which one only wants to be satisfied at worlds in which one has the desires. I learnt of this distinction in respect of (modal) possession-dependence from its application in unpublished work in the 1970’s by André Gallois to the distinction between egoistic and altruistic desires: genuinely altruistic concern for another’s welfare—so the suggestion went—is displayed by a desire that the other should prosper even if one did not want that. Justice is not done to such desires by citing (as on the second interpretation) a ranking of worlds in which what is (there) desired obtains above those in which it does not obtain.

³⁵ Compare the difference between wanting to get along with everyone who is admitted to the club, and wanting only people one gets along with to be admitted. See also Castañeda (1975, p. 67) on “only” in “only if”; the contrast drawn on p. 160 of the same work between the intention expressed by “I shall press button A, unless I jump” and “Unless I press button A, I shall jump” is also relevant: below, use will be made, without explicit further comment, of unless-constructions.

ward off an inappropriate “generality” construal. The intention whose possession is said here to be needed by one purporting to believe that p is the intention only to have that attitude on the condition that p : that is, not to be believing that p unless in fact p . A *general* intention to believe only truths is not to the point, and if the moral of our lie-detector example is accepted, this general intention may well be not possessed—even rationally so—by someone who still has various particular beliefs. (Note that in that example, the present proposal still dictates that the individual who has deliberately inculcated a false belief currently intends-should-that-belief-be-false, *not* to have that belief.³⁶) A similar point applies in respect of desires; see Kenny (1966, p. 95ff).

This matter of generality is one of several which distinguishes the present efforts from forays into the ethics of belief in the tradition of William James. Norms issuing from that tradition have included not only the generalized version of (1) just mentioned—the norm of avoiding false beliefs—but also its mirror-image, that of acquiring true beliefs. A corresponding particularized intention, represented by (1) with the negation signs deleted, appears to lie outside of what we have called the internal axiology of belief: that is, no such intention seems needed in order correctly to be said to believe that p or to believe that *not* p . Of course, *given an interest in the question of whether or not* p , attitudinizing controlled by intention (1) will indirectly be guided to settle on the true belief, since the only alternative (the no-belief-either-way option being excluded) will violate that intention. But no such interest is mandatory.³⁷

Apart from the two respects just mentioned—the question of generality and the norm of gaining true beliefs—another consideration divides the present attempt to base an account of the thetic direction of fit on controlling intentions rather than on the ethics of belief: the role of evidence. The view (opposed by James) according to which it is wrong to believe other than on the preponderance of evidence might suggest that the appropriate controlling intention in the thetic case is not: only to believe what is true, but rather: only to believe what is supported by one’s evidence. It appears, though, that the role of evidence in the present connection is derivative: it is hard to see how one could attitudinize with the aim of not believing something unless it was true other than by taking advantage of whatever evidence came one’s way. A more direct role for evidence seems ruled out, also, by the observation that a true belief retained in spite of overwhelming contrary (but, as it turns out, misleading) evidence is not thought of as

³⁶ According to the present suggestion, the point is not that it is morally wrong to believe falsely, but it is something, which, in respect of any given proposition, a believer must intend not to be doing, in order to qualify as believing. Beware of confusing the intention not to hold a belief which is false with the—vacuous—intention of not holding a belief which one believes to be false. (Compare the difference in respect of what counts as successful execution of the intention, between the intention to purchase a painting by Reubens and the intention to purchase a painting one believes to be by Reubens.)

³⁷ A point nicely emphasized in Heal (1988).

mistaken, even if the subject is criticized as irrational, or held to have made a mistake elsewhere (*e.g.*, as to the strength of the evidence, or its bearing on the case in hand). Finally, a simple reason for not allowing the notion of evidence the role here contemplated comes from the fact that believing something on “blind faith” is still a matter of believing something, so the universality desideratum dictates that evidence-sensitivity cannot play this role.³⁸

Intending to believe only that which is true and intending that what one wants to be the case should be the case, may appear to involve one in conflict of intentions where no such conflict seems plausibly attributed—telling against the suggested account of direction of fit. One might well desire that q but believe that *not* q . How, in this case, the objection proceeds to ask, can one consistently have both the intentions recorded in (1) and (2)? Now if (2) had been given instead as

(2') *Intend* (p/Bp)

then from this, with $\neg q$ for p , together with (1) there would indeed follow:

(6) *Intend* (q/Wq) & *Intend* ($\neg q/B\neg q$)

which gives our subject who believes that p but desires that *not* p contradictory intentions, each relative to some condition which actually obtains, exactly as the objector supposes. But (2) was not given as (2'), as the above discussion of the failure of contraposition for the construction in question emphasized: the whole difference between thetic and telic direction of fit comes down, on the present account, to the difference between what is intended conditionally on what. Making the above moves with (2) in place of (2') leads to the rather different conclusion:

(7) *Intend* (q/Wq) & *Intend* ($\neg B\neg q/q$)

in which the conditional intentions are not similarly contradictory.

Having disposed of the spurious worry about the present account attributing irrationality to our envisaged subject, there remains a genuine, though not rationally criticizable, tension in that subject's state of mind. We can bring this out by calling a conditional intention that α given β *violated* in any situation in which it is true that β & $\neg\alpha$. (These are the “dispreferred” situations on the dyadic-deontic-logic-inspired account of conditional intention sketched above.) In the case of our subject who believes that *not* q but desires that q , we can argue by cases thus, to bring out the respect in which our subject is indeed “intentionally conflicted”:

³⁸ In this last sentence, I am indebted to David Velleman. A related worry—this one a worry for the positive account suggested here—concerns beliefs held not necessarily on faith but perhaps on evidence deemed sufficient to warrant them, yet to which the subject is so firmly committed as to lack the intention not to have them should they be false. I do not know how, short of backing off from the desideratum of universality, to reply to this objection. A similar backdown may also be the only response to a further difficulty, arising over cases in which the falsity of what is believed is—for one reason or another—inconceivable. For example, I believe that $24 + 19 = 43$; but do I have the intention *not* to believe this, conditional on its falsity? What sense can we make of this condition? Both of these difficulties for the present view were put to me by Rae Langton.

Suppose q . Then the intention recorded in the second conjunct of (7) is violated, since $B \rightarrow q$.

Suppose $\neg q$. Then the intention recorded in the first conjunct of (7) is violated, since Wq .

Thus we can see in a way which is a priori relative to whether or not q , that the subject's conditional intentions cannot all be fulfilled (*i.e.*, at least one must be violated). This, as I say, can be regarded as exhibiting a tension rather than an inconsistency in those attitudes: though not irrational, the subject's state is unfortunate. This result is not untoward: any subject with the attitudes in question will regret the putative fact of the unsatisfied desire that q . However, that verdict does not fully exploit the materials at our disposal, drawing only on (i) $B \rightarrow q$ and (ii) Wq . According to the line of objection we now suppose resurrected after conceding that the premature conflation of (2) and (2') was not to the point, any such regret is due to the putatively unfulfilled desire that q , and arguably to the violation of the intention (conditional upon that desire) that q , and not at all to the violation of the theticity-characterizing intention only to believe that *not* q should it be the case that *not* q .³⁹ Suppose that the subject later discovers that in fact q , so the earlier belief was false but the desire is satisfied. According to (7), this should come as a bit of good news and a bit of bad news: the intention attributed in its first conjunct has been fulfilled, while that in its second conjunct has been violated. But in fact—the objection proceeds—isn't the news really all *good* news? Believing your daughter will be convicted while wanting her to be acquitted, how seriously *disappointed* will you be on hearing of the acquittal that your belief was false? Whenever you desire that q but believe that $\neg q$, don't you in fact hope that you're wrong, and won't you be undilutely *delighted* to find that you were? Hardly the sort of response one would expect of an intention violated!

One might attempt a reply to the above objection which stressed the relative seriousness of different intentions, holding that any potential disappointment at having had a false belief is outweighed by relief that, for example, your daughter has been acquitted. But this seems wrong—and to belong more to a puritanical strand in the ethics-of-belief tradition (“Forgive me, for I have sinned against the injunction never to believe falsely”) than to the account suggested in this section. It seems wrong because, for example, the allegedly outweighed concern with believing only the truth leaves no disappointment even when there is nothing to outweigh it when for instance one is (affectively) indifferent towards the proposition one has just learnt one believed falsely. Noting that this is a situation one can only be in after a change of belief, we should consider what happens after a change of desire. There is no disappointment consequent on the non-satisfaction of a discarded desire, a fact which I take it does nothing to undermine the claim

³⁹ I am much indebted to J. A. Burgess for pressing this line of objection.

that one has a higher-order pro-attitude toward the fulfillment of any *current* desire.⁴⁰ (For various reasons, “intention” may not be quite the right word for this attitude; one is that the subject may be alienated from some desires, for which talk of an intention that they be fulfilled sounds too committed; another is that talk of intention can carry the suggestion of intending to *make something the case*, which is not appropriate for the propositional objects of many lower-order telic attitudes.) Thus, in (1), the “*Intend*” and the “*W*” should be understood as temporally indexed to the same time. Likewise in (2), for “*Intend*” and “*B*”. But whereas the past belief has been given up, it might be claimed that one still endorses the past intention to believe only what is true, and that it therefore remains unclear why there is no affective trace of this intention’s having been frustrated. One *might* of course be disappointed at having been “taken in” by what one thought of as good evidence for the belief now abandoned: but any such sentiments would be echoes of an ethics of belief approach, and, in any case, the belief now abandoned might have been precisely appropriate in the light of the evidence then available. The only vestige of the frustrated theticity-characterizing intention is the concession that one was *mistaken*; but if what was said above is correct, there can be no mistake without a thwarted intention. (Thus, in *this* case, “intention” is exactly the right word.) This concludes our discussion of the “no disappointment” objection,⁴¹ and with it, the presentation of the case for seeing the difference in direction of fit between telic and thetic attitudes as constituted by a difference in the direction of conditionality of the controlling conditional intentions which make those attitudes the attitudes they are.⁴²

Department of Philosophy
Monash University
Clayton
Victoria
3168 Australia

I. L. HUMBERSTONE

REFERENCES

- Adams, E. W. 1975: *The Logic of Conditionals*. Dordrecht: Reidel.
 Anscombe, G. E. M. 1957: *Intention*. Oxford: Basil Blackwell.
 Austin, J. L. 1953: “How to Talk—Some Simple Ways”. *Proceedings of the Aristotelian Society*, 53, pp. 227-46.

⁴⁰ §61 of Parfit (1984) on “mere past desires”.

⁴¹ This leaves the “synchronic” form of the objection: how can you at one and the same time hold a belief which you hope is false, when your believing is subject to the intention not to believe the false? But this is just a matter of (say) believing one proposition and desiring that its negation is true, and the discussion surrounding (7), of the consistency of the controlling intentions, applies as above.

⁴² Comments from John (A.) Burgess on an earlier draft of this paper resulted in considerable improvements; I am also grateful to Rae Langton for criticisms. I am especially indebted to Michael Smith and to David Velleman for very full and extremely helpful suggestions on the material, and, in Velleman’s case, for making available a copy of his (1992) before the final version of this paper was completed.

- Barland, G.H., and Raskin, D. C. 1973: "Detection of Deception", in *Electrodermal Activity in Psychological Research*, Prokasy, W. F. and Raskin, D. C., eds. New York: Academic Press 1973, pp. 417-77.
- Braithwaite, R. B. 1947: "Teleological Explanation". *Proceedings of the Aristotelian Society*, 47, pp. i-xx.
- Bratman, M. E. 1987: *Intention, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.
- Castañeda, H.-N. 1975: *Thinking and Doing*. Dordrecht: Reidel.
- Dennett, D. C. 1971: "Intentional Systems". *Journal of Philosophy*, 68, pp. 87-106.
- Fitzgerald, P. J. 1962: *Criminal Law and Punishment*. Oxford: Clarendon Press.
- Garson, J. 1971: "Here and Now", in *Basic Issues in the Philosophy of Time*, Freeman, E. and Sellars, W., eds. La Salle, Ill.: Open Court, pp. 145-53.
- Goldman, A. I., 1967: "A Causal Theory of Knowing". *Journal of Philosophy*, 64, 357-72.
- Harman, G. 1991: "Justification, Truth, Goals, and Pragmatism: Comments on Stich's Fragmentation of Reason". *Philosophy and Phenomenological Research*, 51, 191-5.
- Heal, J. 1988: "The Distinterested Search for Truth". *Proceedings of the Aristotelian Society*, 88, pp. 97-108.
- Houlgate, L. D. 1966: "Mistake in Performance". *Mind*, 75, pp. 257-61.
- Humberstone, I. L. 1982 "Scope and Subjunctivity". *Philosophia*, 12, pp. 99-126.
- 1987 "Wanting as Believing". *Canadian Journal of Philosophy*, 17, pp. 49-62.
- James, F. 1986: *Semantics of the English Subjunctive*. Vancouver: University of British Columbia Press.
- James, W. 1891: "The Will to Believe", in his *The Will to Believe and Other Essays in Popular Philosophy*. New York: Longmans Green and Co.
- Kenny, A. J. 1966: "Happiness". *Proceedings of the Aristotelian Society*, 66, pp. 93-102.
- Kripke, S. A. 1982: *Wittgenstein on Rules and Private Language*. Oxford: Basil Blackwell.
- Lewis, D. K. 1974: "Semantic Analyses for Dyadic Deontic Logic", in *Logical Theory and Semantic Analysis*, Stenlund, S. ed. Dordrecht: Reidel, pp. 1-14.
- Loar, B. 1981: *Mind and Meaning*. Cambridge: Cambridge University Press.
- Mayo, B. 1967: "Belief and Constraint". In *Knowledge and Belief*, Phillips, A. Griffiths ed. Oxford: Oxford University Press, pp. 147-61.
- Millikan, R. G. 1984: *Language, Thought, and Other Biological Categories*. Cambridge, Mass.: MIT Press.
- Parfit, D. 1984: *Reasons and Persons*. Oxford: Clarendon Press.
- Peacocke C. A. B., 1983: *Sense and Content*. Oxford: Clarendon Press.

- Pettit, P. 1987 "Humeans, Anti-Humeans, and Motivation". *Mind*, 96, pp. 530-3.
- Platts, M. 1979: *Ways of Meaning*. London: Routledge and Kegan Paul.
- Price, H. 1989 "Defending Desire-as-Belief". *Mind*, 98, pp. 119-127.
- Searle, J. R. 1979: *Expression and Meaning*. Cambridge: Cambridge University Press.
- 1983: *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Smith, M. A. 1987: "The Humean Theory of Motivation". *Mind* 96, 36-61.
- 1988: "Reason and Desire". *Proceedings of the Aristotelian Society*, 88, pp. 243-58.
- Sober, E. 1990: "Putting the Function Back Into Functionalism", in *Mind and Cognition: A Reader*, Lycan, W. G. ed. Oxford: Basil Blackwell, pp. 97-106. (Excerpted from Sober, E. 1985: "Panglossian Functionalism and the Philosophy of Mind". *Synthese*, 64, pp. 165-93.)
- Sorensen, R. A. 1988: *Blindspots*. Oxford: Clarendon Press.
- Stalnaker, R. C., 1984: *Inquiry*. Cambridge, Mass.: MIT Press.
- Stich, S. P. 1990: *The Fragmentation of Reason*. Cambridge, Mass.: MIT Press.
- Unger, P. 1968: "An Analysis of Factual Knowledge". *Journal of Philosophy*, 65, pp. 157-70.
- Urmson, J. O. 1967: "Memory and Imagination". *Mind*, 76, 83-91.
- Velleman, J. D. 1992: "The Guise of the Good". Forthcoming, *Nous*, 26.
- Williams, B. A. O. 1966: "Consistency and Realism", in Williams, 1973, pp. 187-206.
- 1970: "Deciding to Believe". In Williams, 1973, pp. 136-51.
- 1973: *Problems of the Self*. Cambridge: Cambridge University Press.
- Williams, J. N. 1979: "Moore's Paradox: One or Two?" *Analysis*, 39, pp. 141-2.