Analysis and the Attitudes*

Ian Pratt

If I believe, as I do, that the Loch Ness monster does not exist, then my having that belief is one thing; the nonexistence of the monster (or otherwise) is, of course, another. If I fear, as I do, that war will break out in Europe during the next decade, then my having that fear is one thing; the occurrence of such a cataclysm (or otherwise) is another. So it is with all of the so-called *propositional attitudes*. Whenever I *believe* or *doubt* or *fear* or *desire*, there we have two things: on the one hand, my believing or doubting or fearing or desiring; on the other, that state of affairs which, in having these thoughts, I thereby represent to myself. Question: what, in general, can we say about the relationship between the two?

The problem, known to contemporary philosophy as the problem of intentionality, is to provide an account of the relationship which holds between a state of affairs P, and the state of affairs in which a person entertains a given attitude (typically, belief) towards P. It is widely accepted, among naturalistically-minded philosophers, that a solution can be given which is at once general, succinct and non-trivial. By *general*, I mean an account that works for all possible states of affairs that one may believe: what we seek is a schema that will produce, for any given state of affairs P, a specification of what it means to believe P. By *succinct*, I mean an account which generates these specifications in a principled way: it would hardly constitute an explanation of the propositional attitudes to compile a titanic, perhaps infinite, listing containing a separate entry for each possible content P. By *non-trivial*, I mean an account couched in terms devoid of intentional (or semantic) idiom: if there are any facts about representation, those facts must surely be identifiable within a physicalistic scheme which does not number irreducible psychological (or linguistic) properties among its basic categories. For ease of reference, I shall speak of an account satisfying these criteria—*generality*, *succinctness* and *non-triviality*—as an *analysis* of the attitudes.

*Appeared in Wagner, S. and Warner, S. (eds.): *Naturalism: a critical appraisal*, Notre Dame, IN: University of Notre Dame Press, 1993.

Thus, received wisdom among naturalistically-minded philosophers is that an analysis of the attitudes can in principle be given. The aim of this paper is to unearth the reasons underlying this received wisdom, and to show that these reasons are bad ones. The conclusion I then draw is that, since there is no reason to believe the propositional attitudes to be analysable, it would be a miracle if they were. Therefore they are not.

1 Analysing the attitudes

It is an oft-made point that a normal English speaker's command of the attitude-verbs is thoroughgoingly systematic, in that each of the galaxy of English sentences he can understand can properly form the complement of an attitude-report. Consider the attitude-verb "believes". For just about any declarative sentence P you care to mention, anyone who understands P also understands $\ulcorner S$ believes that $P \urcorner$.[1] Moreover, this command of the attitude-verbs survives the learning of new vocabulary. If someone coins the term "bandersnatch" to denote a newly discovered animal species, then no sooner do we understand the sentence "Bandersnatches are highly intelligent", than we also understand "Karl believes that bandersnatches are highly intelligent". Knowledge of which state of affairs is described by the former brings with it knowledge of which state of affairs is described by the latter. And as for "believes", so for the other attitude-verbs: "doubts", "fears", "desires" etc.

I speak thus of our systematic command of attitude-*verbs* in English (and possible extensions thereof); but this phenomenon is really just the linguistic façade of an underlying conceptual systematicity. For the fact is that anyone who can *think* that P can also think that S believes P, that S doubts P, that S fears P and so on, for just about any proposition P.[2] And, just as our command of attitude-verbs survives the learning of new vocabulary, so our grasp of the attitude-concepts survives the enlargement of our conceptual repertoire. In coming to know what bandersnatches are (whether or not we know the English word for them), no sooner can we entertain the thought that bandersnatches are highly intelligent than we can also entertain the thought that Karl believes that bandersnatches are highly intelligent. That is, there are no gaps in our ability to think about the attitudes of people (either ourselves or others) concerning any of the indefinitely many states of affairs which we can think about ourselves. It is these systematic linguistic and conceptual abilities, I shall claim, that lie at the root of the belief in the analysability of the attitudes.

So much, then, for our systematic grasp of the attitudes. But what sort of thing is a belief or desire? What, in general, does philosophy have to say about the nature of these psychological states? Currently, the most widely accepted answer to this question is *functionalism*, the thesis that psychological states are defined by their functional rôles. Take, for example, the state of *believing that it is raining.* According to functionalism, to believe that it is raining is to be in a state which is endowed with a certain ensemble of causal powers: it must tend to be produced by rain in good perceptual conditions, it must tend to interact with the desire not to be wet and with knowledge about umbrellas and hats so as to produce appropriate behaviour, and much else besides. Notice that a functionalist definition of *believing that it is raining*—if one were ever completed—would make reference not only to environmental causes such as the presence of rain, and to behavioural effects such as the donning of a hat, but also to other psychological states—sundry beliefs, desires and intentions—states which themselves have functionalist definitions in terms of other psychological states, including, perhaps, the state of believing that it is raining. But such circularity is not vicious; it merely requires that psychological states be defined in whole battalions rather than as single spies. As Lewis puts it:

> Think of common-sense psychology as a term-introducing scientific theory, though one invented long before there was any such institution as professional science. Collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses. Perhaps we can think of them as having the form:
>
> > When someone is in so-and-so combination of mental states and receives sensory stimuli of so-and-so kind, he tends with so-and-so probability to be caused thereby to go into so-and-so mental states and to produce so-and-so motor responses.
>
> Add also all the platitudes to the effect that one mental state falls under another—'toothache is a kind of pain', and the like. Perhaps there are platitudes of other forms as well. Include only platitudes which are common knowledge among us—everyone knows them, everyone knows that everyone else knows them, and so on. For the meanings of our words are common knowledge,

> and I am going to claim that the names of mental states derive their meaning from these platitudes.[3]

Commonsense psychological knowledge—*folk psychology*, Lewis elsewhere calls it—can thus be seen as a theory which specifies the functional rôle of every belief, desire, intention etc. it is possible to have. Lewis's suggestion is then that, since it is the platitudes of folk psychology that govern the use of psychological idiom, we should think of phrases such as "belief that it is raining", "desire to stay dry", "puzzlement as to where one's umbrella is" etc. as theoretical terms—terms, that is, whose referents are defined to be the occupants of the functional rôles assigned to them by folk psychology.

It is important to see how functionalism, as characterised above, meshes with the above observations on our systematic grasp of the attitudes. If, as functionalism holds, the defining characteristic of any psychological state is its causal signature, then our ability to grasp what it is for an agent to have any of an indefinite number of beliefs—an ability which, remember, survives the enlargement of our conceptual repertoire—amounts to implicit knowledge of a function mapping any state of affairs P onto the causal signature definitive of the belief that P. Such knowledge cannot, bearing in mind the computational limitations of our brains, be encoded as a separate set of clauses for each possible content P; there are just too many, perhaps infinitely many, possible contents. So there must, it would appear, exist some relatively succinct theory able to generate these mappings, for then our systematic grasp of the attitudes could be explained as knowledge of such a theory. That is, our systematic grasp of the attitudes, together with the limited computational resources of our brains imposes an upper bound on the complexity of the most succinct specification of the relationship between a state of affairs and the functional rôle definitive of some attitude towards that state of affairs.

The natural—and quite reasonable—response of the functionalist to this observation is to go schematic. Rather than have one set of platitudes concerning the belief that it is raining, another set concerning the belief that Caesar crossed the Rubicon, yet another concerning the desire to be rich and famous and so forth, the platitudes of folk psychology are assumed to *quantify* over belief-contents. Indeed, these platitudes might be highly general affairs such as: "People tend, by and large, not to be massively mistaken in their beliefs"[4] or "People tend, by and large, to act in ways which would secure their desires if the world were such that their beliefs were true".[5] As before, ascriptions of psychological states are supposed to be constrained to conform to—or to conform as nearly as possible to—a collection of folk-psychological plattitdes, and this collection of plattitudes is then taken as giving meaning to such ascriptions. But quantifying over attitude-contents as suggested holds out the prospect of a relatively succinct set of such defining plattitudes. Here, the theoretical terms defined by folk psychology are the small band of attitude-verbs themselves—"believes", "desires" etc.—rather than the numberless hordes of complete psychological states—"believes that it is raining", "desires to stay dry" etc. Such a relatively succinct theory would have the desirable property of rendering our systematic grasp of the attitudes explicable: understanding the attitudes is simply a matter of knowing the quantified platitudes that define them. No problems about the effectively infinite supply of content sentences, or about our ability to combine psychological notions with newly acquired concepts.[6] And, of course, to say that such a theory exists just *is* to say that the attitudes are analysable.

The quest for a general, succinct and non-trivial account of the attitudes is by no means confined to functionalist analyses in the style of Lewis. Consider, for example, the approach of Hartry Field. Field proposes to factor the relation of believing—whereby a person is related to a proposition—into two relations: first, a relation called *believing**, which relates a person to sentences in that person's language of thought; and second, a relation of *meaning*, which relates sentences in that language to propositions. Thus, according to Field, for S to believe that P is for S to believe* some sentence σ such that σ means that P. In factoring the notion of *believing* into the notions of *believing** and *meaning*, Field trades the problem of how a person can entertain a thought with proposition P *as content* for the (he thinks) easier problem of how a sentence can *mean* that P. The latter problem is easier because, according to Field, to say what proposition a sentence means is simply to say what possible worlds that sentence is true at.[7] And an account of whether a given sentence is true at a given possible world can be provided by a compositional truth-theory in the style of Tarski. For simple languages, such as those based on first-order predicate logic, such a truth theory would have axioms falling into three classes. First, a large set of axioms specifying the referents of all the proper names in the language: " 'Hesperus' denotes Hesperus", " 'Bismarck' denotes Bismarck", and so on. Second, a large set of axioms specifying the intensions of all the predicates in the language: " 'is a planet' denotes the property of being a planet", "'is a politician' denotes the property of being a politician", etc.[8] Third, a handful of axioms specifying the meaning of the combinator functions and syntax: "$\ulcorner Fa \urcorner$ is true at world w if a denotes an individual ξ and F denotes an intension Ψ such that ξ is in the extension of Ψ at w", "$\ulcorner P \& Q \urcorner$ is true at w if P is true at w and Q is true at w", etc. Of course, languages capable of expressing the full range of human thought might have truth theories considerably more complicated than that just sketched; however, Field believes this fact poses no difficulties in principle for his strategy.

But Field is adamant that to leave matters here world not constitute a satisfactory analysis of the attitudes. For the above account helps itself to the semantic notion of *denotation*; and the litany of axioms specifying the denotation of each of the names and predicates in the language in question (" 'Bismarck' denotes Bismarck", " 'is a planet' denotes the property of being a planet" etc.) hardly constitutes an analysis of the that notion. What is needed, Field urges, is a general, succinct and non-trivial account of what denotation *is*, of *why* a symbol having such-and-such physical characteristics, embedded in such-and-such a physical environment has this or that individual as its referent, or this or that property as its intension.

The kind of account Field has in mind is a causal theory of denotation:

> ...[T]he fact that 'Cicero' denotes Cicero and that 'muon' denotes muon are to be explained in terms of certain kinds of causal networks between Cicero (muons) and our uses of 'Cicero' ('muon'): causal connections both of a social sort (the passing of the original word 'Cicero' down to us from original users of the name, or the passing of the word 'muon' to laymen from physicists) and of other sorts (the evidential and causal connections that gave the original users of the name "access" to 'Cicero' and gave physicists "access" to muons).[9]

There must be something, Field would maintain—and something like a network of causal connections to boot—which unifies the notion of denotation, which binds together the infinitely many pairings of mental symbol with what that symbol is about. And unless we provide an account, however sketchy it may be, of that general causal nexus between symbols and their denotations, unless we can do better than a mere enumeration, we have not given a satisfactory explanation of the phenomenon of intentionality.[10]

The underlying idea is a familiar one which has appeared in various incarnations in the recent philosophy of mind. In its simplest form, the idea is that what makes a type of data-structure in my head the concept of—say—a *horse* is, to a zeroth approximation, the fact that tokens of that type tend to be caused, in good perceptual conditions, by horses, by all horses, and only by horses.[11] Of course, matters must be more complicated than that. For one thing, fake horses can cause beliefs about horses, and camouflaged or concealed horses can fail to do so. For another, whilst the concept of *horse* is at least often activated by real horses (rather than by books or newspapers), the corresponding connection between concept and object is standardly less direct in cases like *muon* or *Cicero*—hence Field's deliberately vague talk of *causal networks*. Nevertheless, defenders of causal theories all take as axiomatic the conviction that it constitutes no sort of explanation of intentionality to give an account of the causal relations holding between "horse" and horses, "muon" and muons, etc. one by one, for all the various concepts we can have. There must be something, the causal theorists maintain, that is common to the cases of all these symbol-denotation pairs. To solve the problem of intentionality is to say something about what this common factor is.

What I have laboured most to draw attention to in this section is the pervasive demand for an analysis of the attitudes—an account, that is, of what it means to adopt an attitude towards a given state of affairs P, which is *general* in that it will work for all possible contents P, *succinct* in that it at least avoids degenerating into an analysis-by-enumeration, and *non-trivial* in that it does not appeal to psychological or semantic notions. I have tried to elicit the natural feeling that only such an account would be intellectually satisfying, and—harder to shrug one's shoulders at—the conviction that only on the supposition that such an account exists would our systematic command of the attitudes be explicable. If these ideas seem too obvious to bear mention, all well and good: I intend to spend the remainder of this paper arguing against them. If, on the other hand, they seem shaky and ill-justified, then I hope at least to have provided evidence of their widespread acceptance.[12]

2 A theory of psychological inference.

In this section, I sketch a theory of psychological inference. By "psychological inference", I understand inference which essentially involves descriptions of persons as doubting, perceiving, affirming, willing, not willing various things—which involves, that is, the attitude-concepts. The object of the exercise is to sketch a possible explanation of our systematic grasp of the attitude-concepts without implying or presupposing their analysability. In so doing, I am not particularly concerned with the plausibility of my 'theory' as a piece of human psychology; for present purposes, I am merely interested in the logical possibilities it opens up. These possibilities will be explored in the next section.

The theory I shall put forward takes as its point of departure an observation made by Hobbes.

> But there is another saying not of late understood, by which they might learn truly to read one another, if they would take the pains; and that is, *Nosce teipsum*, *Read thy self*... [which is meant] to teach us, that for the similitude of the thoughts, and Passions of one man, to the thoughts, and Passions of another, whosoever looketh into himself, and considereth what he doth, when he does *think*, *opine*, *reason*, *hope*, *feare* &c, and upon what grounds; he shall thereby read and know, what are the thoughts, and Passions of all other men, upon the like occasions.[13]

Recommending one's own psychological states as an important source of information about the psychological states of others only makes sense on the assumption that knowledge of the former is less problematical than knowledge of the latter. So Hobbes was taking for granted a faculty—call it *introspection*—whereby we can monitor and know our own psychological states. Though we be not always aware of what we think, though we sometimes deceive ourselves, though our lives obey subliminal urges of which we have not an inkling, we nevertheless enjoy fairly extensive awareness of much of what goes on within our minds. Thus, we frequently know whether we believe something, and we frequently remember what we used at one time to believe, even if we have since changed our minds. And as for beliefs, so for other psychological states as well. But I do not propose to catalogue our introspective abilities and proclivities here. I merely point out that, often at least, if it is possible for us to believe that P, it is also possible for us to be in another psychological state—the belief that we believe P—which can arise through the operation of our faculty of introspection upon the former. Similarly with beliefs about other types of psychological states. The introspectibility of psychological states has tended to be downplayed in the modern philosophy of mind, being portrayed as a contingent fact insofar as it is a fact at all. I propose, by contrast, to promote it to a key rôle.

Introspecting one's attitudes can yield information about their causes. If I can monitor when I came to the belief that there was a cat sitting on the mat in front of me, and if I can thereby observe that I have in the past come to that belief when positioned in conditions of good visibility in front of an unoccluded cat, then I can judge, from this observation and from other of my beliefs, that the latter tends to bring about the former. (Objection: mere observation of conjunctions does not establish causality! Reply: knowledge of the causation of beliefs is, in this regard, no different from our knowledge of the causation of anything else.) Notice that, by remembering what I believed at an earlier time, I can with hindsight come to know under what conditions I falsely believed the cat was on the mat, or under what conditions I failed to believe the cat was on the mat even though it was. Notice also that I may draw conclusions not only about the many possible environmental causes of my beliefs (and other attitudes), but also about their psychological causes. I might for example, observe the patterns of inference that I often follow in reasoning.

There seems to be no limit, in principle, to the sophistication and detail of the knowledge concerning the aetiology of my attitudes that I can garner in this way. And, of course, as Hobbes recommends, these observations of my own psychological states can be generalised to the case of other persons. I can trade on the presumed similarity of other persons' minds to my own by assuming that what caused the belief in me that there is a cat on the mat will cause the same belief in them. So if I know of some person that he is staring at a cat, and I know that that is the sort of thing that causes me to think there is a cat before my eyes, it's a good bet that person is thinking the same thing.

Similar remarks apply to the *effects* of psychological states. I observe that an unqualified intention to move a part of my body in a normal way, to pick up small, liftable objects, to cross the street, to utter sentences of my mother tongue and so on, will standardly be followed by my doing those things. I observe that clashes of desires often precede hesitation. I observe that desires for immediate gratification can often win out over more considered and rational plans to secure my long-term flourishing, and so on. Insofar as I can observe these constant conjunctions, I can form conclusions about the causal powers of my various attitudes, conclusions which can again be generalised from my own case to that of other persons.

The foregoing account—according to which one can reason about the psychological states of others by introspecting one's own psychological states—may, however, be thought implausible. After all, effective reasoning about other people's psychological states would require a large number of these introspective generalisations. And it is arguably unlikely that we make such assiduous use of our introspective faculties that we could form all the requisite generalisations, to say nothing of the problems of storing and accessing the resulting mountain of data. Absent any more detailed proposals as to how such introspectionist methods would actually work, this objection is hard to evaluate. Fortunately, however, there is a far more plausible suggestion as to how we reason about the many and complex interactions between psychological states. And it is to this suggestion that I now turn.[14]

In 1943, Kenneth Craik published a small but influential volume, *The nature of Explanation*, in which he argued that thought should sometimes be understood as a kind of *simulation*. In order to predict the behaviour of physical systems, Craik conjectured, people construct inside their heads mental models of those systems—representations which are, in the relevant ways, functionally isomorphic to the things they represent.[15] Now, as various authors have noted, one obvious domain for reasoning by simulation is that of psychological inference. We have, in the shape of our own brains, ready-made analogues of the brains of others. If only we could take another person and somehow temporarily set ourselves up as believing what he believes, as desiring what he desires and so on, and if only we could observe the progress of these thoughts as they develop in our own minds, we would have a good idea as to how that person's thought is likely to develop.[16]

How might such psychological simulation work? Let us introduce two characters, *Basil* and *Florence*. Basil is trying to reason about what Florence is thinking, and is using psychological simulation to do this. And let us indulge in a little fictional physiology for the purpose of definiteness: we shall imagine that psychological simulations of Florence occur in one part of Basil's brain—call it the "Florence-simulacrum"—and ordinary, non-psychological thought in another. Suppose, then, Basil thinks Florence believes P, Q, R, ..., and desires X, Y, Z, ... In simulating Florence's mind, Basil builds, in his Florence-simulacrum, copies of brain-states which, if they occurred in the part of his brain given over to normal thought, would count as his own beliefs P, Q, R,... and as his own desires X, Y, Z,... . Basil also copies into the Florence-simulacrum certain background beliefs and desires that he takes everyone to share. We might call the ersatz beliefs and desires in Basil's Florence-simulacrum *virtual* beliefs and desires. There may be a certain amount of consistency-checking and maintenance required in order to square Basil's explicit beliefs about Florence's psychological states with his presumption of background knowledge and abilities; but we need not concern ourselves with the details of how this might be achieved. Having cloned a suitable collection of virtual beliefs and desires in the Florence-simulacrum, Basil can can just let matters take their course. Being *copies* of the sorts of psychological states that Basil might be in, they interact exactly as those psychological states would interact in normal thought, producing further psychological states—perhaps the (virtual) decision to perform some action—which Basil, viewing the progress of the psychological simulation, can observe by introspection.

We can assemble the foregoing ideas into a story about how it is that one such as Basil (or such as we) uses his faculty of introspection to determine what it is that another person—Florence, for example—is thinking. Basil's first source of knowledge concerns the *causes* of Florence's psychological states; those conditions—including environmental conditions—which he has found by introspection to cause in him a collection of psychological states Ψ, he may likewise suppose to cause Ψ in Florence. Basil's second source of knowledge concerns the effects of Florence's psychological states; those effects—including behavioural effects—which Basil knows to be caused by a collection of psychological states Ψ in himself, he may likewise suppose are so caused in Florence. Basil's third source of knowledge concerns causal interactions *between* psychological states; when imagining himself in what he takes to be Florence's cognitive predicament, Basil can, in observing the evolution of that imagined state of mind, suppose Florence's real thoughts to evolve along similar lines.

Of course, the supposition that psychological simulation proceeds by producing literal *copies* of the simulator's psychological states in a specially designated part of the brain is only by way of an implementation story. The point I wish to draw attention to here is not the *physical* relationship between, for example, the belief that P and the virtual belief that P, or between the desire that X and the virtual desire that X, but their *functional* relationship. What makes Basil's simulation of Florence a simulation in which Florence is represented as believing P, Q, R ..., and desiring X, Y, Z,... is that, in that simulation, there are states which interact in just the way Basil's *own* beliefs P, Q, R,... and desires X, Y, Z,... would interact if they occurred in normal thought. Now literal copying is one way, but not the only way and doubtless not the most efficient way, to achieve the requisite identity of function between virtual psychological states and their real counterparts. But it is the identity of function that is important here, not the means by which it is achieved. To reason about Florence's belief P, Basil must employ a brain state which functions in his simulation of Florence just as Basil's own belief P would function in ordinary thought. To reason about Florence's desire X, Basil must employ a brain state which functions in his simulation of Florence just as Basil's own desire X would function in ordinary thought. As for beliefs and desires, so for other psychological states: hopes and fears, moods and passions and feelings, and so on: the important thing about simulating them in order to reason about others is similarity of function to their originals in the mind of the simulator.

Now we are in position to describe the characteristic functional relationship between a belief of the form: "S believes P" and the corresponding belief P. Let the functional rôle of the belief P be given; then, as a first approximation, the belief that S believes P functions so as to dispose one to include, in any simulation of the mind of S, a state which, in that simulation, has the functional rôle that the belief P has in ordinary thought. Similarly with desires: let the functional rôle of the desire X be given; then, as a first approximation, the belief that S desires X functions so as to dispose one to include, in any simulation of the mind of S, a state which, in that simulation, has the functional rôle that the desire X has in ordinary thought. And so on, *mutatis mutandis* for other attitudes.

This formulation is nearly correct, but not quite, because the functional similarity between Basil's real belief P and his ersatz belief P used in simulating Florence does not extend to their environmental causes and behavioural effects. Let P be the belief "There is a cat sitting in front of me." For a brain-state of Basil's to qualify as a belief with this content, that brain-state must tend to be caused, under suitable conditions, by the proximity of a suitably positioned cat. That is as important an aspect of the functional rôle of that belief as are the interactions with other of Basil's psychological states into which it may enter. By contrast however, Basil's representation of Florence as believing that there is a cat on the mat should not be caused by the appearance of a cat to Basil. Rather, Basil's disposition to include the ersatz belief P in the simulation of Florence should be prompted by (amongst other things) knowledge of *Florence's* proximity to cats in conditions of good visibility. So we have a difference between the functional rôles of the real belief P and the ersatz belief P used in the simulation of Florence: where the former is caused by perception, the latter is caused by knowledge of Florence's situation.

Corresponding remarks apply to the causation of behaviour by means of psychological states. Suppose Basil believes that the cat is on the mat, and desires to punish the cat for some violation of domestic hygiene; such attitudes might cause Basil's foot to swing aggressively catward. Again, this kind of effect constitutes an important aspect of the functional rôle of the psychological states involved. By contrast, Basil's representation of Florence as believing or desiring these things should not make Basil swing his foot. Psychological simulation of others does not have the immediate and direct effects of ordinary thought. Conclusion: we must amend the above view, according to which the belief that S believes P disposes one to include in any simulation of S a state whose functional rôle within that simulation is identical to the functional rôle of one's belief P in ordinary thought. The identity of function in question must be restricted to internal interactions between psychological states; those aspects of function concerning perceptual causation and behavioural effects are exempt.

We are now in a position to assemble the foregoing observations into an outline of the functional rôles which characterise some simple beliefs involving the attitude-concepts. The belief that S believes P has the following characteristic function: it functions so as to induce one to include, in any mental simulation of S, a psychological state which, input-output conditions excepted, functions within that simulation in just the way that the belief P would function in ordinary thought. Similarly for the other attitudes: the belief that S desires P has the following characteristic function: it functions so as to induce one to include, in any mental simulation of S, a psychological state which, input-output conditions excepted, functions within that simulation in just the way that the desire P would function in ordinary thought. In regard to input-output conditions, where a collection of attitudes is caused by a certain environmental condition, the belief that S has those attitudes should be caused by the belief that S is in a corresponding condition; likewise, that where a collection of attitudes causes certain behaviour, the belief that S has those attitudes should cause the belief that S will exhibit that behaviour. Of course, the above is very much by way of a thumbnail sketch. In addition, it depends upon a simulation-based model of psychological inference whose psychological reality, if true, is surely contingent. However, I hope it illustrates *the kind of* story that could be told about the systematic functional relationship between any belief P and the corresponding belief that S believes P, between any desire P and the corresponding belief that S desires P, and so on through the gamut of the attitudes. And that is all I need for the present.[17]

3 The Problem of Intentionality

Let us take stock. I began by trying to motivate the demand for an analysis of the attitudes, the demand, that is, for a succinct, general and non-trivial solution to the problem of intentionality. Such a demand arises, I said, from the need to explain our systematic grasp of the attitudes. If the defining characteristic of any psychological state is its causal signature, then the ability to understand what it is for an agent to have any of an effectively unlimited number beliefs amounts to implicit knowledge of a function mapping any state of affairs P to the causal signature definitive of the belief that P. But our having this implicit knowledge seems to imply the existence of some relatively succinct theory which generates the required mappings, in which case, analysing the attitudes can then be seen as articulating that knowledge. In the previous section, I sketched my own account of psychological inference, and used that account to draw out the systematic functional relationships which hold between any belief P and the belief that S *believes* P, between any desire P and the belief that S *desires* P, and so on. In brief: the belief that S believes P has the following characteristic function: it functions so as to induce one to include, in any mental simulation of S, a psychological state which, input-output conditions excepted, functions within that simulation in just the way that the belief P would function in ordinary thought. Similarly, *mutatis mutandis*, for the other attitudes. I wish in the present section to consider how the account of psychological inference offered in the previous section might be implemented: what sorts of data-structures might be useful for representing beliefs about people's beliefs, desires, intentions and so forth, and what cognitive mechanisms might be required to process those data-structures in the proper way? In answering these questions, I hope to show how our systematic command of the attitude-concepts can be explained without implying or presupposing their analysability.

I begin with a suggestion made in the previous section: that Basil's belief that Florence believes P might be implemented by giving Basil a disposition to include in a special part of his brain devoted to simulations of Florence (the Florence-simulacrum), a data structure Δ_P which, if it occurred in the part of Basil's brain devoted to normal thought, would encode the belief P.[18] Now, one way for Basil to have such a disposition would be for him to encode the belief that Florence believes P using a data-structure Δ_P^{Flo} which contains Δ_P as a proper part. We should think of tokens of Δ_P^{Flo} as being "wrapped" versions of Δ_P, where the wrapping-paper indexes the data-structure as suitable for simulations of Florence. Setting up a simulation of Florence would then involve copying the kernels of all similarly wrapped data-structures into the Florence-simulacrum, (together, perhaps, with certain background beliefs and desires). By "copying", I mean just that: making physical copies of the kernels of all the data-structures which Basil uses to encode beliefs about Florence's state of mind. This copying having been accomplished, the simulation of Florence will take its course, with the ersatz beliefs and desires within the simulation interacting just as if they were Basil's own beliefs and desires. Small wonder that the ersatz attitudes should have this property: if the relevant data-structures were in another part of Basil's brain, they would be Basil's own beliefs and desires!

To repeat, this is all by way of a fanciful implementation story: something to illustrate how some of Basil's data-structures might have the functional rôles definitive of beliefs about the attitudes of one such as Florence. But—and here is the crux of the matter—on this story, *the systematicity of Basil's grasp of the attitudes is entirely unproblematic.* It is unproblematic, if a belief "S believes that P" is a wrapped version of the corresponding belief P, how the former can be formed whenever the latter can, and can be formed, moreover, in such a way that, input-output conditions excepted (I will come to these presently), it functions in Basil's simulations of Florence in just the way that Basil's belief P would function in normal thought. Moreover, if implemented thus, Basil's ability to engage in psychological inference will survive the enlargement of his conceptual repertoire. When Basil acquires the ability to think about bandersnatches, then the copying and wrapping mechanisms alluded to above can be expected to work on whatever novel data-structures Basil uses to encode his concept of bandersnatch. The situation can be illustrated by considering a computer program whose job is to copy various types of computer files. There are many types of computer file—text, executable, bitmap, raster—and more types yet to be invented. But there is usually no difficulty in writing a program capable of shunting just about any type of file from one location in memory to another: the copying program need have no idea of what the files it operates on are about, or what their functions within the computer system are. So may it be also with the labelling and copying mechanisms that underpin Basil's ability to engage in psychological inference.

We must augment this implementation story to accommodate the differences in functional rôle between the belief that S believes P and the belief P, in respect of their perceptual causes and behavioural effects. Recall that, according to our rough approximation, where a collection of attitudes is caused by a certain environmental condition, the belief that Florence has those attitudes should be caused by the belief that Florence is in a corresponding condition. Likewise, where a collection of attitudes causes certain behaviour, the belief that Florence has those attitudes should cause the belief that Florence will exhibit that behaviour. Let us see whether, on the above implementation story, there is any problem about Basil's systematic command of the attitudes in regard to these aspects of their functional rôles.

I argued in the previous section that knowledge of the the environmental causes and behavioural effects of psychological states could be gleaned by introspection, and that these introspective observations on one's own psychological states could then be simply assumed to apply to other people. How might this process be implemented in Basil conformably with the implementation details that we have fixed so far? Quite simply, as it turns out. What I propose is that Basil, in judging, introspectively, that he believes P, takes the data-structure Δ_P which encodes the belief P, and wraps it in some appropriate packaging, resulting in a data-structure Δ_P^I which contains Δ_P as a proper part. Similarly for Basil's other attitudes: in observing in himself himself the desire X, Basil takes the data structure, Γ_X which encodes the desire X, and wraps it in some appropriate packaging, resulting in a data-structure Γ_X^I which contains Γ_X as a proper part. We can also suppose Basil time-stamps these wrapped data-structures so he later remembers when he had the attitudes in question. These wrappings then protect the data structures against future deliberations that may affect the original attitudes. Thus, should Basil later come to doubt P, he will retain the belief that he once believed P; should he later come to loathe and detest X, he will retain the belief that he once desired X. The important point about this proposal for encoding beliefs about one's own earlier attitudes is that it allows generalisations about the environmental causes and behavioural effects of attitudes to be encoded in a form which makes them suitable for interfacing to psychological simulations. Let me explain.

Let P be the proposition that Basil's cat is sitting on the mat. The previous section offered the following account of how one such as Basil might obtain information about the environmental causes of the belief that P. After a certain amount of experience, Basil may be in a position to judge that, in various situations in the past—e.g. when he was stood in front of the cat sat on the mat in broad daylight—then he came to believe P. Hence or otherwise, he may draw various conclusions about the typical aetiology of this belief—for instance, that whenever he stands in front of his cat sat on the mat in broad daylight, that tends to cause him to believe P. Proceeding in accordance with Hobbes' dictum, Basil may then make judgements about the causes of the belief P in other people—for instance, that if *they* stood in front of the cat sat on the mat in broad daylight, that would tend to cause *them* to believe P.[19] Following the implementation story sketched in the last paragraph, we might suppose that Basil encodes the generalisations about the causes of *his* belief P using data-structures of the form: $C \Rightarrow \Delta_P^I$, where C is a data-structure representing conditions which may cause the belief, and Δ_P^I, as before, a wrapped copy of Δ_P. We might further suppose that Basil encodes the corresponding generalisations about the causes of *Florence's* belief P using corresponding data-structures of the form: $C \Rightarrow \Delta_P^{Flo}$, where Δ_P^{Flo} is a differently wrapped copy of Δ_P. That would be sensible, since then the process of turning the former into the latter could then be effected by systematically exchanging one kind of wrapping for another. Thus, the resulting generalisations about the conditions under which Florence will believe P will be encoded, in Basil's head, using data-structures which contain Δ_P^{Flo}, and hence Δ_P, as proper parts. This means that, should Basil use one of these generalisations to judge that Florence is indeed in such circumstances as are likely to cause her to believe P, he will have a copy of the data-structure Δ_P ready to transfer into his simulation of Florence. That is what I mean when I say that, on the proposed implementation, Basil's beliefs about the environmental causes of Florence's beliefs will be encoded in a form suitable to interface directly to Basil's simulations of Florence. Similarly for Basil's beliefs about the causes of Florence's other attitudes. Similarly also for Basil's beliefs about the behavioural effects of Florence's attitudes: these beliefs will be encoded in a form which allows them to be activated by the occurrence of the relevant data-structures in a mental simulation.

Again, the crux of the matter is that, on this implementation, there is no problem about the systematicity of Basil's grasp of the attitudes. The monitoring and copying mechanisms alluded to in the previous paragraphs will work whatever the functional rôles of the data-structures Δ_P on which they operate. Moreover, such mechanisms will respond effortlessly to any enlargement of Basil's conceptual repertoire: new types of thoughts, connected no-matter-how with the outside world and with other of Basil's existing thoughts, will just be treated like any other data-structure, to be wrapped-up, made the subject of causal generalisations, copied, and run in simulations. Whatever the external aspects of the functional rôle of the new thought are, Basil's introspective mechanisms can in principle recover arbitrarily detailed information about them in a form which can be usefully interfaced to mental simulations of persons whom Basil may later take to have that thought.

Thus, the account of psychological inference set forth in the last section takes our systematic mastery of the concept of belief, indeed, of all our attitude-concepts, in its stride. But nowhere does that account allude to any succinctly specifiable mapping between the state of affairs in which S believes P and the corresponding state of affairs P. Rather, the account is given at the psychological level—in terms of the characteristic functional relationships between the *belief* P and the *belief* that S believes P. I therefore claim that our account of psychological inference bodes ill for the project of analysing the attitudes. For it explains away what I claimed was the only reason for supposing that the attitudes should be analysable at all: our systematic grasp of them. And, I contend, since there is no other reason to believe in a general, succinct and non-trivial mapping between the state of affairs in which S believes P and the corresponding state of affairs P, we should suppose that there isn't one.

The implementation story sketched above makes it clear how Basil can form data structures having the functional rôles definitive of the beliefs that S believes P, that S desires P etc., assuming only that he can form data structures having the functional rôles definitive of the belief P, the desire P etc., respectively. Moreover, this is so no matter what—within limits—those latter functional rôles may be. Thus, if Basil acquires the capacity to think about bandersnatches, then no matter how, exactly, beliefs about bandersnatches are caused by the external world, Basil's introspective mechanisms could, in principle, record arbitrarily detailed information concerning the relevant patterns of causation. Basil might, for instance notice that, upon being confronted with a bandersnatch in broad daylight, then he has the belief "There's a bandersnatch", but that he also has that belief when, as he later discovers, a trickster was projecting realistic holograms of certain sorts in his vicinity, and so on. Certainly, among Basil's beliefs about the typical aetiology of his beliefs about bandersnatches, bandersnatches themselves will, no doubt, loom large: beliefs about bandersnatches (let us pretend) get caused by bandersnatches in certain regular ways. But, on the present view, this fact is quite incidental. If beliefs about bandersnatches were typically caused by quite different things than bandersnatches, Basil could discover that too. Indeed, Basil will make observations concerning the typical aetiology of his beliefs about, for example, *muons* or *Cicero* which do not relate either of these concepts so directly to their originals, to say nothing of his beliefs about prime numbers, hermeneutics or God.

This is where we see the present account of intentionality clashing most sharply with the intuition underlying the traditional view which I tried to elicit in the first section, the view that has it that there must be some succinct and general connection between a representation and what it represents. There must, according to the traditional view, be some core relation which beliefs about muons bear to muons, which beliefs about horses bear to horses, and, if there were such things, beliefs about bandersnatches would bear to bandersnatches. Or, to whistle the same tune in a more propositional key: there must be some core relation which any state of affairs in which S believes P bears to the corresponding state of affairs P. Not so, according to the view developed here. The possible relationships between these two states of affairs can be open-ended; what determines that relationship in the case of a new belief will be the functional rôle of that new belief; and there is no reason, as far as one's abilities to engage in psychological inference are concerned, why that new functional rôle cannot be more or less anything.[20]

The present account, it should be stressed, is compatible with—nay, presupposes—functionalism, considered merely as the thesis that psychological states are defined by their functional rôles. That is, the present account accepts that to believe that it is raining is to be in a brain state which functions thus-and-so, that to believe that the cat is on the mat is to be in a brain state which functions thus-and-so, and so on, for any psychological state one could possibly be in. The thrust of my argument concerns the lack of any need for a common thread running through this list. For there is, I ween, no general, succinct, non-trivial account of the relation between the state of affairs of S's believing that P and the corresponding state of affairs P.

Our conclusion has an idealistic flavour: the primary—and, so far as we know, the only—way to identify the relation between a state of affairs P and the corresponding state of affairs in which S believes P, or desires P, etc., is to describe the way our *thoughts* about believing and desiring etc. function, rather than to reduce believing and desiring to terms and constructs taken from whatever metaphysical catalogue is currently in vogue. If you ask me what it is, in general, for someone to *believe* something, then all I can tell you is what sort of brain-state you will need in order to have the concept of belief; for there is is nothing that believing is, other than that which is thought therewith. Needless to say, this is no idealism which rejects the material world as an illusion, or which denies the objectivity of our judgements about it. No: I really believe there to be no Loch Ness monster; and, as it happens, there really isn't one. Our conclusion is idealistic in this sense: it says that the only general, succinct and non-trivial account of belief takes the form of a specification, at the psychological level, of how the concept of belief functions in thought; and adds thereto that there is no reason to suppose that this specification can be mapped down to the objective level. All that is rejected is the wild goose chase of trying to provide an analysis of the attitudes. There is nothing for philosophy to say about the attitude concepts, about their unity and their utility, that is not best said at the psychological level.[21]

4 Summary

In this paper, I have addressed the problem of intentionality. I characterised that problem as the task of giving an account of the relationship between a state of affairs P, and the state of affairs in which a person entertains a given attitude (typically, belief) towards P. I claimed that the widespread belief in the existence of a general, succinct and non-trivial solution of this problem arises from the fact that our systematic grasp of the attitudes, together with the limited computational resources of our brains, seems to impose an upper bound on the complexity of the best specification of this relationship. I then went on to sketch a theory of psychological inference, and, in doing so, specified the characteristic functional rôles of beliefs of the form "S believes that P", "S desires that X" etc. in terms of the functional rôles of the corresponding belief P, desire X, etc. I showed that these characteristic functional rôles could be implemented without supposing the existence of a general, succinct and non-trivial relationship between a state of affairs P, and the state of affairs in which a person entertains a given attitude towards P. I inferred that we have no reason to believe such a relationship exists.

Department of Computer Science,
University of Manchester,
Manchester M13 9PL, UK
ipratt@cs.man.ac.uk

Notes

1. For brevity's sake, I shall ignore the obviously false implication that normal English speakers can understand sentences with arbitrarily iterated attitude-verbs. I also ignore the grammatical changes, such as those of tense and mood, which English attitude-verbs frequently demand. In addition, throughout this paper, I allow myself some imprecision when it comes to quantifying over the contents of beliefs (e.g. in alternating between "S believes P" and "S believes that P").

2. Provisos corresponding to those of the previous footnote also apply here.

3. David Lewis: "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy*, 50 (1972), p.256.

4. Davidson: "Radical Interpretation", in *Essays on Actions and Events*, New York: Oxford University Press, pp.152–3; Lewis, "Radical Interpretation" in *Philosophical Papers*, vol. 1, p. 112.

5. Lewis, *ibid.* p.113., Stalnaker, Robert: *Inquiry*, Cambridge, MA: MIT Press (1984), p.82.

6. Lewis himself suggests that folk-psychological platitudes might quantify over attitude-contents in "Psychophysical and Theoretical Identifications", note 13. For a more up-to-date account, see, for example, David Lewis: *On the Plurality of Worlds*, New York: Oxford University Press (1986), pp.36 ff. There is one important respect in which I have deliberately misrepresented the views of some of the radical interpretation theorists, including Stalnaker and Lewis. I have represented folk psychology as taking beliefs, desires, intentions etc. to be identifiable, countable internal states or events which stand in certain causal relations to each other. Lewis (at least in his later papers) and Stalnaker urge that the attitudes not be thought of in this way. It does not make sense, they say, to talk of individual beliefs and desires as if they were identifiable states or events: in particular, it would not make sense to *count* the number of beliefs and desires a person has. However, I propose to ignore this feature of their views here, since it does not impinge upon the observations to follow.

7. Actually, Field rejects the possible-worlds construal of propositions, adopting it only for expository purposes. I likewise.

8. By "intension" of a predicate I mean something which determines the extension of that predicate (i.e. the set of things to which that predicate applies) for every possible world. I do not wish to become embroiled in the questions of what such intensions are, or how they 'determine' intensions.

9. Field, Hartry: "Tarski's theory of Truth", *Journal of Philosophy*, 69 (1972), p. 369. See also his "Mental Representation", *Erkenntnis* 13, no.1 (1978), for some discussion of the relation of denotation holding between a predicate and its intension.

10. Interestingly, Stalnaker (*ibid.* pp.30–32.) criticises Field's envisaged analysis for, amongst other things, not being succinct (in our sense) enough. Field proposes only that the plethora of axioms specifying the denotations of names and predicates be generated by some succinct causal theory; Stalnaker points out that to do so leaves the axioms specifying the meaning of the combinators (such as &, $\neg$ etc.) unexplained. A naturalistically plausible theory of truth, Stalnaker urges, ought to explain *why* the symbol & denotes logical conjunction as much as it explains why the name "Cicero" denotes Cicero, a suggestion he bolsters by pointing out that the number of possible combinators is limitless. But Stalnaker's criticism is all grist to my mill: he too is moved by the desire to achieve generality with succinctness.

11. For developments of this idea, see also David Armstrong: *Belief, Truth and Knowledge*, Cambridge, Cambridge University Press (1973), F. I. Dretske: *Knowledge and The Flow of Information*, Cambridge, MA: MIT Press (1981), Jerry Fodor: *Psychosemantics*, Cambridge, MA: MIT Press (1987), ch. 4.

12. For a notable voice of dissent, see Schiffer, S: *Remnants of Meaning*, Cambridge, MA: MIT Press (1987).

13. Thomas Hobbes: *Leviathan*, (ed. C.B. Macpherson) Harmondsworth, Middlesex: Penguin Books (1968), p.82.

14. Notice that we cannot necessarily tell, by introspection, when we are using our introspective faculties. Introspection, as I understand the term here, is merely a *monitoring* process: as such, it need not be conscious (whatever that means); indeed, it need not even be monitored.

15. Craik,K.J.W.: *The Nature of Explanation*, Cambridge: Cambridge University Press (1943), ch. 5.

16. For other writers who have either developed or touched on this idea see: Kahneman, Daniel and Amos Tversky: "The Simulation Heuristic", in Kahneman, Daniel, Paul Slovic and Amos Tversky: *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press (1982); Quine, W.V.O: *Word and Object*, Cambridge, Massachusetts: MIT Press (1960), p.219; Stephen Stich in "Dennett on Intentional Systems", *Philosophical Topics*, 12, 1 (1981), and in *From Folk Psychology to Cognitive Science*, Cambridge, Massachusetts: MIT Press (1983), chapter 5; Lewis, D.K.: "Prisoners' Dilemma is a Newcomb Problem", in his *Philosophical Papers* vol.2, New York: Oxford University Press (1986) pp.299–304; Dennett, Daniel C.: "Making Sense of Ourselves", in his *The Intentional Stance*, Cambridge, Massachusetts: MIT Press, 1987, p. 100; Creary, Lewis G: "Propositional Attitudes: Fregean Representation and Simulative Reasoning", *Proceedings, Sixth International Joint Conference on Artificial Intelligence*, Tokyo (1979); and Haas, Andrew R: "A Syntactic Theory of Belief and Action", *Artificial Intelligence*, vol 28, 3 (1986).

17. In speaking, in the same breath, of the *belief* P, and the *desire* P, I have, in the foregoing explanation, helped myself to the notion of a belief's and a desire's *having the same content*; I need, therefore, to say how I intend this notion to be cashed out. Briefly, what I envisage is a story about the systematic psychological relations holding between a belief P and the corresponding desire P. Such a story would tell, for example, of how the belief P functions so as to terminate planning activity initiated by the desire that P, whilst tending not to affect (or not to affect so directly) the planning activity initiated by desires with content other than P. No doubt such a story would tell much else besides; but I will not attempt to go into the details here. What I do wish to claim is that such a story about the systematic psychological relationships between a belief P and a desire P could then be taken as *constitutive* of what it is for a belief and desire to have the same content. Similarly, with the other attitudes: intending P, doubting P, entertaining the thought P etc. We can give a general account of what it is for two thoughts to have the same state of affairs as content, even though these thoughts are instances of different attitudes. Needless

to say, a general account of what it is to have two thoughts with the same content is far from being a general account of what it is to have a thought with a *given* content.

18. I say that a data-structure Δ *encodes* the belief P for an agent if that agent is so constituted that a tokening of Δ in his brain constitutes his having the belief P.

19. I ignore, for simplicity, variations in the descriptions under which different people are likely to think of Basil's cat, and of the mat on which it is sitting.

20. Might there not be *some* constraints on the new kinds of functional rôles that we are prepared to countenance as definitive of beliefs or other attitudes? Perhaps so. Admissible functional rôles might, for example be required to secure a suitably rich interaction with existing attitudes, and might perhaps be required to respect certain constraints imposed by the framework of beliefs, desires, intentions, etc. Thus, if the belief P is a new belief, then we might require that it be possible to believe conjunctions of the form $P\&Q$ for all existing beliefs Q, that it be possible to form the corresponding *desire* P and *intention* P, and so on. Or again, we might impose a ban on concepts—call them *viral* concepts—whose functional rôle is to gobble up and destroy all one's existing attitudes. To get an idea of what we would be legislating against here, let us consider a concept *tonk* which combines beliefs in the following way: from A, one may infer A *tonk* B; from A *tonk* B one may infer B. The adoption of such inferential policies would soon lead me to infer anything that came into my head, so obliterating all my existing beliefs. Thus, *tonk* has a functional rôle which our minds could not endure; so maybe we should say that it does not count as a proper concept. And of course, it would not be difficult to think up other sensible constraints. But whatever, exactly these constraints turn out to be, they may yet leave a great latitude for beliefs and concepts having all sorts of novel functional rôles—in particular, having all sorts of novel causal relationships with various aspects of the world.

21. There is a connection between the conclusion reached here and the debate over 'procedural' versus 'truth-theoretic' semantics. (See, e.g. Fodor, J: "Tom Swift and his Procedural Grandmother", *Cognition*, 6 (1978), pp.229–247. See also P.N. Johnson-Laird's "What's wrong

with grandma's guide to procedural semantics: a reply to Jerry Fodor", in the same volume, and, for a more recent discussion, Hadley, R.F.: "A Default-Oriented Theory of Procedural Semantics", *Cognitive Science*, 13 (1989), pp.107–137.) The account of the attitude concepts given here might be regarded as a procedural semantics for the attitude verbs. What is interesting about these procedural semantics (if I am right) is their lack of any general succinct, non-trivial, truth-theoretic counterparts.