CANADIAN JOURNAL OF PHILOSOPHY
Volume 31, Number 1, March 2001, pp. 1-22

Are Frege Cases Exceptions to Intentional Generalizations?

MURAT AYDEDE
The University of Chicago
Chicago, IL 60637
USA

PHILIP ROBBINS
Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México
Mexico, D.F. 04510

I Introduction

Let's assume there are psychological generalizations that the folk rely upon in explaining and predicting the behavior of their fellows. Let's further assume these generalizations are intentional, in that they do their explanatory and predictive work by attributing to the subjects in their domain intentional mental states such as beliefs, desires, and the like. Then we can define a *broad* intentional psychology as one that adverts *only* to broad, viz. purely denotational/truth-conditional, mental contents in its generalizations; so the sentences expressing its generalizations should be read transparently. A *narrow* psychology is one that is not so restricted.[1] Accordingly, sentences expressing narrow generalizations will contain opaque contexts, indicated by 'that'-clauses ('believes

1 We don't mean to exclude the position that narrow generalizations attribute states (to the subjects under their scope) simultaneously characterized both referentially *and* non-referentially. It may be that attribution of such states simultaneously takes care of the states' referential properties *and* their 'modes of presentation' (however the latter are understood). In other words, we want to take a narrow psychology as one whose generalizations are to be read opaquely in the relevant contexts, without assuming those contexts to be referentially idle.

that...,' 'desires that...,' and the like). Here is an example of the sort of generalization we have in mind:

(G) If *S* desires that *P* and believes that *S* can bring it about that *P*, then, ceteris paribus, *S* will try to bring it about that *P*.

In recent years, the question of whether such generalizations are broad or narrow has received considerable attention in philosophy of psychology. The general consensus among theorists has been that because generalizations like (G) are false when construed transparently, intentional psychology cannot be broad. For example, when read transparently, (G) seems to be falsified by Oedipus's story. Oedipus wished not to marry his mother and believed that he could achieve this, yet he did not avoid marrying her — on the contrary. So Oedipus satisfied the antecedent and flouted the consequent of (G). In this way, Frege puzzles have served to motivate a narrow intentional psychology, where the intentional properties attributed to mental states are individuated more finely than denotations or truth-conditions.

Alternatively, one could well argue that one of the more pressing reasons to introduce contents more fine-grained than denotations, like Fregean senses, is that we need them to explain the behavior of intentional agents. But if psychological explanation is nomic and intentional, as assumed here, the generalizations involved in the explanation of this behavior must be adverting to such contents; hence those generalizations must be narrow. Indeed, if they were not, we might dispense with problematic notions like sense; denotations or truth-conditions could do all the semantic work needed in psychology. Come to think of it, this conclusion, if cogently reached, would potentially lead to the dismantling of the entire Fregean tradition!

Accordingly, the importance of the question of whether intentional explanation involves broad or narrow generalizations can hardly be overestimated. In particular, arguments purporting to show that such generalizations could in fact be broad are potentially of great significance, and hence should be examined with care. That is just what we intend to do in this paper with a recent argument by Jerry Fodor.

Fodor has argued that intentional generalizations are all broad and the sentences expressing them, like (G), are true. The apparent counterexamples that have made people think that (G) and its ilk must be false on the transparent reading, he says, are not really counterexamples. They are merely exceptions of the sort special science laws typically admit, given that such laws are hedged with 'ceteris paribus' clauses. So Fodor suggests: Read the sentences transparently and treat the apparent counterexamples to the generalizations they express as exceptions, not defeaters.

Of course, Fodor needs to justify this strategy. He needs to explain why certain prima facie counterexamples to transparently expressed generalizations should be treated rather as exceptions without already assuming that sentences like (G), when read transparently, express truths. The burden of argument here is on the broad-minded, for unless the price of nomic exceptionhood is kept sufficiently high, confirmation of ceteris paribus laws becomes too cheap, and the laws themselves become vacuous.[2] So we need to be stingy about granting exceptions. In the case of psychological laws, say, we might choose to limit such grants to cases which are relatively rare and which involve some sort of pathology. But Oedipus's case, and Frege cases more generally, don't meet this standard. Hence the broad theorist's burden.

Fodor sets out to discharge this burden in the second chapter of *The Elm and the Expert: Mentalese and Its Semantics* (Cambridge, MA: The MIT Press 1994). His main argument isn't easy to follow. But the general outline goes like this:

(I) The following principle is true:
Principle of Informational Equilibrium (PIE). 'Agents are normally in *epistemic equilibrium* in respect of the facts on which they act. Having *all* the relevant information — having all the information that God has — would not normally cause an agent to act otherwise than as he does.' (42)

(II) Since PIE is true, any psychology, broad or narrow, must accept it.

(III) No psychology that accepts PIE can count Frege patients as subjects covered by its (relevant) generalizations; that is, Frege patients are outside the proper domain of intentional explanation/prediction.[3]

---

2 As Fodor himself might put it, exceptionhood on the cheap threatens to collapse laws of the form
(i) Fs cause Gs ceteris paribus
into laws of the form
(ii) Fs cause Gs unless they don't. (J.A. Fodor, 'Making Mind Matter More,' *Philosophical Topics* **67** [1990] 59-79)

3 We will use the term 'Frege patient' to refer to agents who
(a) suffer from ignorance of the identity of the referents of some pair of co-denoting concepts in their possession, and
(b) are apt to act on this incomplete information in a way which jeopardizes the success of their behavior.

Though this much seems clear, how it is supposed to work is not. In particular, it is not clear how (III) is supposed to follow from (I) and (II). The rest of this paper is an attempt to work out the details. We think that PIE is pretty clearly false under the relevant reading(s), but we'll put this worry aside for the moment and come back to it below. First, we would like to see how the truth of PIE is supposed to make Oedipus an exception to (G),[4] or more generally, how (III) is supposed to be justified on the basis of PIE. We will begin by untangling two main readings of PIE, depending on how 'normally' is construed; then we'll discuss whether the principle supports (III) on either construal. At that point we turn to the motivation given for PIE. Our ultimate goal is to show that Fodor's attempt to shield broad psychology from Frege puzzles does not succeed.

II How To Read PIE

Though simple in formulation, Fodor's statement of PIE does not wear its intended meaning on its sleeve. In this section, therefore, we propose to survey a range of possible interpretations of the principle, in order to locate one which best fits the larger argumentative context. This preliminary discussion will also serve as an exegetical warm-up to our subsequent critique of Fodor's brief for PIE.

On the first reading, PIE is a descriptive generalization over a population of intentional agents. So its truth depends entirely on empirical facts about this population. It is a statistical reading.

> (R1) PIE: Agents are *usually* (i.e., on most of the occasions on which they act) in epistemic equilibrium in respect of the facts on which they act. Having all the relevant information — having all the information that God has — would not *usually* cause an agent to act otherwise than as he does.

This seems to be the most natural reading. But though it is plausibly true, PIE so read lends no support to (III). From R1 we get:

All agents in Frege cases satisfy (a), but only some of them satisfy (b). So Frege patients constitute a proper subset of agents involved in Frege cases.

4 Oedipus is a Frege patient. He is not in epistemic equilibrium: he would have acted otherwise if he had known all the relevant facts, here the identity of Jocasta and Mom.

(IV) Agents who are not usually in epistemic equilibrium are statistically atypical or abnormal.

In order to derive (III) from (IV) we need something like the following auxiliary claim:

(V) Statistically atypical or abnormal agents are outside the proper domain of intentional generalizations.

And it is not obvious why (V) should be true. In fact, we doubt that it is true. At a minimum it must be explained and argued for, and Fodor does neither. But even if we grant both (IV) and (V), (III) still doesn't follow unless most Frege patients are statistically abnormal in the R1 sense. And this is doubtful.

Take Oedipus for instance. Let's suppose that the only two occasions on which he acted in such a way that he would have acted otherwise had he had all the relevant information about identities are those involving his marrying Mom and killing Dad: i.e., he didn't know that Jocasta = Mom, and that this quarrelsome and arrogant traveler = Dad. So, he ended up marrying Mom and killing Dad — neither of which events he wished for, at least so described. But let's assume, which seems plausible anyway, that on the overwhelming majority of occasions, Oedipus knew all the relevant identities of his (co-denoting) concepts. If so, he will turn out to be a statistically normal agent by the standard of R1.

Of course, there may be agents who by luck or irrationality or sheer lunacy *usually* suffer from identificational ignorance of the Fregean variety. Such agents typically will not be in epistemic equilibrium with respect to the facts on which they act, so perhaps they can be excluded from the domain of intentional generalizations. But most Frege patients are not like this. So on this reading PIE turns out to be true but harmless.[5]

---

5 There is an alternative empirical reading of PIE: '*Most* agents are *always* in epistemic equilibrium in respect of the facts on which they act. Having all the relevant information — having all the information that God has — would *never* cause most agents to act otherwise than as they do.' PIE on this second reading seems to be false, especially if R1 is true. But even if it were true, it is unclear how it could support (III). Again, some connecting premises are needed, and their truth would be moot. But most importantly, if the goal is to exclude Frege patients, then most Frege patients will turn out to be outside the scope of intentional psychology, which seems preposterous. Oedipus, like most Frege patients, lives an otherwise perfectly rational and epistemically responsible life. Even if (G) didn't cover him on two occasions, there were many other occasions it did; likewise for other intentional generalizations. To exclude someone from the domain of intentional psychology simply because he has been a Frege patient on a few occasions is simply unaccept-

Here is a second reading:

> (R2) PIE: It is constitutive of *normal* agency that an agent be in epistemic equilibrium in respect of the facts on which she acts. Having all the relevant information — having all the information that God has — would not cause a *normal* agent to act otherwise than as she does.

This is a normative reading, as opposed to R1, which is descriptive. The sense of normality here is slightly technical. It is one according to which the applicability of psychological generalizations requires that the agent be normal, i.e. in epistemic equilibrium in the relevant sense. So abnormal agents — e.g. agents who are not in equilibrium — are ipso facto not covered by intentional generalization.[6]

On this reading, (III) immediately follows from (I) and (II). But the question becomes why any psychology should accept PIE so read. What makes PIE true? We'll take up Fodor's argument for it shortly (in section III). But first we need to confront a few interpretative questions.

As it stands, PIE is too strong. To see why, consider Oedipus again, as a statistically typical (albeit dramatic) example of a Frege patient. Is Oedipus normal in the R2 sense? No. So, are we to put Oedipus outside the proper domain of psychological generalizations? On a literal reading of R2, the answer to this last question is yes. But can we really treat Oedipus as outside of an entire body of intentional lore (folk or otherwise) simply because on a few occasions he didn't know all the relevant identities on which the success of his behavior depended? This would be tantamount to treating Oedipus as not sharing our psychology, which is simply not credible. What makes Oedipus's story so compelling is precisely that he is one of *us*, that he shares our psychology. Sophocles' success in telling his story relies on just this fact.[7]

To block this undesired consequence, agent normality needs to be relativized to *occasions* of acting:

---

able. (In personal communication, Fodor has confirmed that this was not his intention.)

6 Fodor sometimes gives the impression that he thinks agents who are not in epistemic equilibrium are irrational and that is why they are not covered by intentional generalizations. We'll come to this below.

7 Compare Fodor's own remarks, in the first chapter of *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, MA: The MIT Press 1987), about the psychology of Hermia, Lysander, and Demetrius in Shakespeare's *A Midsummer Night's Dream*.

> (R3) PIE: It is constitutive of *normal* agency relative to a given occasion that an agent be in epistemic equilibrium in respect of the facts on which she acts on that occasion. Having all the relevant information — having all the information that God has — would not cause a *normal* agent to act otherwise than as she does on a given occasion.

So on the occasions on which Oedipus killed his father and married his mother, Oedipus was not in equilibrium in respect of the facts on which he acted. He was therefore abnormal *with respect to those occasions*. Hence, the relevant generalizations that cover these *particular* actions should exclude Oedipus from their proper domain.

Notice that, intuitively, PIE on this reading is intended to exclude agents from the scope of *only* certain generalizations *on only certain occasions of action*. For instance, Oedipus is not to be excluded from the scope of (G) with respect to those acts where he is in epistemic equilibrium, even though he is to be so excluded with respect to his behavior towards his mother and father. And the only reason he is to be so excluded with respect to the latter occasions, according to R3, appears to be because he didn't know the identities relevant on those occasions, i.e., because he was a Frege patient. So, on this reading, the same generalization may simultaneously be both applicable and inapplicable to an agent with respect to different occasions of action.

Note that without an independent argument for it, R3 would be question-begging. Remember that we are trying to see why, according to Fodor, Frege patients are to be excepted from the domain of certain intentional generalizations. And the answer Fodor would give us here — if he had no independent argument for PIE — is this: Oedipus should be excepted from the domain of these generalizations precisely because he is a Frege patient!

Any intentional agent, according to Fodor, should be excepted from the domain of the relevant generalizations *on the occasions of her being a Frege patient*. Fodor seems to take this principle to be valid for any sort of psychology, broad or narrow. By his lights, the issue is whether or not intentional agents, on the occasions when they are being Frege patients, ought to be covered by any sort of intentional generalizations, regardless of their breadth. In other words, it appears that Fodor's argumentative strategy is to take this issue as prior to the issue of how intentional generalizations are to be read. If so, the main burden of his strategy must be carried by the argument he gives for R3. We turn now to its discussion.

III The Argument for PIE: Validity

We submit that Fodor's argument for PIE is unsound, on two counts. First, because it is invalid (this section). Second, because at least one of its premises is false (next section). Here is the argument in question:[8]

(T1) You cannot choose A over B unless you believe you would prefer A to B if all the facts were known to you.[9]

(T2) The success of an action is accidental unless the beliefs that the agent acts on are true.

(T3) No belief/desire psychology can view the normal success of rational actions as accidental.

Fodor says that it follows from T1-T3 'that no belief/desire psychology can fail to accept PIE ... broad or narrow, [belief/desire psychologies] are all committed to treating Frege cases as aberrations' (*The Elm and the Expert*, 42).

Again, there are interpretative problems, especially since at least one of the premises must be read normatively (given that PIE is so read for present purposes). Before we discuss how PIE is supposed to follow from T1-T3 — and why it fails to do so — let's try to clarify the premises first.

Taken at face value, T1 is plainly false: I can (am able to) choose A over B without believing that I would prefer A to B were all the facts in. If it is not to be immediately falsified, T1 must be read something like this:

(T1i) You cannot *normally/rationally* choose A over B unless you believe you would prefer A to B if all the facts were known to you.

Here the addition of 'normally/rationally' would make T1i properly normative, putting it in harmony with R3.[10] In other words, if I choose

---

8 T1 and T2 appear verbatim in Fodor labeled as such (*The Elm and the Expert*, 42). T3 also appears there, but we added the label for expository convenience. Fodor calls the first two premises 'truisms,' and he appears to regard T3 as more or less self-evident.

9 Fodor also offers a substantially weakened version of this thesis in a footnote (*The Elm and the Experts*, 122-3n.3). We will come to it below, in section IV.1.

10 That this reading is the intended one is suggested by Fodor's parenthetical remark on T1, where he notes: 'if an agent has no views about what he would prefer if all the facts were in, then if he is forced to choose, the rational thing for him to do is flip a coin' (*The Elm and the Expert*, 42).

A over B without believing that I would prefer A to B if all the facts were in, then I must be an abnormal or irrational agent, and so be subject to certain censures. Note that in the discussion to follow we will stick to 'rational/irrational' understood in what we take to be the usual internalist sense (hence T1*i*). On this conception, standards of epistemic justification are in an important sense internal to the agent; thus, for example, rationality does not require that all the beliefs out of which an agent acts are true.[11]

According to T1i, if I am rational and choose to do A over B, then I believe that I would still prefer A to B even after full updating. Though this conditional statement follows from T1i, it is considerably weaker. Given the normativity at stake in T1i, the premise is no mere indicative conditional. The 'cannot' in T1i does not indicate a contingent inability; rather, it signifies that rationality/normality *necessarily* requires having the relevant higher-order belief.[12] Thus we will read the conditional as follows:

(1) Necessarily, if I am rational and I choose to do A over B, then I believe that [if all the facts were known to me, I would still prefer A to B].

Call the higher-order belief whose content is expressed by the sentence inside the square brackets, 'HB.' In the next section, we will argue that T1i so construed is false.

Although we doubt that T2 is true, we will leave its discussion aside in this paper.[13]

It is not clear how to take T3. Sometimes Fodor writes as if it is incumbent upon intentional psychology to explain why rational actions are *usually* successful, in the sense that they usually promote the realization of the goals they were intended to promote. But the explanandum at issue is quite different from the usual explananda of such a psychology, viz. intentional behaviors, irrespective of whether they succeed. That such behaviors usually succeed is an interesting fact, and as such invites explanation. But there is scant reason to think that the explanation *must* be

---

11 For discussion of rationality of a different sort, see section IV.2, below.

12 For convenience, we call the belief involved a 'higher-order belief' even though it is a belief about one's preferences, not about one's beliefs.

13 But see J.J. Prinz, 'Is Narrow Content Superfluous?' for an interesting elaboration of doubts about T2. Available on-line at http://csmaclab-www.uchicago.edu/philosophyProject/LOT/jjp1.html

given within intentional psychology itself.[14] Rather, the normal success of intentional behavior appears to be a precondition for the existence of intentional psychology — insofar as the existence of such behavior historically (viz. evolutionarily) depends on its routine success.[15]

Perhaps that is all that Fodor wishes to say. But then it is not clear whether there is any support for PIE in the idea that intentional behavior is *normally* successful. Both narrow and broad psychologies depend on this fact for their continued existence. The explanation of it can be given in terms of the general tendency of rational agents to justify their beliefs well, together with the general tendency of well-justified beliefs to be true. So there is no route here from the normal success of intentional behavior to broad psychology per se.[16] On the contrary, a narrow psychology seems able to do a better job here, since it can cover the occasional unsuccessful behavior, or accidentally successful behavior, to which Frege patients are prone. So a narrow psychology would have a wider scope, hence — ceteris paribus — greater explanatory and predictive power. And that surely suggests its superiority to the broad alternative.[17]

Notice that we have assumed what appears to be the most natural reading of 'normal' in T3, i.e., the statistical reading:

(2) Rational actions statistically tend to be successful (i.e., most rational actions are successful).

T3 assumes (2), and implies that

---

14 D. Arjo, 'Sticking Up for Oedipus: Fodor on Intentional Generalizations and Broad Content,' *Mind and Language* **11** (1996) 231-45, makes roughly the same point.

15 We don't mean to suggest that a historical/evolutionary explanation would be the only proper kind of explanation. Certainly, ahistorical and mechanistic explanations might also apply. These two kinds of explanations are not incompatible.

16 Note, however, that Fodor's officially stated aim is not to show that broad psychology *is* true, but to show that it *might* be true. He writes: 'I therefore propose to argue, in this lecture, that it is *plausible* — not unreasonable to believe — that ... for all we know, the laws of intentional psychology may well be broad ... I pause for emphasis: I'm not going to argue that psychological laws *should* be broadly construed.... What I *am* going to argue is this: the considerations that have been supposed to show that an externalist construal of content won't meet the purposes of psychological explanation are, on balance, unconvincing. So maybe narrow content is *superfluous*' (*The Elm and the Expert*, 28).

17 Assuming, of course, that cetera *are* paria in this case — which Fodor and others have given us independent reasons to doubt (see note 34, below). But such considerations will be largely bracketed for the purposes of this paper.

(3) It is no accident that rational actions statistically tend to be successful.

Since we agree with Fodor that (2) and (3) are true, we agree that

(4) Any belief/desire psychology ought to accept (2) and (3).

If this is all T3 says, namely, that the conjunction of (2)-(4) is true, we have no quarrel with it. However, when it is read this way, Fodor's argument for PIE becomes invalid.

To see this, suppose that

(S) I am rational and I choose A over B.

Then we get

(5) I believe that if all the facts were known to me, I would still prefer A to B [= HB; from (1)]

and

(6) If A is successful and its success is not accidental, then the beliefs out of which I acted are true. [from T2]

It is clear that the following is a suppressed premise in Fodor's argument (which we will also argue against, below):

(7) HB is among the beliefs out of which I acted in doing A.

So,

(8) If A is successful and its success is not accidental, then HB is true. [from (6) and (7)]

What's needed at this stage is the consequent of (8), namely:

(9) HB is true.

As for getting (9) from (8), T3 looks like Fodor's best bet. But that won't do.[18] What T3 yields is the following pair of claims:

(10) A is likely to be successful. [from (S) and (2)]

---

18 It is even doubtful that T3 entails that most rational actions are non-accidentally successful. But since the latter claim seems fairly intuitive, we won't make heavy weather of this point.

(11) The truth of (10) is no accident. [from (S), (3), and (10)]

In other words, the statistical tendency of rational actions to succeed confers a higher probability on A's success than on its failure, and not by accident. To emphasize: what follows is *not* that A *is* successful. To secure PIE, Fodor needs the antecedent of (8), to wit:

(12) A is successful and its success is not accidental.

This is quite strong, compared to (10) and (11).

In short, (12) is needed for the pro-PIE argument to go through, yet it does not follow from T1-T3. Nor can Fodor just assume (12), for to do so would be to require not only that rational actions *always* succeed, but also that they *always* do so in a non-accidental way. But this requirement is both independently implausible and question-begging. It is implausible for two reasons. First, because it would exclude cases of rational actions which succeed by accident; and it seems fair to suppose that such cases exist.[19] Second, because it would exclude *all* cases of unsuccessful action from the explanatory/predictive domain of psychology, which borders on the absurd. Relatedly, the requirement also begs the question against friends of narrow psychology, who want to count Frege patients, like Oedipus, as rational on the relevant occasion(s) of action. For there can be little doubt that Oedipus's actions do *not* succeed, in the sense that they do not further the goals they were undertaken to promote (e.g. finding a suitable mate). This is evident from the fact that, having realized what he's done, Oedipus blinds and banishes himself — rather than, say, patting himself on the back. Expressions of regret like these are typically symptomatic of unsuccessful action.

This leaves Fodor with a dilemma. Either his argument for PIE is invalid, in which case the jig's up; or he has to revert to a statistical reading of PIE, which says that rational actions are, as a statistical rule, carried out by agents who are in epistemic equilibrium in respect of the facts on which they act on those occasions. From this the most that would follow is that Frege cases are not statistically typical. But that is less than Fodor needs, since — pending an argument to the contrary — there's no reason to think that mere statistical atypicality *suffices* for exceptionhood.[20] Either way, he loses.[21]

---

19 For a lovely Gettier-style example, see Prinz.

20 Though one might well suppose that pathological ('breakdown') cases, which ordinarily are atypical in this sense, should count as exceptions. The problem is that atypicality need not imply pathologicality; Oedipus is a case in point. And there are

IV The Truth of the Premises

Though Fodor regards T1i as a truism, we take quite a different view of the matter. Not only does rationality not require that whenever I choose A over B I believe I would prefer A to B if all the facts were known to me; rationality sometimes requires that I *lack* such a belief — and even, on occasion, that I have a belief to the contrary. Or so we will argue.

1. Rationality without conviction

Here is the general description of the sort of situation we have in mind.[22] Circumstances are such that I have to act, i.e., I have to choose A over B by time *t*, but by that time and/or because of the nature of the circumstances, I cannot gather all the *relevant* information pertaining to my choice, which I know I could if the circumstances were more favorable; so I have to make do with the scant evidence available. I do my best to use that evidence in epistemically responsible ways, and I eventually choose A over B before *t*. You come and ask me, just before *t* and after my choice: 'Do you think you would have chosen A even if you had had all the relevant information?' We submit that there are situations like this in which I am agnostic about whether I would choose A over B if all the facts were in, so I lack the higher order belief in question. Moreover, we would also like to claim that there are situations like these in which it is sometimes rational to be agnostic about the relevant HB, i.e. situations in which rationality demands that I be agnostic about the HB. But most interestingly, we will show that there are also situations like these where

---

various other serious problems with a statistical reading of PIE. It's not clear, for instance, how a reading could be given for R1 that applies only to *occasions of acting*, as in the case of R3. (In personal correspondence, Fodor has denied that the statistical reading was intended.)

21 In personal communication, Fodor has conceded that he needs (12), or something like it, to be the 'unmarked case,' but insisted that this is no problem for his view, since PIE is to be read as a ceteris paribus claim. But it's unclear how the principle could be read in this way. For PIE is supposed to specify an umbrella constraint on other things' being equal in the intentional realm — a constraint governing the acceptability of candidate intentional generalizations *in general*. This makes it difficult to see how the usual sort of nomic hedging could be appropriate to it.

22 The cases we will describe are cases involving *decisions under risk*, where choices are made on the basis of the *expected utility* of each option. The cases we are interested in are those where the expected utility of the choice actually made comes out to be greater than that of the alternatives *because* the risk associated with the latter is very high, even though the subjective probability of its occurrence is low. See below.

you actually believe that you would probably have reversed your choice (viz. chosen B over A) had all the facts been in! If this is right, then T1i is false, and Fodor's argument is unsound, regardless of whether or not it is valid.

One set of examples is provided by activities like gambling, playing the lottery, and voting. In such activities, situations routinely arise in which one may act rationally while being agnostic about whether one would do the same if one had all the facts. Sometimes, time is not a constraint, but for various reasons we may knowingly be unable to collect all the relevant evidence, and so must act on the basis of what we believe to be very incomplete evidence.[23] Ought we to believe that we would act the same if we knew everything? We think that we typically don't, and that sometimes we ought not to.

Interestingly, something like this seems to be acknowledged by Fodor himself:

> More precisely, the strength of your preference for A over B should equal the strength of your conviction that you would prefer A to B if all the facts were in. Offered a bet on a fair coin, you shouldn't prefer heads to tails; and you shouldn't think it more likely that you *would* prefer heads to tails if you knew which way the coin will land. (*The Elm and the Expert*, 122-3n.3)

Fodor offers this as a precisification of T1, but it seems considerably weaker. If your conviction that you would prefer A to B if all the facts were in is less than 100%, as is typically the case, then you believe it possible that you would act differently if all the facts were in. Moreover, there is a spectrum of situations in which one might be decreasingly sure that one would prefer A to B if all the facts were in, and T1, even thus weakened, steadily loses plausibility as one continues along the spectrum.

So let's try to accommodate Fodor's remark. Here is a version of (1) which adds parameters for strength of preference and degree of confidence, as suggested:

---

23 Fodor might perhaps object that if your degree of conviction that you wouldn't change your mind after updating is low enough, then you are not making a rational choice after all. But remember the circumstances we are imagining are such that it is not optional for the agent to gather more information: she just can't. Nevertheless, she uses, in an epistemically responsible way, all the evidence she can responsibly gather. The demands of rationality, we take it, extend no further than this. See K. Bach, 'Default Reasoning: Jumping to Conclusions and Knowing When to Think Twice,' *Philosophical Quarterly* **65** (1984) 37-58 for discussion of the inevitable trade-offs between reliability and efficiency that real-world (i.e., resource-bounded) rational choice involves.

(1*) Necessarily, if I am rational and I prefer to degree n to do A over B, then I believe to degree n that [if all the facts were known to me, I would still prefer A to B]

with $0 \leq n \leq 1$, where 1 represents perfect preference/certainty.

But this can't be quite right. Here is a counterexample. Say I'm diabetic and I need to digest some sugar at once, but I need to watch my cholesterol because I also have a heart condition. I am running late for an important meeting and rushing like mad to get there. I stop at a nearby convenience store to get something sweet to eat en route to the meeting. At the counter I see a display of candies in different-colored wrappers. I grab the orange one after hesitating briefly between that and a green one, not bothering to look at the labels. I decide against the green one because I suspect there is roughly an even chance that it contains mint, a flavor I dislike; whereas I've no such worry about the orange one, since I am confident that orange-wrapped candies never contain mint. I make this choice despite my belief that green-wrapped candies tend to be cholesterol-free.

Am I rational in choosing the orange candy over the green one? Intuitively, the answer is yes. But what do my parametrized beliefs and utilities look like? Well, let's suppose that I believe that if a candy on display is wrapped green, then the probability of its being minty is about 50% (as against 1% for the orange alternative) and the probability of its containing cholesterol is about the same (as against 95% for the orange). All else being equal, I'd prefer to stay away from mint. And though I watch my cholesterol level closely, there are times I don't care much about whether the food I eat contains cholesterol if I think the amount of cholesterol it contains is negligible compared to other benefits I might get (e.g., a pleasant taste). So the strength of my preference for the orange candy is quite high; close to 1, say, 0.9. But certainly, this parameter is not matched by the degree of confidence in my relevant HB. If I believe it considerably less likely for green candies to contain cholesterol than it is for orange ones to do so, and I believe that green candies are as likely to be mint-free as they are to be minty, then the strength of my conviction that I would prefer the orange candy to the green one if all the facts were in will surely be *less* than 0.9 — this on the assumption that if I had just a bit more time to look at their labels, I would have made sure to pick out a candy that was both mint- and cholesterol-free.[24] So (1*) is false.

---

24 Also, note that in this example it is intuitively implausible to suppose that when I chose the orange candy over the green one, I had a relevant HB which was among the beliefs *out of which I acted*, in the sense that it was causally implicated in the

There is no shortage of examples of ordinary choice behavior which fit the above general form, including cases involving much tougher choices made under much graver circumstances.

But we can show something even stronger. Let's change the example slightly. Suppose that on the counter there are only orange and green candies, and I believe that all the orange candies contain cholesterol but are not minty. I also believe that the green ones have only a slight chance of being minty but in all likelihood don't contain cholesterol (say the subjective probability of their being either minty or non-cholesterol-free is considerably less than 0.5 for each, say 0.1). Further, avoiding mint flavor is overridingly important to me, since I have a strong allergic reaction to mint which may potentially be fatal. Under these circumstances, my confidence in choosing the orange candy is quite high, but certainly my HB that I would have chosen the same way if I knew all the relevant facts is much less than 0.5. But if this is true, then not only is it rational for me to lack an HB of the relevant sort — that is, to be agnostic about the stability of my preference for the orange candy after updating — it's also rational for me to believe to degree $n > 0.5$ that I would choose the green candy if I had all the information!

Here is another counterexample to (1*).[25] Let's go back, for a moment, to the days before genetic screening was available, and imagine a pregnant woman who believes she has been exposed to high doses of radiation early in her pregnancy. The woman is concerned to avoid giving birth to a deformed baby, though she believes (because she has been told by experts) that, under the circumstances, such an outcome is unlikely, but still better than the average chance if she had not been exposed to radiation. So she decides to have an abortion — despite a strong suspicion that, were all the facts in, she would choose otherwise. It seems clear

---

production of my choice behavior. In personal communication, however, Fodor has indicated that he meant the relevant HB to be (at least) *implicit*, in the sense that were the agent to believe its negation, she would choose differently. Thus an agent would have the relevant HB implicitly just in case she would choose differently were she to believe that she might not stick with her original choice after full updating. But there are difficulties with this suggestion. Most obviously, it remains to be explained how, or in what sense, such implicit HBs could contribute to the production of an agent's choice behavior. To see why, suppose I *lack* the relevant HB: that is, suppose that, even were I to believe that I might not stick to my original choice after full updating, I might not switch. What would be the likely impact on my choice behavior? As far as we can tell, not much. Of course, Fodor might counter that the implicit HB is one out of which I act insofar as my choice of action would not qualify as rational were I to lack it. But this again seems implausible.

25 Thanks to David Malament for suggesting the material for this paragraph.

that, given the woman's beliefs and utilities, this would be a perfectly reasonable decision. But it just as clearly runs afoul of Fodor's principle.[26]

Our routine daily lives might not (fortunately) be dominated by situations like that, situations where we knowingly have to operate with considerably less than desirable evidence. But there are certainly enough of them to refute T1i, since, remember, T1i, as reformulated in (1), was introduced as telling us something about the essence of rationality.[27]

---

26 Notice that the last two counterexamples are just less dramatic versions of the situation involved in Pascal's Wager. I'm trying to choose, on practical grounds, between two beliefs: the belief that God exists and the belief that He doesn't. As a rational epistemologist sensitive to using evidence in responsible ways, I think it unlikely that God exists; my confidence is not complete of course, but it is closer to 1 than to 0.5. But despite this confidence, I decide to adopt the stance of a religious believer on the basis of (a) a very powerful aversion to infinite torture in Hell, and (b) the assumption that such torture is exactly what unbelievers can expect if it turns out that God does exist. It is plausible that my choice is rational, but I certainly don't believe that if all the facts were in I wouldn't switch (and embrace atheism). What's more, I regard it as likely that if all the facts were in I *would* switch! I just can't take the attached risk. (Here we're assuming that under certain conditions 'belief to degree n that P' can be systematically translated to 'belief that the probability of P is k.')

27 Note that in both of these cases, the counterexemplification of T1i depends upon the fact that the agent assigns steeply asymmetric utilities to the possible outcomes of action. One might suppose that such cases need not be taken all that seriously, in the sense that intentional psychology can safely idealize away from them (personal communication with Fodor). But there are problems with this move. First, it's not clear that T1i can be read as a garden-variety ceteris paribus claim, any more than PIE can (see note 21, above). Second, at this point appealing to ceteris paribus clauses has become just too cheap, and in fact, ad hoc. What's needed here is an independent motivation for abstracting away from the cases in question, that is, something other than a prior commitment to broad psychology. But what that motivation might be is anyone's guess. At least it's unclear why asymmetries in an agent's utilities should be thought to interfere with the realization of decision-theoretic laws in anything like the way that, say, friction and air resistance can be held to interfere with the realization of laws governing the motion of spheres on inclined planes. (For more on the general topic of ceteris paribus laws and exceptionhood, see P. Pietroski and G. Rey, 'When Other Things Aren't Equal: Saving *Ceteris Paribus* Laws From Vacuity,' *British Journal for the Philosophy of Science* **46** [1995] 81-110.)

Still, consider the pregnant woman again. Change the situation this way: she now believes (on her doctor's word) that her chances of giving birth to a deformed baby are 50-50. Despite her strong desire to have a baby she aborts on the basis that her desire to avoid a deformed baby is slightly stronger. Here we have about the same probability assignments to the outcome of her actions (aborting or not) with just a *slightly* different utility assignment to each. In this kind of situation, it is rational for her to lack even an implicit HB (see note 24) to the effect that were she to believe that all the facts were in, she would still abort.

2. *Rationality refigured?*

So far, we have operated with an intuitive notion of rationality where the standards of epistemic evaluation are in an important sense internal to the agent. Thus conceived, rationality doesn't require that the beliefs out of which agents act are pretty generally true. It only requires that agents justify those beliefs as best they can;[28] the rest depends on the cooperation of the world. This doesn't render the routine success of rational behavior an accident, for beliefs that are well justified also tend to be true.

Nevertheless, Fodor may now be operating with an externalist notion of rationality, on which all that irrationality requires is the falsity of one or more action-basing beliefs, regardless of one's epistemic efforts.[29] We think this is not our ordinary notion of rationality, but we don't want to quarrel about terminology. We want to grant the legitimacy of the notion as defined, but call it 'e-rationality' to distinguish it from the more familiar internalist notion, which we'll call 'i-rationality.' Let's say that an agent is e-rational on a given occasion only if all the beliefs on which she acts on that occasion are true. Thus we can revise Fodor's first thesis as follows:

---

28 Where the notion of justification is internalist, as noted. At least until recently, Fodor himself used to insist on an internalist notion of rationality: 'according to the present view, questions of rationality are assessed with respect to the vehicle of a belief as well as its content; whereas questions of truth are assessed with respect to content alone.... It's because the vehicle of his belief that his mother was eligible was, say, "J is eligible" rather than, say, "Mother is eligible" that [Oedipus]'s seeking to marry his mother was not irrational in face of his abhorrence of incest' ('Substitution Arguments and the individuation of Belief.' Reprinted in *A Theory of Content and Other Essays* [Cambridge, MA: The MIT Press 1990], 17n.10). See also 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology,' reprinted in *Representations: Philosophical Essays on the Foundations of Cognitive Science* (Cambridge, MA: The MIT Press 1981), 241-3, where the same claim is elaborated in terms of an internalistically understood notion of content, rather than in terms of vehicles of content.

29 We don't think this is likely, but since Fodor's writing on this issue isn't very clear and there are passages which seem to suggest that he regards Oedipus' behavior as rationally defective (or at least 'rash'; see *The Elm and the Expert*, 46), and since it would nevertheless be instructive to see whether an externalist notion of rationality can come to Fodor's rescue, we will proceed with the discussion. Such an externalist notion of rationality is explicitly in play in the work of Ruth Millikan; see, e.g., Millikan, 'White Queen Psychology; or, the Last Myth of the Given,' reprinted in *White Queen Psychology and Other Essays for Alice* (Cambridge, MA: The MIT Press 1993), especially sections 1-5.

(T1e) You cannot e-rationally choose A over B unless you believe you would prefer A to B if all the facts were known to you.

Suppose, as Fodor wants us to accept, that whenever agents responsibly act, among the beliefs and desires out of which they act is the higher-order belief about what they would prefer to do if all the facts were known to them. Now, as we have seen, such a belief will not sometimes be i-rational for the agent to hold if, depending on the situation, she believes it may as well be false for all she knows. (Fodor himself must grant this point if he is serious about the note on T1 cited above.) But if held nevertheless, it may still be e-rational if it turns out to be true on a given occasion. If it happens to be false then the agent will be e-irrational. Indeed, supposing that Oedipus held higher-order beliefs of the relevant sort on the occasions he married Jocasta and killed his father, he was e-irrational on these occasions according to T1e.

One immediate problem with T1e is this. If we are right about the falsity of T1i, then many agents on many occasions simply won't have the relevant higher-order beliefs about their preferences. In fact, if they are self-reflective and careful, there will be situations in which they would deny having such a belief if queried directly. But then there is no telling whether such an agent is e-rational on the basis of whether her higher-order belief is true, for in all likelihood she won't be holding any such belief. But then T1 just collapses: if most agents are i-rational, then many agents will sometimes lack any higher-order belief of the sort T1e specifies. So many otherwise i-rational agents will likely become e-irrational *whether or not they are Frege patients.*

So it turns out that T1e is too strong even for Fodor: many ordinary i-rational agents will turn out to be e-irrational and will therefore be potentially outside the proper scope of intentional generalizations (i.e. if Fodor succeeds in deriving the truth of R3 from T1-T3, which of course he doesn't — see above).

V Conclusion

When read as R1, PIE does not imply that Frege patients are, or ought to be treated as, exceptions to the relevant intentional generalizations. When read as R3, PIE would give the desired result — if it were true. Taken on its own merit, PIE would seem to beg the question against opponents of broad psychology. And Fodor's reasoning does not establish the principle. So no reason has been given for accepting it.

Can anything be salvaged from Fodor's argument for PIE? To answer this question, we propose to look at what sort of intuitions might be driving his attempt to excuse intentional generalizations from having to

cover Frege patients. The following passage seems to indicate the bottom line:

> I assume that any intentional psychology that we can imagine taking seriously will construe a creature's behavior as largely determined by causal interactions between its beliefs and its utilities. My point is that *no* such belief/desire psychology, *broad or narrow*, can tolerate a general proliferation of Frege cases. *Any* intentional psychology, broad or narrow, has to take for granted that identicals are generally de facto intersubstitutable in belief/desire contexts *for those beliefs and desires that one acts on*. For if this isn't granted, there is nothing to connect the rationality of an action with the likelihood of its success. Suppose Fa is something that Smith wants and Fb is something that he doesn't want. Then Smith will find that Fa tastes of ashes if a=b. This is all right as far as it goes; such things happen. Getting what you want can be awful. The problem, however, is that if a=b and Smith doesn't know it, even perfectly prudent behavior in respect of *a* (viz., of *b*) won't tend towards the satisfaction of Smith's desires *except by accident*. And, surely, no serious belief/desire psychology could treat the routine success of prudent behavior as *accidental*. (*The Elm and the Expert*, 40)

Fodor seems to assume throughout that Frege cases are statistically *rare* (≈ R1) and *unsystematic* (≈ that they are rare is no accident).[30] We agree with this assumption — call it A1 — provided that Frege *cases* are taken to be the occasions for actions involving Frege *patients* such that their actions on those occasions are either unsuccessful or, if successful, then accidentally so.[31] We also agree that no belief/desire psychology can tolerate a general proliferation of Frege patients — call this A2 — in the sense that their systematic proliferation would undermine the general utility of intentional behavior and hence, in the long run, remove the ontological precondition of such a psychology. However, it doesn't automatically follow from these assumptions that 'any intentional psychology, broad or narrow, has to take for granted that identicals are generally de facto intersubstitutable in belief/desire contexts *for those beliefs and desires that one acts on*' — call this target conclusion T. In order for T to follow from A1 and A2, Frege patients must already have been excluded from the proper domain of intentional explanation. What is needed, then, is an *independent* argument to show that Frege patients are exceptions to psychological generalizations in the general case.

---

30 See *The Elm and the Expert*, 43-7.

31 We think that Frege cases are ubiquitous in a broader sense according to which most intentional agents have many co-denotational concepts or extensionally equivalent thoughts — some of which may not be known to be so to their hosts, and only a few of which happen to provide the occasion for action, in which case their hosts become Frege patients (see note 3, above).

We have seen that Fodor's main attempt to meet this need fails. But in the passage above he seems to be pointing to the idea that drove his argument by claiming that, if the point is not granted, there is nothing to link the rationality of an action with the likelihood of its success. We deny this. Let's say we accept A1 and A2, and refuse to grant T. Are we then left with no way to connect the rationality of an action with its tendency to bear fruit? Surely not.

As his remarks on 'perfectly prudent behavior' suggest, Fodor here seems to mean by the rationality of an action i-rationality.[32] So let's take the claim as demanding an explanation of what links the i-rationality ('prudence') of an action to the likelihood of its success. In that case we have already mentioned what we take to be the right answer. It's just this: i-rational behavior is behavior which is grounded in beliefs that tend to be well justified; such beliefs tend to be true; hence the generally rosy prospects for behavioral success.[33] Of course, even perfectly prudent behavior occasionally fails. But this is not only intelligible but in fact is predicted on a narrow psychology, so long as A1 and A2 are true. We think that this makes a narrow psychology a better choice than a broad one.[34]

---

32 If Fodor has e-rationality in mind, his claim doesn't make much sense (and may be question-begging). For supposing that e-rational actions are ipso facto caused by true beliefs, it is already built into the claim that successful actions result from true beliefs; so what is the point of claiming that if T is not granted there is nothing to connect the e-rationality of an action to its success? Irrespective of whether T is granted or not, if the question is what connects the e-rationality of an action to its success, then the question answers itself: the truth of the beliefs out of which the agent acts.

33 Prinz makes much the same point.

34 This assumes, of course, that the notion of content required by a narrow psychology is non-problematic and the project of constructing one is viable. We haven't touched on this issue. If Fodor's long-standing arguments against the viability of such a notion — especially if worked out in terms of functional-role semantics — are cogent, then we may have to learn to live with the fact that intentional psychology can't explain why Frege patients do what they do. (Note too that Fodor's own notion of narrow content as a mapping from contexts to broad contents is no help with the explanation of Frege cases involving concepts expressed by proper names; see M. Aydede, 'Has Fodor Really Changed His Mind on Narrow Content?' [*Mind and Language* **12** (1997) 422-58] for details.) This *might* be the real insight behind Fodor's argument. On the other hand, the successful folk practice of explaining *inter*personal Frege cases suggests that there must be a viable notion of narrow content — at least if we assume that this practice is, as Fodor likes to say, 'intentional through and through' (*Concepts: Where Cognitive Science Went Wrong* [New York: Oxford University Press 1998], 7).

We conclude that Fodor's core idea, though based on fair assumptions about the permissible frequency of Frege cases, leads him astray.[35]

Received: July, 1999

Fodor has an alternative proposal about how to treat Frege patients: they are to be explained at the sub-intentional level, via differences in the 'syntax' of the relevant concepts, understood as Mentalese terms. If this is right, handling these cases need not involve subsuming them under intentional generalizations of any sort. But see M. Aydede, 'Fodor on Concepts and Frege Puzzles,' *Pacific Philosophical Quarterly* **79** (1998) 289-94 and 'On the Type/Token Relation of Mental Representations,' *Facta Philosophica* **2** (2000) 23-49 for a critique of this proposal.

35 We would like to thank Sara Bernal, Jonathan Cohen, Jerry Fodor, Melinda Hogan, David Malament, Eric Margolis, Mark Moyer, and Jesse Prinz, as well as an anonymous referee for this journal, for helpful feedback. Portions of this paper were delivered at the Pacific APA meeting and at the 91st meeting of the Southern Society for Philosophy and Psychology (SSPP) in April 1999; we would like to thank the audiences for their comments and questions.