Concepts Are Beliefs About Essences

Ulrike Haas-Spohn[*], *Wolfgang Spohn*[**]

1. Introduction

Putnam (1975) and Burge (1979) have made a convincing case that neither meanings nor beliefs are in the head. Most philosophers, it seems, have accepted their argument. Putnam explained that a subject's grasp of an expression's meaning is often unable to fix its reference, she needs help from her natural and social environment; and Burge explained that having a belief, even in the de dicto sense, is really a relational property which may change when merely the implicit relatum, the linguistic community, changes.

To accept this, however, does not necessarily mean to accept all the antiindividualistic consequences Burge has drawn from these insights. On the contrary, these consequences have met much more reluctance. Many share the view, and we do as well, that there must be something in the head, not only a brain, but also a mind, indeed a mind with internal or intrinsic representational or semantic properties. This view was also backed up by arguments, having to do mainly with psychological explanation and the causation of individual behavior on the one hand and self-knowledge on the other. Of course, these arguments have been disputed; but the dispute has not shattered our prejudice.[1] Here, we would simply like to presuppose the correctness of this view without any further comments.

Thus, all those sharing the prejudice set out to characterize what's in the head, i.e., so-called narrow contents. Now, narrow contents are rather expressed by, or

[1] We had and have good company: Loar (1986), Fodor (1987, ch. 2) (though ch. 4 apparently got the upper hand in the end – cf. Fodor 1994), Perry (1988), Block (1991), Lewis (1994), Chomsky (1995), Chalmers (1999), and others. Moreover, most of cognitive science certainly sees itself as an individualistic enterprise.

associated with, whole sentences. But sentences are composed of parts, basically a singular and a general term, and hence narrow contents seem to be composed in the same way. We reserve here the term "concept" for those entities which a subject expresses by, or internally associates with, singular and general terms (and maybe other expressions, too). Having a concept is hence *defined* to be an internal, non-relational property. In the absence of a generally agreed usage of the term "concept" this stipulation is certainly legitimate, and it is often made. The terms "narrow content" and "concept" thus stand for the same thing; the only difference, which we do not strictly observe, however, lies in the associated kinds of expressions.

For internalists like us the existence of concepts and narrow contents is thus beyond doubt; the question is rather a constructive one: how precisely to conceive of them? This is, as the title indicates, the topic of this paper. However, the offers are so far rather problematic than impressive. We shall refer to two major options in the sequel:

First, the dominant view concerning the mind-body problem has been, and perhaps still is, functionalism, which says that internal mental states are functional states, i.e., to be individuated by the place they occupy within a large functional net spanned between perceptual input and behavioral output. Insofar mental states have narrow content, their content is then also to be characterized in a functional way. This gave rise to the program of so-called conceptual or functional role semantics which may thus be conceived as an attempt to establish internalism.[2]

Second, one may build upon the epistemological reinterpretation of Kaplan's character theory which was not fully endorsed by Kaplan (1977), but acquired prominence through Fodor (1987, ch. 2), though it is first recognizable in Stalnaker (1978)[3] and Perry (1977). According to the character theory, semantics has to recursively specify a character for each expression assigning to it its extension relative to a context (of utterance) and an index (or point of evaluation). And according to the epistemological reinterpretation, the diagonal of the character of an expression

---

[2] Cf., e.g., Field (1977) and Block (1986). Functional role semantics is not necessarily individualistic, though. It is ambiguous between short-arm and long-arm functional roles; cf. Harman (1987).

[3] Stalnaker is certainly not an internalist, as his (1989) and (1990) clearly show. However, his (1978) may well be interpreted as making more internalist sense of the character theory as Kaplan did.

represents the cognitive significance of, or the concept associated with, this expression.[4]

Here we shall pursue only the second approach via the epistemologically reinterpreted character theory. Our main reason is that functional role semantics failed to give a clear and precise theory of how concepts and narrow contents build up in a recursive way. By contrast, the character theory has a clearly specified formal structure which is easily connected with linguistic semantics; in particular, characters combine recursively in much the same way as intensions do in intensional semantics. Prima facie, these formal virtues are overwhelming reason for our choice (although we are well aware that formal structure alone does not determine its interpretation).

However, there are difficulties with the character theory as well. If one considers their interpretational questions, two serious problems emerge, as Schiffer (1990) and Block (1991) have forcefully made clear. The first problem is that the character theory seems to be either inadequate or superfluous. Schiffer argues that the character theory cannot avoid having recourse to functional roles or states. But then it seems to be only a detour; one could have explained narrow contents rather by directly appealing to functional roles. We call this Schiffer's problem. The second problem set up by Block is that the character theory can apparently take only one of two inadequate forms. Either it must specify narrow contents by reference to linguistic expressions themselves, i.e., fall prey to syntacticism. Or it must specify narrow contents in a profoundly holistic way, i.e., fall prey to an unacceptable degree of holism. We call this Block's dilemma.

Thus it seems that the character theory, whatever its formal virtues, cannot get off of the ground unless it offers some good response to these challenges. This is, more specifically, the task we want to address here. We tackle it in four sections: section 2 explains the epistemologically interpreted character theory and its problems in more detail, section 3 presents the solution we want to propose, section 4 explains it in a bit more detail, and section 5 explains that this solution indeed avoids Block's dilemma as well as Schiffer's problem.

2. The Problems Specified

---

[4] Perhaps one should also mention the very dense account of Lewis (1986, sect. 1.4) which is related to all three approaches mentioned, but not identical with any of them; to consider it seriously would, however, require a separate discussion.

We cannot go on after this rough and general introduction without referring to some specific statement of the epistemologically reinterpreted character theory. Let us therefore briefly look at the statement it receives in Haas-Spohn (1995) where it is dealt with in book length. We shall see that her account is also susceptible to the two problems just mentioned, but this will pave the way for improvement.

What is a character? A character is a function assigning to each possible context of utterance, or context for short, an intension, which is, in turn, a function from points of evaluation, or indices for short, to extensions. Equivalently, the character of an expression is a function assigning to each context and index the extension the expression has at this context and index. The characters of complex expressions build up recursively in the way we are used from intensional semantics.

We take a possible context c to be just a centered world, i.e. a triple $<s_c,t_c,w_c>$ such that the subject s_c exists at time t_c in the world w_c and may (but need not) utter the relevant expression. A possible index i consists of all items which may be shifted by operators of the given language; here it will suffice to put only a possible world w_i into the index i.[5]

Sentences, in particular, are true or false at contexts and indices, according to their character. This entails a notion of truth at a context simpliciter: A sentence is true at the context c if and only if it is true at c and the index which consists of the context world w_c itself. The function assigning to each context the truth value the sentence has at the context is called the diagonal of the sentence. Similarly we may define the diagonals of other expressions. Note that this definition works only on the condition that for each item of indices there is a corresponding item of contexts and that our definitions meet this condition.

Now we can say what the epistemological reinterpretation of the character theory is supposed to be. Basically, it just consists in considering possible contexts at the same time as possible doxastic alternatives of some subject. Thus, what a subject believes is always to be in one of the contexts of a certain set of contexts. And if a subject believes a sentence to be true, she believes to be in a context in which the sentence is true; that is, the sentence's diagonal is a superset of the set of the subject's doxastic alternatives.[6] All this agrees well with the characterization of contexts as centered worlds since centered worlds are known to be needed for the representations of beliefs de se and de nunc.[7]

---

[5] Cf. also Lewis (1980).

[6] Here, and elsewhere, we do not distinguish between a set and its characteristic function.

[7] Cf. Lewis (1979) or Haas-Spohn (1995, sect. 2.2-3). But see Spohn (1997, sect. 9 and 10) for arguments that doxastic alternatives need a variable assignment as a further component.

Now, to be a bit more specific, consider a certain natural language L like English and some referring expression α of L; one best imagines α to be a name like "Aristotle" or a one-place predicate like "water", "table", or "red".[8] Then Haas-Spohn (1995, pp. 99, 136, 150) explained the (*objective*) *character* of α in L in the following way:

> $||\alpha||_L$ (c,i) = the object or the set of objects at the index i which is the same or of the same kind, i.e., has the same essential properties as the object or the objects from which the usage of α in the language L originates in the context c.[9]

The crucial term is here "the usage of α in L". In the context c it stands for the whole communicative pattern in c associated in L with the expression α. However, what is essential to this pattern are not all of its ramifications it actually has in the context, but only the methods of identifying or recognizing the reference of α which are available to the community speaking L. These methods may be those of Putnam's experts for gold as opposed to the laymen, or those of Evans' producers of a name who are acquainted with its bearer as opposed to the consumers of the name[10], or indeed those of almost everybody in the case of chairs and tables in which nobody has privileged knowledge. Thus, such usages are in principle well described in the relevant literature.

Two points are important about such usages as conceived by Haas-Spohn (1995). First, the expression α itself is not essential to its usage; the very same usage may be associated with another expression as well. This entails in particular that different languages may have the same usage of different expressions; this is crucial for their translatability.[11]

Secondly, the extension, the object or objects from which the usage originates, is also not essential to the usage; in different contexts or context worlds different objects may fit the same usage. In our world it is H_2O which fits the usage of "wa-

---

[8] We shall not address definite and indefinite descriptions and all kinds of indexicals and demonstratives since they involve a number of further problems which we better avoid.

[9] Obviously, the "i.e." is only justified if the essential properties are necessary and sufficient for individuating the object or the kind. This may be false. But at least it seems true that they are necessary and that nothing else (except the pure thisness) achieves the individuation. So we may ignore the point in the following.

[10] Cf. Putnam (1975, pp. 235ff.) and Evans (1982, ch. 11).

[11] In fact it is often the other way around. Translation *merges* the usages of different languages and thus *makes* them identical; cf. Haas-Spohn (1997, sect. 3). This is an insight which seems to put the indeterminacy of translation and related issues into a very different light.

ter". But for all we know, or have known 250 years ago, it may as well be XYZ from which our usage of "water" originates. Likewise, in the actual context world our usage of "Aristotle" originates from the actual Aristotle. But there may be another context world in which somebody else had the same career as our Aristotle and has triggered the same usage of "Aristotle". In this way, then, the extension of α may vary with the context; and so, Kaplan's strategy of explaining the informativity of identity sentences between overt indexicals[12] may be carried over to hidden indexicals like "water" or "Hesperos". Hence, Haas-Spohn (1995) intends a usage to be something which may properly be called a communal concept which is internal to the relevant language community and does not change by merely changing the community's environment.[13]

The above explanation of the character of α in L is still incomplete; we have not yet specified its domain. Concerning the indices, we may assume that all indices or possible worlds belong to its domain. Concerning the contexts, however, the explanation presupposes that the very usage of α in the language L exists in the context; otherwise the character of α in L is undefined simply because there is no origin of the usage if there is no usage.[14] Thus, if we understand a language to be the collection of all the usages of its terms, the recursive explanation of the characters of its expressions works only for those contexts in which the language exists.

So, what is, finally, the *diagonal* of the expression α in the language L? It is the function which is defined for all contexts in which the usage of α in L exists and which assigns to each context the extension α has there according to its usage. This indicates the heavy burden the notion of a usage has to carry, and in view of this the explanations given may well be felt to be insufficient. We shall return to this point.

For the moment, however, we have to attend to another crucial point. Since usages are communal concepts which, as explained, summarize not what everybody knows, but what all together know about the relevant extensions, they are unsuited for describing concepts and narrow contents which are intended here to be internal to a given subject; the subject need not fully know about usages or communal concepts. This was indeed the basic point of Burge (1979): that a subject may have an

---

[12] Cf. Kaplan (1977) who explains in sect. XVII how "this = that" may be informative and refrains in sect. XXII to generalize the method to "Hesperos = Phosphoros" because he considers names to be absolute. According to the above explanation, however, names are *hidden* indexicals.

[13] Anti-individualists will find this notion of a usage to be question-begging, whereas we attempt here to provide individualistic foundations to such communal concepts.

[14] The counterfactual question what the origin would have been if the usage had existed does not generally make good sense.

incomplete or false linguistic understanding and still be amenable to de dicto belief ascriptions. So, how do we get down to the level of individual subjects?

A natural idea, indeed the one Haas-Spohn (1995) pursued, is the following; we repeat it here because it makes Block's dilemma very perspicuous: If a subject's knowledge of her own language may be incomplete, and indeed severely incomplete without clear lower boundary, then, it seems, we have to completely abstract away from such knowledge and to add it again for each subject according to her individual measure.[15] But what survives such abstraction? It seems the only thing we can hold fixed is the knowledge of the grammar, i.e. of the (purely morphologically conceived) words and their ways of composition. Thus we end up with what Haas-Spohn (1995, sect. 3.9) defines as *formal characters* which belong to a grammar G, the syntactic skeleton of a natural language:

$\|\alpha\|^G(c,i)$ = the object or the set of objects at the index i which is the same or of the same kind, i.e. has the same essential properties as the object or the objects from which the usage of α in the context language l_c originates – which is the language of s_c at t_c in w_c and has the grammar G.

In continuation of the parallel, the domain of a formal character consists first of all indices and second of all contexts in which the subject of the context speaks a language with the expression α or, indeed, with the whole grammar G. From this *formal diagonals* are again easily derived.

Formal diagonals describe the minimal semantic knowledge accompanying the syntactic knowledge of the grammar. In order to know the formal diagonal of the expression α one merely needs to know the triviality that α refers to whatever it is used for in one's language.

Thus, formal diagonals have at least some features desired by the internalist. Insofar knowledge of grammar is internal, knowledge of formal diagonals is internal as well. Moreover, there is no problem of intersubjectivity; all subjects mastering the grammar G thereby master the same formal diagonals. However, if we identify concepts with formal diagonals, we clearly fall prey to syntacticism, the one horn of Block's dilemma, since the words themselves and only the words are essential to concepts so understood. This is an understanding which is intuitively too narrow

---

[15] This strategy and the quantification over possible languages it involves goes back to Stalnaker (1978). Thus, the formal characters to be defined immediately are our way of capturing the idea behind Stalnaker's propositional concepts.

and too wide at the same time. It is too narrow because it entails that speakers of different grammars must ipso facto have different concepts. And it is much too wide because any two persons associating whatever they want with the same word ipso facto have the same concept. By moving to formal characters we have therefore lost the two virtues usages or communal concepts seemed to preserve.

This is no surprise because we have so far realized only the first part of our strategy, the step of maximal abstraction. However, a subject has beliefs about usages in her language just as about any other empirical matter, and only these beliefs add substance to the formal diagonals. Hence, we have to take the second step and to enrich the picture by the subject's individual beliefs. Our first attempt to do so will turn out to be too coarse, but without it one cannot understand the later refinements.

For the representation of beliefs we propose to follow the standard line formalized in doxastic logic. There, a subject's doxastic state is simply represented as a set of so-called doxastic alternatives, her *belief set*, and each proposition which is a superset of the belief set is then believed in that state. This representation has well-known problems: it neglects the fact that beliefs come in degrees; it cannot account for mathematical, but at best for empirical beliefs; it seems to presuppose logical omniscience since it assumes propositions to be believed regardless of how they are expressed; and so on. However, rival accounts are beset with other and no less grave problems; we therefore stick to this representation.[16]

In order to understand it properly, however, one needs to get clearer about what a doxastic alternative is supposed to be. We already said that it is simply a possible context $c = <s_c, t_c, w_c>$. But what precisely does it mean that c is a doxastic alternative of a given subject s at a given time t in the world w? The usual explanation is that it means that all the beliefs of s at t in w do not exclude that s_c is s itself, t_c the present time t, and w_c the actual world w. This explanation is certainly correct, but not very helpful because it inherits all the ambiguities of the notion of belief of which it makes essential use; in particular, it leaves unclear whether or not a doxastic alternative can be understood in an individualistic way. The following explanation (cf. also Haas-Spohn 1995, p. 34) is more explicit:

> $c = <s_c, t_c, w_c>$ is a doxastic alternative of s at t in w iff the following
> holds: suppose that s would fully investigate the world w_c – where this
> includes that it may take the perspective of every individual in w_c during

---

[16] Spohn (1997), sect. 5, distinguishes two large families of representations of belief, the computational and the semantic ones, each of which has their characteristic problems, and he argues that from the perspective of a rationality theory of belief semantic representations have primacy – which we join here.

its entire existence (so far it has a perspective at all), that it may learn all the languages there are in w_c, that it may subject each part of the world w_c to any investigations and experiments it can think, etc. – and suppose further that s would make its most considerate judgment about w_c after this investigation; then the assumption that it is now s_c at t_c in w_c is compatible with the doxastic state of s at t in w. More briefly, $c = <s_c, t_c, w_c>$ is a doxastic alternative of s at t in w iff s' maximally experienced and considerate judgment about $<s_c, t_c, w_c>$ is just an extension and not a revision of s' doxastic state at t in w.

The counterfactual suppositions in this explanation are impossible to satisfy in various respects. However, the explanation is thereby not rendered incomprehensible; it just drives common counterfactuals to the extreme, as philosophers are used to do. The explanation conforms to the common idea to characterize a belief state (or any kind of partial assignment) by the set of all its maximal extensions. What it adds to this idea is to describe more fully what is involved in such a maximal extension: not a maximally consistent set of sentences of a given language, not a totality of facts with facts being individuated from some external point of view, not just a possible world in the sense of a maximally inclusive object, but rather our normal ways of belief formation using experience and judgment which are maximally expanded with respect to such a possible world.

It is philosophically highly important to have a clear picture of what is involved in such a maximal extension.[17] Let us point out three consequences for the moment:

First, the above explanation of doxastic alternatives does not entail that the subject has to be able to express her belief set in her own language; some context may not qualify as a doxastic alternative simply because it does not feel or look like the actual world in some inexpressible way. Second, the explanation leaves entirely open how the subject's belief set (and its supersets) relates to our external belief ascriptions couched in our language; this is a question full of incompletely understood intricacies which we may, and are well advised to, leave aside. Third – this is most important for the rest of the paper –, *to have a given belief set is an internal, non-relational property* according to our explanation. The fact that the context c is,

---

[17] Christopher Peacocke pointed out to us that a lot of idealization is involved in what the subject's judgment would be under such counterfactual circumstances (perhaps it has only headache then and no judgment any more) and that the standards of idealization are not easily explained, maybe only in a circular way. Cf., e.g., the analogous critical discussion of the dispositional analysis of meaning in Kripke (1982, pp. 22-37). This is a legitimate worry which, however, we cannot further address here.

or is not, a doxastic alternative of *s* at *t* in *w*, does clearly not depend at all on any-thing in *w* outside *s* at *t*; it merely depends on *s'* internal disposition at *t* in *w*. In any case, we find it obvious that *s'* belief set as characterized above is entirely unaffected by exchanging all H_2O by XYZ or *s'* linguistic community by a slightly different one which can change, in some other sense, *s'* water and arthritis beliefs into twater and tharthritis beliefs.

This characterization of a subject's beliefs will be used throughout this paper. The present purpose, however, is only a limited one, namely to carry out the above idea of abstracting away usages as they actually are and to substantiate the formal diagonals thus obtained by the subject's beliefs about these usages. This raises the problem of distinguishing her beliefs about the usages of her linguistic community from all her other beliefs. Since the problem appeared unsolvable (see, however, footnote 28) Haas-Spohn (1995, pp. 58f.) mentioned the possibility to restrict the domain of formal diagonals not to some selected beliefs, but to all beliefs, i.e. to the belief set of the subject. This has the desired effect that the subject's terms do not apply to any arbitrary objects so-called in the various contexts, but only to objects conforming to the subject's beliefs. In this way, the substance we have lost by in-troducing formal characters has returned to a subjectively relativized extent.[18]

Since formal diagonals as well as belief sets are internal, the restricted formal diagonals are internal as well; insofar they could qualify as concepts in the sense in-tended here. However, we are obviously stuck now with the other horn of Block's dilemma. If a subject's concepts are formal diagonals restricted to her belief set, any change in her beliefs changes her concepts; this is indeed an extremely holistic conception of concepts. Likewise, any two subjects are almost guaranteed to have different concepts; this makes communication and psychological generalizations seem like a mystery. Moreover, we have not get rid of the flaw that subjects having acquired different grammars are bound to have different concepts. All this seems unacceptable.

These considerations may suffice as a concrete exemplification of how Block's dilemma arises for a theory of internal concepts. So far, we have apparently avoided Schiffer's problem – individualistic functional roles nowhere entered the picture –, but we did so only by doing very badly on Block's dilemma. White (1982) has already anticipated a way out of that dilemma. His framework is quite similar to the one presented so far.[19] The domain of the partial characters he defines consists of

---

[18] Similarly, the propositional concepts of Stalnaker (1978) are not defined for all contexts what-soever, but only within the context set consisting of the common presuppositions.

[19] For more detailed comparative remarks see Haas-Spohn (1995, pp. 141f.).

what he calls contexts of acquisition which are pairs of a possible world and some functional state the subject acquires in the world. In order to avoid the emptiness of syntacticism White restricts the domain of the partial character of a given expression to equivalent contexts of acquisition having one and the same functional state as a component, and by associating a separate functional state with each expression White has prima facie avoided holism. However, these very sketchy remarks show already that it is the functional states which are doing the work here, and that the possible escape from Block's dilemma immediately leads into Schiffer's problem.

3. How to define concepts: a proposal

Should we give up, hence, trying to explain concepts and narrow contents via the epistemologically reinterpreted character theory? No, we have not yet tried hard enough. So far, we have considered only two extreme options: the minimal option that the concept a subject associates with an expression contains only the trivial belief that the concept refers to whatever the word is used for, and the maximal option that the concept contains all beliefs of the subject, in particular all the beliefs the subject has about the reference of the concept. This leaves open a huge range of middle courses which deserves inquiry.

Block (1991, p. 40) uses a nice picture by distinguishing between the lexicon entry and the encyclopedia entry of an expression. The encyclopedia entry of an expression corresponds to our maximal option; intuitively, however, it is rather the lexicon entry which corresponds to the associated concept. So, this is what we seek to characterize.[20] There seem to be two basic ideas, with ramifications, for driving a middle course towards an adequate notion of concepts:

One idea is to appeal to communal standards, e.g., to define a concept to contain just the social minimum of beliefs about its reference which is required for being recognized as a partner in communication; this is Putnam's idea of a stereotype. One may doubt whether such social standards of semantic competence exist in a salient way; but if they do they certainly provide a useful notion for some purposes.

However, the idea seems to be the wrong one for our present purposes. If concepts are explained with reference to communal standards, then all competent spea-

---

[20] The task reminds of the task of explicating the analytic/synthetic distinction which is confronted with the objections so forcefully initiated by Quine; indeed, Block's despair of drawing his distinction in a more precise way may be seen as a late effect of Quine's negative conclusion. However, this conclusion is still contested; moreover, the two tasks are not identical. So we should not be discouraged by these objections.

kers in the group have automatically the same concepts. This seems unwelcome for individualistic psychology which should presumably allow for more conceptual variation across subjects. There is a further decisive objection which will be explained later on because it applies to other proposals as well.

The other basic idea, which appears sufficiently individualistic, is to appeal to the subject's recognitional capacities, i.e., to define a subject's concept of an object or a property to consist of those features on the basis of which the subject recognizes the object or instantiations of the property. What this means, however, depends crucially on what one takes here recognition to mean.

There is space for interpretation since the strictest and simplest understanding of recognition does not work at all. The strictest understanding would be to say that a subject is able to recognize an object if and only if, under all possible circumstances, it is always and only the object itself which the subject takes to be the object. This is much too strict because we are hardly ever able to recognize objects in this sense; there are always circumstances under which we mistake the best known objects and even ourselves. Recall also how absurdly narrow Russell's acquaintance relation turned out to be. The same holds, mutatis mutandis, for the recognition of properties. Recognition therefore needs to be understood in a much weaker way. There are various options from which varying notions of a concept result:

Our recognitional capacities may first be seen in our normal means of recognition which work, according to our familiarity, fairly reliable in the circumstances we usually encounter. Something like this comes presumably next to our intuitive notion of a recognitional capacity. However, it is utterly vague; and it entails that a recognitional capacity may be something very varied. For instance, I[21] may identify my son on the basis of my belief that he is somewhere in the crowd before me and that nobody else in the crowd is likely to wear the same kind of satchel. The example shows that my normal means of recognition use a lot of specific knowledge varying from situation to situation. By contrast, concepts are intuitively something more stable or invariant; it would certainly be strange to say that it belongs to my concept of my son[22] that he has that kind of satchel.

---

[21] There is a kind of pragmatic contradiction in writing a joint paper on subjective concepts, since the authors cannot use the stylistic device of taking themselves as an example. So we decided to use "I" in examples, and the reader is free to choose which of the authors he takes to be speaking.

[22] This is the first time that we slipped into talking of the concept of an object (or a property) instead of the concept associated with a given term – certainly a convenient, but also very dangerous slip, since it imports the de re/de dicto ambiguity and its proliferating consequences. Throughout, however, "the concept of a" is to mean the very same as the more clumsy phrase "the concept associated with 'a'", though we are perfectly aware, of course, that the two phrases should be distinguished.

There seem to be two ways of gaining more specificity here. One way is to narrow down a subject's means of recognizing an object or an instantiation of a property to the way the object or the instantiation looks to the subject under various circumstances. This line leads to what are called the subject's perceptual concepts. The other way is to raise a subject's recognitional means from those she normally applies to the best and most considered means which she usually cannot, or does not care to, apply.

So far, the available options are very sketchy. But there is an argument which tells even at this sketchy stage which direction to pursue more thoroughly. The argument is this: Clearly, we want our beliefs to be closely connected with our concepts; the contents of our beliefs should recursively build up from the concepts involved. For instance, in the primitive case of predication we would like to characterize the content of a belief such as that *a* is *F* in the form of a truth condition, i.e., as something like the function from contexts to truth values given as follows:

> A subject's belief that *a* is *F* is *subjectively true* in a context *c* if everything and at least something that conforms to the subject's concept of *a* in *c* also conforms to the subject's concept of *F* in *c*.

The strange term "subjectively true" indicates that some explanation is still missing. For the moment, however, we may ignore this and take the adverb "subjectively" as redundant; we shall return to the point at the very end of the next section.

Moreover one may stumble about the quantifier "everything and at least something". This is an attempt to do justice to the fact that there may be no or several objects in a single context *c* which conform to the subject's concept associated with the name "*a*". The attempt is certainly plausible. Finer analysis would show, however, that we run here into similar problems as does the counterpart theory of Lewis (1968) with non-identity-like counterpart relations. The corresponding proposal has proved there to be insufficient, and more complicated solutions are required.[23] But we need not dwell on this point; our attempt will do for the rest of the paper.

The crucial point of the truth condition is that it seems to yield inadequate results when it is based on anything else but the subject's best and most considered means for recognizing *a*. For instance, if the subject's concept of *a* would consist in some communal stereotype of *a*, the subject could possibly believe that *a* does not satisfy its stereotype or that many things different from *a* satisfy this stereotype, and then

---

[23] See Hazen (1979) and Kupffer (1999, ch. 3 and 4).

the above truth condition assigns truth or falsity to the belief that *a* is *F* in contexts in which the subject would intuitively not count it as, respectively, true or false. The same holds in the case where the subject's concept of *a* consists of the criteria normally used to recognize *a*. Again, it seems possible that the subject knows or believes in a given situation that *a* does currently not have the features normally used for recognizing *a* or that things different from *a* satisfy the criteria normally used for recognizing *a*, and then the above and the intuitive truth condition for the subject's belief that *a* is *F* diverge again. The only way to avoid this discrepancy seems to be to base the subject's concept of *a* on her best means for recognizing *a*, as we have proposed. The same holds, mutatis mutandis, for the property *F*.[24]

One may fear, though, that the best recognitional means available to a subject with respect to an object *a* or a property *F* come close to what Block called an encyclopedia entry. Should the subject not optimally use *all* her beliefs concerning *a* or *F* for recognizing *a* or an *F*? The answer is decidedly no. There are, for sure, many possible contexts in which the subject would recognize something as *a*, though it lacks there many properties the subject believes *a* to have; the subject has her ways, whatever they are, of distinguishing contexts which contain *a*, but with other than the believed properties from contexts which do not contain *a* at all. This is a crucial assertion without which the rest of the paper would not make any sense.

The following explanation captures this subjective distinction, or the subject's best recognitional means, or indeed the subject's concepts in a more explicit way:

> Let α be a name or a one-place predicate and $@ = \langle s,t,w \rangle$ the actual context (which may be any context). Then the concept $\beta_@(\alpha)$ which s associates with α at t in w is the function which assigns to each possible context $c = \langle s_c, t_c, w_c \rangle$ the set of objects in w_c which, according to s' judgment at t in w, might be the object, or instantiate the property, denoted by α in $@$.
>
> Or to spell out the phrase "according to s' judgment" in analogy to the above explanation of doxastic alternatives: $x \in \beta_@(\alpha)(c)$, or x is a *doxastic counterpart* in c of what α denotes in $@$, iff the following holds: suppose that s would fully investigate the world w_c – where this includes that it may take the perspective of every individual in w_c during

its entire existence (so far it has a perspective at all), that it may learn all the languages there are in w_c, that it may subject each part of the world w_c to any investigations and experiments it can think, etc. – and suppose further that s would make its most considerate judgment about w_c including x after this investigation; then the assumption that x is the object, or instantiates the property, denoted by α in @ is compatible with s' judgment at t in w about what is denoted by α in @.

This may look imperspicuous to some and trivial to others. But it will unfold its meaning and power in the following explanations and arguments.

4. Explanations

The final section will argue that this notion of a concept indeed helps the internalist against Schiffer's problem and Block's dilemma. This section is devoted to three kinds of explanations: some remarks about the features concepts have according to our definition, an explanation that the title of this paper is indeed appropriate, and a clarification of the relation of this definition to the proposals discussed in section 2. So, let us first explain five more or less expected and instructive features of concepts which are entailed by our definition:

(1) Concepts are usually not egocentric. Thereby we mean that, usually, things can conform to one's concept associated with α in a context c without there being anything in c which could be oneself. Hence, insofar modes of presentation and acquaintance relations have been thought to be usually egocentric, they are not concepts in the above sense.

(2) To which extent is the look, sound, or feel of things important for their conforming to one's concepts? It depends. Often it is conditionally important. Consider my concept of my son. Clearly, there could be many possible things in possible contexts which look perfectly like my son without possibly being my son according to my concept of him. Conversely, however, something could hardly be my son according to my concept without looking very much like him. Hardly! Of course, my son could look very different from his present look, not only actually, but also according to my concept of him. But if I encounter in a possible context c such a differently looking object, it could only be my son if there is something in the context c explaining why that object started or emerged to deviate from my son's look so well known to me. In this sense, the look of my son (the sound of his voice, etc.)

is a conditional part of my concept of him. In a similar way, the look of species, substances, and also individual things is a conditional part of my concepts of them; for instance, no doxastic counterpart of the black ball-pen in my drawer could be red during its entire existence. But there are other cases as well. It seems, for instance, that the look of things is not essential for their conforming to the concept I associate with the word "table"; what is essential is only what is done with the things in the relevant context. If there are culturalized beings in the context which use a given object only for sitting down *at* it, then that object counts as a table according to my concept, even if it never looks like a table; and conversely if something looks like a table, but is only used as something else, say, for sitting *on* it, then my concept does not count it as a table, but, say, as a seat.

(3) To which extent does the place of objects enter into one's concepts of them? Again, it is very often conditionally important. According to my concept of him, my son could be (almost) anywhere in the universe. However, the context must then provide some plausible story how he got there; any object, however intrinsically similar to my son, could not be my son if it is far away from Earth, or Germany, for that matter, during its entire existence. The same holds for many concepts of many other objects; after all, most objects we know are on the surface of Earth. The same may even hold for predicates. One may think, for instance, that a species which develops somewhere else in the universe, but, as it happens, intercross with our tigers, does nevertheless not consist of tigers; if so, one's concept of tigers includes their emergence on Earth.

Hence, very many of our concepts are, so to speak, geocentric. This entails the question what my concept of Earth may be. It seems to be quite poor. According to my concept, at least, the history of and on Earth so richly known to me is highly contingent to Earth; almost any planet of comparable size, age, and composition revolving around a sun of comparable size, age, and composition in the Milky Way could be Earth. And, of course, my concept of the Milky Way is even poorer; it contains hardly more than the Milky Way being some spiral galaxy.

(4) Their causal origin is essential to many objects. This is also reflected in our concepts of them. For instance, nothing which is not procreated by us could be our son, and since I also believe so, my concepts of myself and my spouse enter into my concept of our son. The same holds with respect to ourselves and our parents. Of course, my concepts of our ancestors get soon very dim, still all of them are part of my concept of my son. In fact, my son could not exist without history being pretty much as it is. Thus, a lot I believe about history enters into my concept of my son. This makes for a perhaps unexpected richness of that concept. In the same vein, my

concept of things is quite poor when I know very little about their causal preconditions, as is the case, for instance, with Earth. In fact, what we just said about the conceptual role of location presumably reduces to the present point about causal origin. Our son could not be born outside Germany or Earth, unless we, or our parents, etc., travelled. The same holds, mutatis mutandis, for tigers and other kinds if their causal origin is essential for them.

(5) Do concepts involve social relations, are they mutually connected by communication? Yes, of course; there is a clear relation between the concept I associate with a certain expression and the concepts others associate with that expression, a relation which Putnam (1975) has described as division of linguistic labor. Consider my concept of an elm, to take one of Putnam's examples. Elms might exist without mankind; in such a context, the extension of my concept of an elm would alternatively contain elms, beeches, and, maybe, other deciduous trees, since I, by myself, cannot distinguish elms from beeches and, maybe, other trees. This may entail that my perceptual concept of an elm is the same as that of a beech, but it does not entail sameness of the two concepts in our sense. On the contrary, since I believe elms and beeches to constitute different kinds and since I am allowed to identify the various kinds of trees in that context, my concept of an elm has any one of these kinds as extension in this context, and my concept of a beech any other kind, though I do not know which.[25] In other contexts there is an even clearer difference in the extensions of the two concepts, namely in contexts in which there is a linguistic community which generally resembles my actual community as I know it and which I observe during my full investigation of these contexts to apply the term "elm" only to certain trees and not to others (to which I might have been inclined to apply it as well). Then there are two possible cases: Either, these applications of the term "elm" contradict my concept of an elm so flatly – say, the community applies it to coniferous trees – that I conclude that this could not be my linguistic community after all and that its judgment cannot help mine in this matter; in this case, my judgment is as bad and the extension of my concept of an elm as wide as before. Or the linguistic community in the context behaves like mine in every relevant respect and in particular with respect to the term "elm" so that I conclude that this community could be mine and that I may trust its judgment; in such a context, the extension of my concept of an elm is as narrow as the usage of the community and certainly different from the extension of my concept of a beech (though, of course, it would be

---

[25] This idea of alternative extensions in one and the same context is not mentioned in our above definition of concepts, because it entails additional complications. But it seems required in order to overcome the difficulties referred to in footnote 23.

compatible with my concept of an elm that this counterfactual community applies "elm" only to beeches).

In this way, the division of linguistic labor is reflected in subjective concepts. This entails in particular that referential and deferential aspects are often inextricably mixed in subjective concepts; the simple reason is that subjects often trust the judgment of their fellows more than their own. Of course, the degree to which semantic deference enters into subjective concepts may considerably vary. For instance, my concept of Indian deity, say, or of multiple sclerosis is so poor, that I would follow almost any opinion if it presents itself as a consistent opinion of our experts; in such cases the deferential component of concepts is overwhelming. By contrast, I may be convinced to know more or less as well as all others what tables are; in such cases my own most considerate judgment is hardly helped by others, and the deferential component of my concept of tables largely vanishes. However, it seems that it never vanishes completely in concepts associated with linguistic concepts; it seems present even in the concept associated with the predicate "x looks red to me" in the phenomenal reading.[26]

Let us next explain the appropriateness of the title of our paper. Our aim was, we said, to drive a middle course between the minimal and maximal option both of which we found to be inadequate. So, which beliefs are contained in the concept a subject associates with the expression α if they are more than that α has an extension and less than all beliefs about that extension? Our title gives a simple and informative answer which runs as follows:

G is an essential property of a if and only if it is metaphysically or ontologically necessary that a is G, i.e., if nothing which is not G could be (identical with) a; for instance, being human or having the parents our son has are essential properties of our son. This is the common definition; it is full of niceties, which we better skip over, however. We can extend it to a relation between properties: G is essential for F if and only if it is metaphysically necessary that every F is G. For instance, being unmarried is essential for being a bachelor (though it is not essential for bachelors, no bachelor is essentially a bachelor), or consisting of hydrogen and oxygen is essential for being water.

Now, one may express our definition of concepts also in the following way: The concept a subject associates with "a" is the conjunction of all concepts G, or the strongest G, such that the subject believes that G is essential for a. Similarly, the

---

[26] This point is made already by Austin (1962) (see his magenta example on p. 113). Cf. also Spohn (1997a and 1997/98. sect. 5).

concept a subject associates with "*F*" is the conjunction of all concepts *G*, or the strongest *G*, such that the subject believes that *G* is essential for *F*.

When one compares this with the original definition, it is rather obvious that this is an equivalent characterization. Indeed, it is trivial in view of the fact that being identical with *a* is the strongest essential property of *a*, and being *F* is the strongest property essential for being *F*. The characterization would become more interesting if we would introduce restrictions on the metaphysical side, for instance by excluding identity from genuine properties and relations, or on the epistemological side, for instance by postulating that all concepts are ultimately qualitative in some suitable sense. We would in fact be prepared to make such restrictions, but it would take us too far to go into this issue.

Let us rather briefly check whether this characterization agrees with the five features of concepts just noted. What we said about beliefs about causal origin often being part of concepts, fits perfectly, of course. We also stated that the look of objects or kinds often enters into our concepts of them. But, as a rule, looks are certainly inessential. However, we qualified our statement. Often, the look of an object or of a kind displays its essence provided that it is allowed to unfold its normal look; and it is only this complex concept which is part of the concept of an object or a kind. Finally, what about the deferential component of concepts? What others believe about an object or a kind is certainly not essential to it. Sure, but insofar we trust others, we believe what they believe, and if we take the experts' beliefs about essences as trustworthy and they believe essences to be such and such, we also believe these essences to be such and such. So, the present characterization agrees well with the earlier observations.

Viewed in this way, is our proposal for defining concepts not a familiar one? We are not aware of this. To our knowledge it is mentioned only in Block (1995, sect. 4) where he attributes the view to two lines in Fodor (1987), discusses it on one page, and dismisses it right away. The paper is about one example, namely the concept a teenie associates with the word "grug" which denotes beer in his assumed dialect. The teenie knows very little about grug; he knows, e.g., that it comes in six-packs. Block simply assumes that this belief is part of the teenie's concept of grug, and he is certainly right to claim that it is not essential to grug to come in six-packs. But Block has a different notion of concepts here. His notion seems to be the one we have already mentioned, namely that concepts are something like normal means of recognition, and the teenie's poor means of recognizing grug refer to its packing. However, we have already argued that this is not the best notion of a concept, and indeed we would flatly deny that the belief that grug comes in six-packs is part of

the teenie's concept of grug. So, as we say, there does not seem to be much of a discussion of the line of thought we are proposing here.

Let us finally explain how the present definition of concepts relates to the two kinds of diagonals in section 2; in a way this is for our own records, but it also illuminates our definition in some important respects.

Recall that the objective diagonal of an expression α in a given language L was the function which is defined for all contexts in which the usage of α in L exists and which assigns to each such context the extension α has in that context; in particular, it assigns to the actual context the actual extension of α in the familiar sense. The formal diagonal of α as part of a grammar G did the very same. The only difference was that the formal diagonal has a larger domain consisting of all contexts in which the grammar G exists which might be realized in different languages; the specific language having the grammar G and spoken in the context was implicitly fixed by the subject of the context.

By contrast, concepts as defined above are functions defined for any contexts whatsoever (though most of them will be alien so that we find hardly anything in them conforming to our concepts). The all-important question is therefore: do they agree with diagonals on their common domain? And the crucial answer is: conditionally yes!

Imagine that the subject having the concept associated with α may investigate a context in which her actual language L exists with its very usage of α. Then we may expect that the subject judges that the context's linguistic community might indeed be her own, at least as far as the usage of α is concerned, that the community as a whole is more competent than she himself with respect to α, and that she should therefore follow the community's final judgment. In this way, semantic deference enforces an agreement of the extensions of the subjective concept and the public usage. This expectation may be wrong, however. The subject may also find the usage of α in L, as compared with her concept of α, so strange that she (falsely) concludes that this is not her actual linguistic community rather than that she is the victim of a severe misconception; only then the subject's judgment about α or α's may deviate from that of the community. Semantic deference is thus an important ground for the subject's agreement with the community.

Note, however, that the extension of the subject's concept and the experts' concept may agree even in a context in which the subject does not defer to the experts or in which none of them exists at all. The context may be kind, so to speak: The subject may believe that her concept refers to a single natural kind, though she

knows very little about that kind and the experts may know very much. But suppose the context provides only one natural kind which conforms to the little the subject knows about it. Then only this kind is in the extension of the subject's concept, just as in the extension of the much better informed communal concept. And again the two agree.

Let us illustrate this with the two standard examples "water" and "arthritis". The actual extension of the concept Oscar presently associates with "water" consists of all H_2O and nothing else, even if Oscar knows nothing about chemistry. The primary reason for this is that Oscar believes water to be a natural kind amply instantiated in his environment and that there is no natural kind in the actual world which he would confuse with water in his maximally informed and considerate judgment. For the same reason, the actual extension of the concept which Oscar's ancestor associates with "water" 250 years ago also consists only of H_2O. Semantic deference becomes relevant in a context in which there are two kinds of liquid which Oscar by himself might take for water. If he finds there a linguistic community which might be his own and which acknowledges only one liquid to be water, then his subjective concept has only this liquid as extension in this context. If he finds there a community which is as indiscriminate as he is, then both liquids constitute alternative extensions of his concept in this context. And the same is true, if he finds there two trustworthy communities as in Putnam's twin earth story to which he might defer, which, however, refer to different liquids. All this shows that there is, on the one hand, a lot of agreement in the extensions of various subjective and communal concepts at different times, and that, on the other hand, the differences among all these concepts show in suitable counterfactual contexts.

What about the actual extension of the concept Fritz associates with "arthritis"? This case is more delicate. If Fritz' belief that arthritis is an ailment which may occur in the thigh is conditional on the agreement of his community, then he will defer also to his actual community which denies this, and the extension of his subjective concept will congrue with that of the communal concept. If Fritz' belief about the essence of arthritis is unconditional, then he will not acknowledge the actual community to be his community, and his judgment will be unassisted. In this case the subjective and the communal concept may diverge. But it may also be the case that after fully investigating the actual world his judgment is that arthritis occurs only in the joints, because the investigation shows that there is a natural kind of appropriate ailments in the joints, but none which extends to thighs.

If we return to comparing our definitions of diagonals and concepts, a further important difference emerges. Concerning objective diagonals in L we said that the

extension of α consists of the object(s) with the same essential properties as the object(s) from which the usage of α in L *originates*. This is a clear heritage of the causal theory of reference on which Haas-Spohn (1995) relies; and therefore the usage of α in L had to exist in the context c in order for α's extension being defined in c. By contrast, the extension of α in c according to a subjective concept consists of the object(s) in c *conforming* to the concept. There, the causal aspect has disappeared and with it the restriction of the concept's domain. But how then can the two functions, the objective diagonal and the subjective concept, agree within their common domain?

The question does not really concern subjective concepts. It rather points to a tension in our notion of a usage. On the one hand, a usage has, we said, an extension only where it exists and has causes; on the other hand, we said that a usage is something like a communal concept internal to the community, and then objects in any world should be able to conform to the usage. The tension hides a confusion of metaphysical and epistemological matters. Metaphysically, it is inessential to most objects or kinds of objects that they are actually conceived of, i.e. that they cause an intelligent species to form specific concepts. We also believe this. So, if a (communal or subjective) concept assembles beliefs about the essence of its reference, such a causing does not belong to it. Within an epistemological perspective, however, the belief in such a causing is an a priori company of the concept; any community (or subject) which acquires a concept associated with some term thereby acquires the belief that the concept and the term refer to the object(s) in confrontation with which the concept was acquired. This is so at least to the extent in which a causal theory of reference applies. Hence, insofar the concept or the usage exists in a context, its extension is described by our objective diagonals in the same way as by our definition in this section.[27]

Similar remarks apply to the comparison of formal diagonals with the concepts of a subject s. Again, the two functions agree for those contexts in which the subject s_c of the context speaks a language with the grammar G (otherwise the context would not be in the domain of the formal diagonal) and in which the community

---

[27] This does not seem to agree with Putnam's Twin Earth stories. Suppose the English and the Twin English community exist in the same world and associate the same internal communal concept with "water", as Putnam suggests. Everybody agrees that the concept has different extensions in the two communities, namely, respectively, H_2O and XYZ. But we seem to have to say that both extensions consist of all H_2O *and* all XYZ since both, H_2O and XYZ, conform to this concept. This is not so, however. According to the concept, its extension in this world consists *either* of all H_2O *or* of all XYZ. We don't know of which; if our extension is H_2O, theirs is XYZ, and vice versa. The decision is made by the context which, by being a centered world, says which community is in the center.

speaking that language in that context could be s' community as far as s believes (so that s can defer judgment to the community). And the causal ingredients in the formal diagonal give an a priori condition on the extension within this common domain and do thus not constrain the extension as specified by the concept. What about the fact that the formal diagonal of α essentially involves the expression α itself, whereas the concept associated with α does not? Again, this does not create a difference within the common domain, since it is a further a priori condition on the concept associated with α that it is associated with α.[28]

All this enables us, at last, to explain what is subjective about our above truth condition of a subject's belief recursively built up from the concepts involved. In contexts in which the subject can defer her judgment to the surrounding community, there is nothing subjective about the truth condition. To that extent the truth condition is intersubjective and indeed objective (since the relevant contexts may be fully investigated, with no space for error left), i.e. to that extent the subject's belief that a is F is subjectively true if and only if the sentence "a is F" of her language is true. The difference shows in other contexts without an appropriate community. There, the poverty of a subject's concepts and a large divergence from the concepts of her community may come to the fore. Hence, there is a difference in subjective and objective truth conditions, as it should be, but not a critical one.

5. Individualism Rescued?

To which extent does the proposal explained in the previous sections promote the individualist's project? Four points are worth discussing:

---

[28] A further thought which we owe to Manfred Kupffer: In section 2, when restricting formal characters and their diagonals to the subject's belief set, we have, it may have appeared, given up too soon on distinguishing the subject's beliefs about her linguistic community and its meanings from her other beliefs. We may conceive of a doxastic alternative c in a richer way, consisting not only of an indiviudal s_c, a time t_c, and a world w_c, but, given s_c has the language l_c at t_c in w_c, also of the objective character function $\|.\|_{l_c}$. A subject's belief set then consists only of doxastic alternatives thus enriched (because she believes to have a language), and only of those enriched by a character function which might be, for all she believes, the character function of her own language. In this way, we may explicitly distinguish the subject's beliefs about the meanings of her language, and we could explain the subject's concepts by restricting the formal character and its diagonal not to the subject's belief set, but only to the larger set of doxastic alternatives enriched by a suitable character function.

There is no conflict, though. Our previous considerations rather imply that concepts thus explained (= the diagonals of the larger set of enriched doxastic alternatives) agree with concepts in our sense on their common domain; the difference is only that concepts in our sense have, desirably, a wider domain. However, the agreement supports our case; it is nice to see that this different line of thought arrives essentially at the same result.

(1) Our proposal provides something of a definition at all; this is more than one usually finds in the literature. It does so mainly because it firmly rests on the epistemologically reinterpreted character theory which has by far the best formal grip on these matters. This theory also provides concepts and narrow contents with a recursive structure following (essentially – see the qualification giving rise to footnote 19) the recursive structure of the expressions with which they are associated. No negligible advantages.

(2) Again, the crucial point is, of course, that concepts are individualistic according to our definition in the same way as belief sets are individualistic according to our definition; to have a concept is an internal, non-relational property. Which function from contexts to extensions a subject associates with an expression depends solely on its internal cognitive state, does not presuppose the existence of anything outside the subject, and does in no way change when the environment of the subject changes without affecting her internal state. For instance, Oscar, Twin Oscar, and the (appropriate) Swamp Man would display precisely the same dispositions, they would respond in our huge counterfactual test in precisely the same way, and they have hence precisely the same concepts. Of course, agreement will usually be at most partial; the Frenchmen may associate with ”Londres” the same concept as I associate with ”London”, while our concepts diverge elsewhere.

Defining concepts and contents in a narrow way is one thing, describing them is another. We have to build a theory how concepts combine to contents, how contents become attitudinized, how perception acts upon the attitudes, how the attitudes result in action, and so on. Thus we say how this huge array of counterfactuality integrates into factuality, and conversely this makes this array accessible from the facts we observe on the street and in the lab; of course, theory is vastly underdetermined by the data, here as everywhere. We have not said a word about how this theory goes and which ways of describing all these internal entities go along with it; but this would clearly be a different task which does not impair the internality of its starting point.

On the contrary, spelling out this theory would fully display the strategy of individualism which consists in defining the momentary states (i.e. state types) of subjects in such a way that they are connected with past and future only through causal laws. By contrast, externalists take such connections to be part of the identity conditions of these states, by defining them either as being caused in a specific manner, as does the causal-information theoretic account of Dretske (1981), or as dispositions or attitudes analytically tied to their manifestations or intentional objects – a false understanding of (most) dispositions, as Spohn (1997b) argues. Even the

functionalist is externalistically biased insofar he defines a mental state by its functional role, by its place in a causal net extending from past to future. In our account, however, the narrow mental states of a subject are not defined by their causal ancestry, but as dispositions which are only causally related to their actual manifestations; and it is only the envisaged rich theory which conjectures the functional role of these states; that role is not definitionally fixed to begin with. These remarks show at the same time that our proposal has not led us into Schiffer's problem; our proposal is so far independent of functionalist conceptions.

(3) The next question, then, is how we fare with respect to Block's dilemma. Here, it is clear that we have perfectly avoided the syntacticist horn of that dilemma. Which expression a subject associates with a concept is fully contingent and does in no way add to the identity of the concept. This entails in particular that members of different linguistic communities may nevertheless have the same concepts. Of course, the deferential component of a subject's concept makes reference to her own linguistic community, and this distinguishes concepts in unconnected languages. However, translation has the effect of merging the experts of the communities and thus of merging their usages or communal concepts, and thereby differences of subjective concepts due to deference vanish as well.

(4) Whether we are equally successful with respect to the holistic horn of Block's dilemma is less clear; this is the final point to be discussed at more length. We shall not attempt to clear up the term "holism"; there seems little agreement on its precise meaning. However, it is very clear that concepts as we have explained them are thoroughly interconnected. It would be extremely important to study the architectonics of concepts in detail – though nothing we can achieve here. We see no reason, though, to expect the conceptual connections to be unidirectional, i.e. that there is a set of basic concepts from which all the other concepts are defined step by step, as Carnap (1928), for instance, has tried to establish in an exemplary way. Rather, all kinds of circular dependencies among concepts are to be expected. Insofar concepts will certainly turn out to be holistic.

The essential reason for this holism is that ontology is holistic, in the first place. There is rich ontological dependence among objects and properties; we mentioned the example that many objects and maybe even properties ontologically depend on Earth, i.e., could not exist or be instantiated, if Earth would not exist. Hence, if essences are thoroughly intertwined, beliefs about them, i.e. concepts, will be so as well.

However, if we follow Block's and Fodor's definition of holism concepts as we have explained them are not holistic. Block (1991) says "that narrow content is

holistic if there is no principled difference between one's 'dictionary' entry for a word, and one's 'encyclopedia entry'" (p. 40). But the whole point of this paper was to propose such a principled difference! The lexicon entry for a word contains only one's beliefs about the essence of its reference, whereas the encyclopedia entry contains all other beliefs about the reference as well

The case is similar with Fodor (1987). What he says about holism does not exactly fit our present discussion. He there defines that "meaning holism is the idea that the identity – specifically the intentional content – of a propositional attitude is determined by the *totality* of its epistemic liaisons" (p.56). This does not exactly fit, first because Fodor addresses only the narrow content of propositional attitudes and not that of subsentential expressions, and second because the term "epistemic liaisons" refers to confirmatory of justificatory relations between propositions – something we have not touched at all. If, however, we straighten out the definition by taking the epistemic liaisons of a word to consist in the beliefs in which it occurs, we are back at Block's definition.

Let us look, hence, at what Fodor (1987) dubs the Ur-argument for meaning holism which runs as follows: "Step 1: Argue that at least some of the epistemic liaisons of a belief determine its intentional content. Step 2: Run a 'slippery slope' argument to show that there is no principled way of deciding *which* of the epistemic liaisons of a belief determine its intentional content. So either none does or they all do. Step 3: Conclude that they all do" (p. 60).

Fodor goes on to discuss three versions of the Ur-argument and tries to argue that in all three of them step 1 has erroneously been taken for granted. Given the above straightening out we have no quarrel with step 1, however. Rather, step 2 is faulty. There may be vagueness or indeterminateness in the beliefs about essences or perhaps even in the essences themselves. But there is no slippery slope.

However, it is not important whether or not concepts should be called holistic according to our definition; holism as such is not bad. The question is rather whether or not the unacceptable consequences for which holism is blamed in this area are avoided by our definition. Let us look at four such consequences.

A first bad consequence of holism appeared to be that belief change ipso facto meant conceptual change. This, however, is not so at all with our proposal. Take my concept of my son, again. I acquire new beliefs about him all day long and forget many old ones. But, according to our explanation, my concept of him has in no way changed in the last years; all the beliefs I have acquired or forgotten concerned contingent matters and did not add to, or subtract from, my beliefs about his essence. The same holds, say, for my concept of tables. Almost every day I learn

something about tables, for instance, at which places tablehood is instantiated. But my concept of tables is fixed since long.

A second bad consequence seemed to be that holism renders impossible intrapersonal and interpersonal psychological generalizations. This is an objection we never understood. Each individual constellation may be unique, but this does not prevent it from being subsumable under general laws. It was always clear that, strictly speaking, there is only one application for Newton's theory of gravitation, namely the whole universe. But this did not deprive it of its lawful character. Block (1991, p. 41) makes similar remarks to the effect that there is not really an objection here.

A third bad consequence of holism was said to be that it makes communication miraculous because the concepts of different subjects are almost guaranteed to differ, preventing them to understand each other. There are several remarks to be made about this point.

To begin with, we are not sure whether subjects need to have the same concepts in order to understand each other. It rather seems to be sufficient to know which matter the others talk about, i.e. to which objects and properties they refer. As long as this is secured, it does not do much harm when we have a different grasp of the objects and properties referred to; communication may also serve to assimilate the differing grasps. In this perspective, sameness of concepts is required only insofar concepts are constitutive for ontology. This may indeed be a relevant aspect in abstract realms; but we do not think it has much relevance in everyday matters.

Still, it would be good to know the extent to which we have the same concepts according to our proposal. The answer is a mixed one. Take my son again. I know his grand-parents, others don't. So, our earlier remarks imply that there are diverging concepts of him. Take Bill Clinton, by contrast. Most of us know him just from TV. Certainly, we have looked at TV at different times; hence, we believe different things about him. But there is no reason to assume that our concepts of him differ in any way; we believe quite the same about Clinton's essence. Take tables. Again, there is no reason why our concepts of tables should differ despite our differing beliefs about tables. If we compare the functions from contexts to extensions which we associate with the word "table", our guess would be that the variance keeps well within the range of vagueness of that word. Take elms, finally. Presumably, many of us are still roughly in the poor state Putnam describes. But some of us may have been ashamed of this and have informed themselves. Insofar, then, their concept of an elm differs from that of the rest. Hence, there is neither a guarantee nor an impossibility of agreement in concepts.

However, one should observe that there is considerable conditional agreement. We argued in the previous section that subjective concepts and objective diagonals agree on their common domain (if the relevant condition is satisfied). Since this holds for all subjects, we find the same conditional harmony among their concepts.

These remarks do not add up to a satisfactory discussion of the question how communication is possible on the basis of concepts as beliefs about essences. But we may tentatively conclude that there is no clear evidence at all for a serious objection to be forthcoming here.

The fourth and final bad consequence of holism seems to be what Fodor (1987, p. 102) calls the disjunction problem, which is the problem how error is possible – which it must clearly be – according to one's theory of meaning, content, or concepts. This problem arises in particular for the causal-information theoretic account of Dretske (1981), and in this way it is treated by Fodor (1987 and 1990, ch. 3 and 4). However, the problem of error also plagues holistic accounts. Suppose Fodor's Ur-argument quoted above would be sound. Then all the epistemic liaisons of a content which I believe, i.e. hold to be true, would be constitutive of that content. Now suppose I change these epistemic liaisons. Could this result in a different balance of reason for that content and even in a different judgment, e.g., that this content is really false? No, because it would be a new content which I would judge false; the old content would cease to exist. That is, the old content can exist only as held true. Similarly, if a concept is an encyclopedia entry in Block's sense. I believe all parts of that encyclopedia entry to be true. Now, for some reason, I want to change my mind and to discard some parts. Because they have proved wrong? No, we cannot put it this way. If I change my encyclopedia, I change my concepts, and my beliefs change content. So, again, I can put together my concepts only to form contents with a fixed truth assignment; all contents would be conceptual truths or falsehoods. These would be fatal consequences indeed.[29] Of course, I often err even by my own lights, and any adequate theory must be able to account for this.

It should be clear, however, from our comments on the first possible objection that our proposal has none of these absurd consequences and enables us to change our mind without changing our concepts. In particular, our explanation of concepts and our subjective truth condition for beliefs clearly allow us to have beliefs which are false by our own lights; our most considerate judgment may well falsify our actual judgment. There is no error problem for our account.

---

[29] In fact, this issue was first raised in relation to the account of the meaning of theoretical terms in Kuhn (1970), pp. 111ff. and 198ff., and, for instance, Feyerabend (1965).

So, to sum up: have we escaped the holistic horn of Block's dilemma? Our discussion does perhaps not firmly establish a positive answer, but it shows, we think, that the prospects for our proposal are bright – all the more so as it was clear that the syntacticist horn of the dilemma was definitely avoided and that there was no danger of stumbling into Schiffer's problem.

References

Austin, John L. (1962), *Sense and Sensibilia*, Oxford: Clarendon Press.

Block, Ned (1986), "Advertisement for a Semantics for Psychology", in: P. French, T. Uehling, H. Wettstein (eds.), *Midwest Studies in Philosophy, vol. X, Studies in the Philosophy of Mind*, Minneapolis: University of Minnesota Press pp. 615-678.

Block, Ned (1991), "What Narrow Content is Not", in: B. Loewer, G. Rey (eds.), *Meaning in Mind. Fodor and His Critics*, Oxford: Blackwell, pp. 33-64.

Block, Ned (1995), "Ruritania Revisited", in: E. Villanueva (ed.), *Philosophical Issues*, vol. 6, Atascadero: Ridgeview, pp. 171-187.

Burge, Tyler (1979) "Individualism and the Mental", in: P. A. French, T. E. Uehling jr., and H. K. Wettstein (eds.), *Midwest Studies in Philosophy, Vol. IV: Metaphysics*, Minneapolis: University of Minnesota Press, pp. 73-121.

Carnap, Rudolf (1928), *Der logische Aufbau der Welt*, Hamburg: Meiner.

Chalmers, David (1999), "The Components of Content", *Mind* 108, to appear.

Chomsky, Noam (1995), "Language and Nature", *Mind* 104, 1-61.

Dretske, Fred (1981), *Knowledge and the Flow of Information*, Oxford: Blackwell.

Evans, Gareth (1982), *The Varieties of Reference*, Oxford: Clarendon Press.

Feyerabend, Paul (1965), "Problems of Empiricism", in: R.G. Colodny (ed.), *Beyond the Edge of Certainty*, Englewood Cliffs: Prentice-Hall, pp. 145-260.

Field, Hartry (1977), "Logic, Meaning, and Conceptual Role", *Journal of Philosophy* 74, 379-409.

Fodor, Jerry A. (1987), *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*, Cambridge, Mass.: MIT Press.

Fodor, Jerry A. (1990), *A Theory of Content and Other Essays*, Cambridge, Mass.: MIT Press.

Fodor, Jerry A. (1994), *The Elm and the Expert*, Cambridge, Mass.: MIT Press.

Haas-Spohn, Ulrike (1995), *Versteckte Indexikalität und subjektive Bedeutung*, Berlin: Akademie-Verlag.

Haas-Spohn , Ulrike (1997), "The Context Dependency of Natural Kind Terms", in: W. Künne, A. Newen, M. Anduschus (eds.), *Direct Reference, Indexicality and Propositional Attitudes*, Stanford: CSLI Publications, pp. 333-349.

Harman, Gilbert (1987), "(Nonsolipsistic) Conceptual Role Semantics", in: E. LePore (ed.), *New Directions in Semantics*, London: Academic Press, pp. 55-81.

Hazen, Allen (1979), "Counterpart-Theoretic Semantics for Modal Logic", *Journal of Philosophy* 76, 319-338.

Kaplan, David (1977), "Demonstratives. An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals", in: J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press, 1989, pp. 481-563.

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

Kupffer, Manfred (1999), *Counterparts and Qualities*, Dissertation, University of Konstanz.

Kuhn, Thomas S. (1970), *The Structure of Scientific Revolutions*, Chicago: University Press, 2nd ed.

Lewis, David (1968), "Counterpart Theory and Quantified Modal Logic", *Journal of Philosophy* 65, 113-126.

Lewis, David (1979), "Attitudes de dicto and de se', *Philosophical Review* 88, 513-543.

Lewis, David (1980), "Index, Context, and Content", in: S. Kanger, S. Öhman (eds.), *Philosophy and Grammar*, Dordrecht: Reidel, pp. 79-100.

Lewis, David (1986), *On the Plurality of Worlds*, Oxford: Blackwell.

Lewis, David (1994), "Reduction of Mind", in: S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell, pp. 412-431.

Loar, Brian (1986), "Social Content and Psychological Content", in: R. Grimm, D. Merrill (eds.), *Contents of Thought*, Tucson: University of Arizona Press, pp. 99-110.

Perry, John (1977), "Frege on Demonstratives", *Philosophical Review* 86, 474-497.

Perry, John (1988), "Cognitive Significance and New Theories of Reference", *Noûs* 22, 1-18.

Putnam, Hilary (1975), "The Meaning of 'Meaning'", in: H. Putnam, *Philosophical Papers, Vol. 2: Mind, Language and Reality*, Cambridge: Cambridge University Press, pp. 215-271.

Schiffer, Stephen (1990), "Fodor's Character", in: E. Villanueva (ed.), *Information, Semantics, and Epistemology*, Oxford: Blackwell, pp. 77-101.

Spohn, Wolfgang (1997), "Über die Gegenstände des Glaubens", in: G. Meggle (ed.), *Analyomen 2. Proceedings of the 2nd Conference "Perspectives in Analytical Philosophy". Vol. I: Logic, Epistemology, Philosophy of Science*, Berlin: de Gruyter, pp. 291-321.

Spohn, Wolfgang (1997a), "The Character of Color Predicates: A Materialist View", in: W. Künne, A. Newen, M. Anduschus (eds.), *Direct Reference, Indexicality and Propositional Attitudes*, Stanford: CSLI Publications, pp.351-379.

Spohn, Wolfgang (1997b), "Begründungen a priori – oder: ein frischer Blick auf Dispositionsprädikate", in W. Lenzen (ed.), *Das weite Spektrum der Analytischen Philosophie. Festschrift für Franz von Kutschera*, Berlin: de Gruyter, pp. 323-345.

Spohn, Wolfgang (1997/98), "How to Understand the Foundations of Empirical Belief in a Coherentist Way", *Proceedings of the Aristotelian Society, New Series* 98, 23-40.

Stalnaker, Robert C. (1978), "Assertion", in: P. Cole (ed.), *Syntax and Semantics, Vol. 9: Pragmatics*, New York: Academic Press, pp. 315-332.

Stalnaker, Robert C. (1989), "On What's in the Head", *Philosophical Perspectives 3, Philosophy of Mind and Action Theory*, pp. 287-316.

Stalnaker, Robert C. (1990), "Narrow Content", in: C.A. Anderson, J. Owens (eds.), *Propositional Attitudes. The Role of Content in Logic, Language, and Mind*, Stanford: CSLI Publications, pp. 131-145.

White, Stephen L. (1982), "Partial Character and the Language of Thought", *Pacific Philosophical Quarterly* 63, 347-365.