

Intersubjective externalism*

Peter Pagin

April 16, 2006

Abstract

In this paper I present the idea of a kind of externalism different from what is usually considered in the internalism / externalism debate. It is intersubjective in nature, since it concerns representation determined by *relations* between cognitive subjects. More precisely, it concerns linguistic expression types. The basic idea is that the meaning of linguistic expression types often depends on external factors over and above any external factors that serve to determine the contents of the thoughts of speaker and hearer. In particular, this kind of intersubjective externalism can hold even *internalism* is true of thought content itself.

1 Introduction

In this paper I shall try to show that with a particular view of communicative success, a kind of content externalism results that is *intersubjective* in nature. In its weak form, it will not concern contents of thoughts of individual subjects, but the contents of linguistic expressions by means of which the subjects communicate. In its regular form, it is also set in relation to the contents of thoughts. I shall first explain what I mean by labeling content externalism (from now on I'll skip the 'content' qualification) 'intersubjective'.

Current debate on externalism and internalism is standardly concerned with *thought content*. An imprecise definition of internalism would be

(SI) The content of a thought of a subject S is completely determined by factors internal to S

and the corresponding definition of externalism would run

(SE) The content of a thought of a subject S is at least partly determined by factors external to S

*The work on this paper was supported by a research grant of The Bank of Sweden Tercentenary Foundation, for the project *Meaning, Communication, Explanation*. I have benefited much from discussing the ideas presented here with esteemed colleagues, especially in Prague, at conferences on interpretation, organized by Petr Kotatko, 2001–2005.

‘Thought’ is here of course used in a psychological sense: to have a thought is to be in a certain mental state, whether the thought is occurrent or dispositional. What makes the definition imprecise is, primarily, that the wording leaves it open *what to count* as internal to the subject.

On one interpretation, it means factors internal to the subject’s *body*. On that definition, objects outside the boundaries of the subject’s body, or events taking place outside them, count as external. On this view, *distal stimuli*, i.e. objects or events in the environment of sentient subject are external, while *proximal stimuli*, i.e. stimulations of sensory receptors, are internal.¹

On another interpretation, internal factors are factors internal to the subject’s *mind*. Everything physical counts as external on that interpretation. Internal factors would be such things as sensations, sense impressions, intentions and various attitudes. It must be made sure that to the extent that e.g. an attitude A_1 is a content determining factor for another attitude A_2 , it is not in virtue of an externally determined content of A_1 .

In both cases, bodily internalism and mental internalism, further distinction can be made that would induce new subdivisions. I shall not go deeper into this issue. I shall here choose mentally internal factors as defining internalism [externalism]. However, what I have to say further on can where needed be reformulated in terms of bodily internal factors. The main ideas are intended to be neutral with respect to the two alternatives.

I shall call the generic view *subjective internalism/externalism*, because in both cases the factors considered are internal or external to a particular subject. By contrasting *intersubjective* internalism/externalism with subjective internalism I intend a primary distinguishing feature:

Intersubjective internalism/externalism concerns representations determined by *relations* between subjects

The representations I shall be concerned with are *linguistic* representations, i.e. linguistic expression types. And the basic idea about contents of linguistic expressions, i.e. linguistic *meaning*, is that it is determined by the *communicative relations* between speakers. The further question whether internalism or externalism holds in this case, depends on the nature of the communication relation: does it or does it not involve more than what is involved in determining thought content.

I then define *regular intersubjective internalism* as follows:

- (RII) The meaning of an expression e used in communication between speakers S and S' is completely determined by factors that determine thought contents of S and S' .

¹It is strictly better to speak of the excitatory *states* of sensory cells than of the stimulations that are the causes of these states. Depending on how things are categorized, the causes may be external to the receptors, and hence external simpliciter. However, since the talk of proximal stimuli is well established, I’ll keep the terminology.

I shall here also sidestep the issue of stimulations of anatomically internal receptors.

Correspondingly, *regular intersubjective externalism* is the view that:

- (RIE) The meaning of an expression e used in communication between speakers S and S' is at least partly determined by factors external to S and S' other than those that determine thought contents of S and S' .

Some comments are in order. First, if two speakers S and S' belong to a larger linguistic community, it may be that the meanings of expressions that they use for communicating with each other indeed are completely determined by factors that determine thought contents of *other* speakers of the community. In this case, by the formulations given, (intersubjective) externalism would hold for S and S' but maybe internalism would hold for other members. This is a complication that I simply abstract from, in order to keep the presentation simple. It is convenient to focus on two speakers rather than on larger speech communities.

Second, the formulation ‘the meaning of an expression e ’ has a natural *generic* reading, by which it is at least meant that the property in question applies to expressions in general. However, it is conceivable that for *some* expressions internalism holds while externalism holds for others. Strictly speaking, then, intersubjective internalism as a *general thesis* should be the view that the meaning of *all* expressions are completely determined by what determines thought content. Similarly, intersubjective externalism should then strictly be the view that there is at least *some* expression whose meaning is determined by other factors. However, since I shall put forward the view that the meanings of at least *most* expressions are externally determined (either in the sense of (RIE) or in the sense of (WIE) below), I hold a view that is much stronger than the strict existential claim. Since there is then no reason in the present context to make it more precise, I shall keep the formulations above.

Third, there are two related views which I shall call *strong intersubjective internalism* and *weak intersubjective externalism*, respectively. The strong internalist view combines regular intersubjective internalism with subjective internalism:

- (SII) The meaning of an expression e used in communication between speakers S and S' is completely determined by factors that are internal to S or S' .

This is to be read to the effect that every determining factor is internal either to S or to S' . The weak externalist view combines regular intersubjective externalism with subjective internalism:

- (WIE) The meaning of an expression e used in communication between speakers S and S' is at least partly determined by factors external to S and S' .

Strong intersubjective internalism follows from the conjunction of subjective internalism and regular intersubjective internalism. Conversely, it is in principle possible to accept strong intersubjective internalism while rejecting *subjective*

internalism, but that is an intuitively strange combination. I shall not pay more attention to that option.

The distinction between regular and weak intersubjective externalism is of greater relevance. The regular version clearly implies the weak version, but not conversely. I am inclined to be a subjective internalist. On the assumption of subjective internalism, regular intersubjective externalism is true iff weak intersubjective externalism is true, for then any external factor that determines meaning is external to what determines thought contents. However, I am not going to argue for subjective internalism, and therefore I shall not rule out the possibility that the external factors I take to determine linguistic meaning *also* determine thought content. If that is true, then *regular intersubjective externalism* is false, while *weak intersubjective externalism* is true. Accordingly, I shall assert the weak version, but be inclined to accept the regular version as well.

Fourth, since the meaning of linguistic expression is intended to be determined by the communicative *relations* between S and S' , weak intersubjective externalism is true iff the relevant communicative relation *involves* objects, events, properties or other entities that are external to S and S' . That is, if we let μ be the function from expression to their meanings we can schematically depict the strong internalist view as

$$(1) \quad \mu(e) = m \text{ iff } R(e, m, S, S')$$

for some relevant 4-place relation R . By contrast, the weak externalist view would be represented as

$$(2) \quad \mu(e) = m \text{ iff } R'(e, m, S, S', w)$$

for some 5-place relation R' . The difference is that in the externalist case, there is an extra *world* argument: whether m is the meaning of e depends on properties of “the world”, over and above the features of the world represented by the speaker arguments. Of course, since the inner structures of the relations R and R' are left wholly unspecified, the (1) and (2) schemata are more for illustration than for specifying logical form.

Fifth, and last, it should be evident by now that the intersubjective externalism suggested here is not a variety of *anti-individualism*, or *social externalism*, as this notion has been explained in Burge 1979 and subsequent literature. Social externalism is, in the present terminology, a kind of *subjective* externalism, only that the external factors involve other speakers. The same goes for Hilary Putnam’s idea of the *division of linguistic labor*, as presented e.g. in Putnam 1975a, except that in Putnam’s case, what is determined by way of relations to *other* speakers, the experts, is the meaning of linguistic expressions *as* used some particular non-expert speaker. The primary meaning is not determined by a *relation* between speakers. It is just that some speaker inherit the meaning of the expression from others to which they do stand in certain relations. It is again a kind of subjective (linguistic) externalism.

In the following two sections, I shall set out my views on communicative

success, and my views on the explanatory role of semantics, respectively, and then in the final section return to the consequences for externalism.

2 Communicative success

I shall here present, without argument, my views on communicative success.² First, what I take as successful or unsuccessful are *communicative events*. A communicative event involves a sender (speaker), a signal (utterance), and a receiver (hearer). In case there are several hearers there are also several communicative events (and the signals will not be identical, even if they have common initial segments). I shall not here go into the further conditions that need to be met for a signal to be *communicative* over and above being a causal process.

The signal is partly caused or motivated by a particular *inner state* of the sender, the *initial state*. In the case of human communication this is most often, and perhaps invariably, a *mental state*. The signal also causes a particular inner state of the receiver, the *terminal state*, again typically or always a mental state in the case of human communication. I shall here just assume that the identity of the initial and terminal states is determinate.

Whether the communicative event is successful or not depends on the relation between the initial and terminal states. It does not depend e.g. on preceding or subsequent actions or behavior of speaker and hearer, but only on those mental states that are directly involved in the communication.

Moreover, in the case of typical human communication, the initial and terminal states are states with *representational content*, and in such cases success depends on a relation between the respective contents. That is, the communicative act is successful just if the content of the hearer's terminal state is suitably related to the content of the speaker's initial state. From now on I shall refer to those states as *thoughts*. Communicative success therefore turns on a relation between thought contents.

Before turning the question what the appropriate relation between thought contents is, we need to have an idea of how thought contents are *individuated*. For instance, an identity relation between contents under one scheme of individuation will not be an identity relation under another scheme. I shall here opt for the simplest alternative: individuate thought contents as possible worlds propositions, i.e. as sets of possible worlds.

There are two well-known difficulties that ultimately will force a modification of this choice. One has to do with distinguishing contents that are true in the same possible worlds. As has been pointed out many times, in a regular possible worlds framework there is e.g. just one necessary proposition, even though intuitively there are infinitely many non-synonymous sentences with different meaning. One way of distinguishing such meanings is to introduce *impossible* worlds next to the possible ones, and another is to introduce structured meanings (e.g. see Lewis 1972, Cresswell 1986). I think there are several independent reasons for having structured meanings next to ordinary propositions (see Pagin

²I have argued extensively for these views in Pagin 2006b and Pagin 2006a.

2003a, 2005, Pagin and Pelletier 2006). Since more elementary issues are under consideration at present, however, I shall set this difficulty aside.

The other difficulty concerns the need to account for essentially *indexical* thoughts: the content of Alfred's thought that *he* (himself) is poor, is not identical with any content in which Alfred is identified in a non-indexical way. This was above all stressed by Hector-Neri Castañeda (1966, 1968). David Lewis (1979) suggested that this phenomenon be handled by an alteration to the possible worlds framework, with *centered worlds* rather than ordinary worlds. A centered world is a pair of a world and a space-time point in that world. Again, I think this modification is ultimately needed, but for present purposes this complication can be ignored.

Given that we have settled for an individuation scheme,³ the most natural idea for the relevant relation is that it is the *identity* relation: for communication to succeed it is both necessary and sufficient that the content of the hearer's thought is the *same* as that of the speaker's. There are, however, a number of reasons for thinking that with most thought contents that are of an empirical nature, identity is too strong a requirement.⁴ Instead, I propose a relation of *similarity*.

It is natural to object that *similarity* is not a transitive relation, and that choosing a *similarity of content* standard of success allows that *A* might successfully communicate with *B* and *B* with *C* while *A*'s communication with *C* *unsuccessful*, even though each speaker's content is the same across the different events.⁵ However, the idea of similarity I shall propose is in fact an equivalence relation.

If you have an identity standard in a possible worlds framework, you have communicative success if, and only if, the speaker proposition (the content of the speaker's thought) and the hearer proposition (the content of the hearer's thought) have the same *truth value* with respect to every possible world. By switching to a similarity standard one *could* instead say, loosely, that we have success just if the two proposition agree in truth value with respect to *most* worlds. That would not, however, yield a transitive standard of success, and it is not what I shall say. Whether agreement holds in a majority or a minority of worlds is also of lesser importance. What is important is that they agree in *relevant* worlds.

The idea is that when two speakers communicate successfully, their thoughts contents are so similar that they *reliably* agree in truth value. For two proposition to *reliably* agree in truth value, they must first of all actually agree, but should also agree with respect to all the *relevant* ways in which the world might have been different. That is, it is required that for a given set W_N of *relevant alternative worlds*, the two propositions agree in truth value with respect to all members of W_N . I shall refer to the relevant alternative world as the *nor-*

³Note that a scheme of individuation leaves it open e.g. whether to employ a Russellian or a Fregean interpretation of proper names, since both kinds of interpretation can be represented in the possible worlds framework.

⁴For these reasons the reader is referred to Pagin 2006a.

⁵Jerry Fodor and Ernie Lepore (1999) have especially emphasized this.

mal worlds, because what is relevant depends on similarity to the actual world, in the sense of Lewis 1973; normal worlds are closer to the actual world than abnormal ones.

En example might help at this point. Suppose that by ‘tree’ Alfred understands *plant with woody tissue and annual growth renewal*, while Elsa by ‘tree’ understands *plant with bark surface, branching top, nourishment through root system, and with either leaves or needles on the branches*. As a matter of facts everything that Alfred counts as a tree is counted also by Elsa as a tree, and *vice versa*. Moreover (or so I shall assume), in all “normal” alternative worlds, Alfred’s and Elsa’s respective concepts of a tree *still* coincide: only in worlds with a basic difference in biological basis would the extensions of two concepts come apart. Thus we can hold that when Alfred says

(3) I have five trees in my backyard

(and there is sufficient agreement with respect to the other components of the sentence), Alfred’s proposition (what Alfred understand by (3)) will agree in truth value with Elsa’s proposition in all members of W_N , i.e. with respect to all normal alternative worlds.⁶

It should be kept in mind that the truth value agreement at issue is an agreement between the propositions that are the contents of the relevant speaker and hearer thoughts. It is irrelevant what *attitude* speaker and hearer takes towards that proposition. Thus, in case Alfred is sincere in his utterance of (3) and Elsa believes what he says is false (because she thinks there are only three trees in Alfred’s backyard), then there is not in this case truth value agreement with respect to what they *believe*. Nonetheless, there *is* agreement between the propositions they entertain in connection with the utterance: the proposition that is the content of the thought that generates the utterance, and the proposition that is the content of the thought that is generated *by* the utterance.

Finally, what counts as a “normal” alternative to a possible world is meant to be an *objective* matter. It is not relative to context, to speaker or hearer interests. It depends on basic regularities of physics, chemistry, biology, perhaps metaphysics, and maybe more. Exactly what should go in there (basic features of human history, current state of technology?) is an open question, and one cannot expect a sharp boundary.

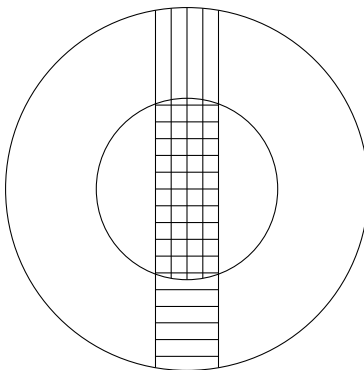
Nevertheless, the idea that what is a normal alternative is an objective fact, determined by the modal profile of the world, independent of the thoughts of

⁶There is a complication that I shall not elaborate upon here. If Alfred says

(i) It is necessary that trees have woody tissue

this would intuitively be *false* by Elsa’s *tree* concept. So there cannot be communicative success. Under a necessity operator, for success to obtain, it is required that the two propositions are similar enough, not just with respect to the *actual* world, but with respect to all *accessible* worlds. That is, with respect to every accessible world, there is a set of “normal” alternatives to *that* world, and the two propositions will have to agree in all worlds in every such set of alternatives.

speaker and hearer, has the consequence that it often depends on *the world* whether or not two propositions are similar enough. Hence, it often depends on the world whether or not a communicative event is successful. Alfred's proposition and Elsa's proposition might diverge, even radically, in the domain of worlds *outside* W_N , i.e. outside the range of normal worlds. As long as they agree within W_N , communicative success is unimpaired. In this sense the shared external world itself can be seen as The Great Equalizer, the factor that filters out semantic differences, establishing communicative success despite personal semantic idiosyncrasies of speaker and hearer.



Vertical proposition and horizontal proposition coinciding in the normal domain.

This world dependence is the first major driving force of intersubjective externalism of linguistic representations.

3 Semantics and the explanation of success

The *second* major driving force of intersubjective externalism is the idea that the role of natural language semantics is to provide a core part of a theory of communication that has the potential of *explaining* the high rate of communicative success in linguistic interaction. By common sense standards, i.e. judging by the frequency of achieved practical coordination, we usually communicate successfully, when we communicate linguistically. In more abstract matters, it is often more difficult to judge. This is compatible with present ideas, since in more abstract matters the connection with worldly, stabilizing facts is looser and more indirect.

What is the *explanation* of this high rate of success? Part of the present idea of explanation is that *semantics* plays a central role. I regard this not as any kind of *a priori* truth, nor as self-evident in any other sense or respect, but rather as a well-motivated empirical hypothesis: *the meaning hypothesis*:

- (MH) Semantic properties of natural language expressions play a central role in the explanation of the rate of success of communication performed by means of utterances of these expressions.

(MH) is not self-evident, since it is conceivable that the rate of success could be explained without reference to semantic properties, but with reference only to physical, mental and mental-representational properties.⁷

The main reason why semantics is needed is the productivity of human communication. As Frege observed (Frege 1923, opening passage), human speakers often manage to communicate new propositions (propositions that have not previously been thought by speaker or hearer, and often not by anyone) by means of new expressions (new word combinations):

It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the very first time can be put into a form of words which will be understood by somebody to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thought corresponding to the parts of a sentence, so that the structure of the sentence serves as an image of the structure of the thought. (Frege 1923, opening paragraph)

Frege apparently thought that only a certain *isomorphism* (or structural sameness) between sentence and proposition could make this possible. I think Frege's claim about the necessary condition was mistaken, but that he nevertheless was right in pointing to a structural similarity between sentence and interpretation for the correct *explanation* of success.

Although Frege does not use the standard formulation of *the principle of compositionality*, it is reasonable to interpret him as intending this principle to be satisfied. A standard formulation of the principle runs

(PoC) The meaning of a complex expression is a function of the meanings of its parts and its mode of composition.

The natural simple intuition is that the hearer is able to arrive at the intended meaning of the new complex expression by putting things together: the hearer knows the meanings of the simple parts (individual words or other primitive expressions), and also knows the semantic *significance* of the syntactic modes of composition employed, and putting these factors together the hearer is able to arrive at the interpretation of the complex expression.

In fact, it is even plausible that Frege intended something stronger than standard compositionality. Compositionality does not require an isomorphism. Also, Frege emphasizes not only the task of the hearer to arrive at an interpretation, but the task of the *speaker* to arrive at a suitable *expression* for communicating the proposition. For this we would need a principle that is inversely related to standard compositionality:

(IC) The expression of a complex content is a function of the expressions of its parts and its mode of composition.

⁷One attempt to achieve an explanation without relying on semantics was proposed in Schiffer 1987. I argue at length against Schiffer's view in Pagin 2003b.

In a corresponding fashion, the principle of *Inverse Compositionality* can explain how the speaker manages to find a suitable linguistic expression for the proposition or other complex content that the speaker has in mind. The corresponding idea is that the speaker associates primitive expressions with some basic contents (concepts) and certain syntactic constructions with conceptual operations, and thus is able to put the complex expression together out of the simpler parts.

The explicit appeal to *knowledge* is intuitively natural, but ultimately dispensable and in some respects even harmful. The relation between semantics and the mental should be pictured differently. I have proposed (Pagin 2003a,b), in part following Schiffer (1987), that public language semantics be related to contents of mental representations by way mental operations. The speaker / hearer has a set of mental representations of public language *expressions* (or syntax). As suggested by Schiffer, a set of mental operations effect mappings between the mental syntax representations and the mental content representations. Then, given such a picture, we can say that a working explanation of novel communicative success is *possible* if syntax and semantics are so related, according to our syntactic theory **T** and our semantic theory **M**, that we can add a *theory* of mental representation **G**, such that

- i) **G** systematically assigns a representation of each element in the domains of **T** and **M** respectively, and
- ii) **G** systematically assigns a mental operation to each part of the definition in **M** of functions between the domains of **T** and **M**.

G is a theory of the mental operations that map representations of syntactic elements on representations of semantic elements (and *vice versa*). If our theories of syntax and semantics are such that a theory **G** can be added which meets these conditions, then we do have an account of how actual communicative success *can* be achieved. And for this we need only consider syntax and semantics themselves.

In order that there be success in communicating representational content there must be something of a representational nature that is *shared* between speaker and hearer. The mental representations themselves are not shared. On the simplest picture, what is shared is the representation content itself: the speaker proposition is the same as the hearer proposition. Then we can regard the natural language semantics as providing an interpretation of a sentence *s* that (given contextual additions) is identical with both the speaker and the hearer proposition associated with the use of *s*. That is, if *F* is a function mapping mental representations of content r_c on mental representations of natural language syntax r_s , and *H* is a function mapping mental representation of syntax on public language expressions,⁸ we would get the following equation:

⁸It should really be grammatical *terms* or *trees*, for the meaning function can only operate in full generality on *disambiguated* expressions, and ordinary surface forms are often structurally ambiguous.

$$(Eq) \quad C(r_c) = \mu(H(F(r_c)))$$

That is (disregarding pragmatic modifications), the content of a mental content representation is exactly the meaning of the public language expression that corresponds to the mental syntax representation that in turn corresponds to the mental content representation.

However, the idea of public language semantic as explanatory does not require that this equation be true. It is enough that the *extension* of the mental content and the *extension* of the public language expression are the same in the actual world *and* in all *normal* alternative worlds.

In a possible worlds framework, where intensions are functions from worlds to extensions, (Eq) corresponds to a claim about sameness of extension in all worlds:

$$(Eq_w) \quad \forall \mathbf{w} ((C(r_c)(\mathbf{w}) = \mu(H(F(r_c))))(\mathbf{w}))$$

If we restrict the requirement to sameness of extension in normal worlds, we instead have

$$(Eq_n) \quad \forall \mathbf{w} \in W_N ((C(r_c)(\mathbf{w}) = \mu(H(F(r_c))))(\mathbf{w}))$$

What is shared between speaker and hearer, therefore, is the *restriction* of mental content to the domain of normal worlds. This they both share with the meaning of the linguistic expression itself (in context). It is decisive for the externalism at issue that the domain of normal worlds is a feature of the world, not a feature of mental contents.

4 Externalism of the semantics

Let's go back to the example of the term 'tree'. As far as I understand, 'tree' is not a natural kind term by current taxonomic standards of biology. But even if it isn't, there are nomic biological regularities that keep features of trees together: bark layer with woody tissue, size of stem with size of root system, nourishment through root system with photosynthesis in needles or leaves etc. Such features cluster as a matter of natural regularities. That is, these features cluster in the biologically normal worlds.⁹

It is part of the regularities of the world that features cluster. Suppose that there is a certain kind of entity, K associated with a set of clustering features M_K . In virtue of the clustering there will be a number of proper subsets $S_i \subset M_K$ such that, in all normal worlds, the set of object having all the S_i features is the same as the set of objects having all the M_K features. That is, if treating the feature sets as intensions, we have

$$(N) \quad \forall \mathbf{w} \in W_N (S_i(\mathbf{w}) = M_K(\mathbf{w}))$$

⁹Of course, some but not all features are shared with other plants.

Then, if we have two proper subsets S_i and S_j both satisfying (N), Alfred may associate a term, say ‘tree’, with S_i and Elsa may associate it with S_j . This difference will not be detectable in ordinary discourse about actual entities having the features (say, trees). If Alfred and Elsa disagree about actual states of affairs regarding these entities, at least one of them will be mistaken, and it will not turn on the difference between their respective conceptions. Only when they move into a more advanced discourse, e.g. modal discourse, may such a difference be detected.

The technical terms ‘concept’ and ‘property’ are used in a variety of ways. I shall here propose the following: each of M_K and the various S_i subsets is, or corresponds to, a *concept* of the same underlying *property* (in the example: K). What that property is, is well-defined in the normal domain, but need not be well-defined outside it. The term in question refers to this property, or kind, regardless of any particular way of selecting it. That is, none of the feature sets is treated as privileged: no “essences” are imposed.

For instance, to connect this with Putnam’s Twin Earth thought experiment (Putnam 1975a), different speakers may associate different features with the term ‘water’. Some may focus on macroscopic and some on microscopic properties, some with everyday and some with scientific properties. None of these will be treated as corresponding to an underlying essence of water. The term ‘water’ denotes the stuff that actually has all the features, and it remains an open question how to extend the term to possible substances that have some of these features and lack others.¹⁰ Speaker converse about water without the need of an underlying essence of *sameness of liquid* that identifies water in any possible world.

What the term in question, such as ‘tree’ or ‘water’, denotes does depend on the modal profile of the world, for it does depend on what possible variations are *normal*.¹¹ Since the semantics of the expressions are determined by what explains communicative success, it is determined by a relation between speakers that does involve the shared external world. Clearly, this is a kind of intersubjective externalism. In the terminology of section 1, we have at least a *weak* intersubjective externalism, since there is a world dependence in the meaning determination.

It is less clear that we get a *regular* intersubjective externalism. That will depend also on principle that determine thought content. I would like to conclude, however, with the claim that the kind of weak intersubjective externalism proposed is compatible with subjective *internalism*. For we can regard the S_i feature sets as being completely *subjective*. For instance, Alfred’s conception of a tree may be that of *the property, if any, that corresponds to Alfred’s tree-impressions*, or perhaps *the property, if any, that corresponds to features α, β, \dots of Alfred’s tree impressions* or something yet further, perhaps involving some theory as well. If there are such conceptions, they are semantically independent

¹⁰Of course, the thought experiment concerns a different region of the actual world, but as long it *is* a thought experiment, it is best treated as an exotic, probably abnormal possibility.

¹¹The examples that are most intuitive concerns natural kinds or kinds that are “almost” natural kinds, but the scope of the idea itself is not limited to natural kinds.

of features of the environment. And, if subjective internalism is true, we have regular intersubjective externalism if we have weak intersubjective externalism.

On such a picture, the respective inner worlds of Alfred and Elsa may be radically different. Still, because of the regularities of the causal interaction with the environment, and because of the physical regularities of the environment itself, there will be clusters of representations in Alfred's mental life that correspond to clusters of representations in Elsa's mental life. Alfred's and Elsa's conceptions might correspond to different external properties in possible worlds *outside* the normal domain, but inside the normal domain what they correspond to coincide. The terms they both use denote precisely those coincidental properties. On this picture, that Alfred and Elsa are able to communicate successfully at all depends on the stabilizing contribution of The Great Equalizer, the external world itself.

Department of Philosophy
Stockholm University

References

- Burge, T., 1979, 'Individualism and the mental', in P. A. French, T. E. Uehling and H. K. Wettstein (eds.), *Midwest Studies in Philosophy*, 4, 73–121, University of Minnesota Press, Minneapolis.
- Castañeda, H.-N., 1966, 'He': a study in the logic of self-consciousness', *Ratio* 8:130–57.
- Castañeda, H.-N., 1968, 'On the logic of attributions of self-knowledge to others', *Journal of Philosophy* 65:439–56.
- Cresswell, J., Max, 1986, *Structured Meanings. The Semantics of Propositional Attitudes*, MIT Press, Cambridge, Mass.
- Fodor, J. and Lepore, E., 1999, 'All at sea in semantic space: Churchland on meaning similarity', *Journal of Philosophy* 96:381–403.
- Frege, G., 1923, 'Gedankengefüge', *Beiträge zur Philosophie des Deutschen Idealismus*. 3:36–51. Reprinted in Frege, *Logische Untersuchungen*, Vandenhoeck & Ruprecht, Göttingen, 1976. Translated by R. Stoothoff as 'Compound thoughts', published in *Mind* 72 (1963), 1–17.
- Lewis, D., 1972, 'General semantics', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Reidel, Dordrecht.
- Lewis, D., 1973, *Counterfactuals*, Basil Blackwell, Oxford.
- Lewis, D., 1979, 'Attitudes *de dicto* and *de se*', *The Philosophical Review* 88:513–43. Reprinted in Lewis 1983, 133–56. Page references to the reprint.

- Lewis, D., 1983, *Philosophical Papers. Volume I*, Oxford University Press, Oxford.
- Pagin, P., 2003a, ‘Communication and strong compositionality’, *Journal of Philosophical Logic* 32:287–322.
- Pagin, P., 2003b, ‘Schiffer on communication’, *Facta Philosophica* 5:25–48.
- Pagin, P., 2005, ‘Semantic and pragmatic composition’, paper presented at the semantics-pragmatics workshop in Context 05.
- Pagin, P., 2006a, ‘When does communication succeed?’, in P. Kotatko (ed.), *The Constraints on Interpretation*, Cambridge Scholar Press, Cambridge.
- Pagin, P., 2006b, ‘What is communicative success?’, under review.
- Pagin, P. and Pelletier, J. F., 2006, ‘Content, context and communication’, in G. Preyer and G. Peter (eds.), *Content and Context. Essays on Semantics and Pragmatics*, Oxford University Press.
- Putnam, H., 1975a, ‘The meaning of ‘meaning’’, in K. Gunderson (ed.), *Language, Mind and Knowledge*, Minnesota Studies in the Philosophy of Science VII, University of Minnesota Press, Minneapolis. Reprinted in Putnam 1975b. Page references to the reprint.
- Putnam, H., 1975b, *Mind, Language and Reality. Philosophical Papers vol. 2*, Cambridge University Press, Cambridge.
- Schiffer, S., 1987, *Remnants of Meaning*, MIT Press, Cambridge, Mass.