Propositions, propositional attitudes and belief revision*

Rohit Parikh

Computer Science, Philosophy and Mathematics

CUNY Graduate Center and

Brooklyn College of CUNY

33 West 42nd Street, New York, NY 10036

email: ripbc@cunyvm.cuny.edu

February 8, 1999

Abstract

In this paper I will propose a new approach to certain semantic puzzles due to Frege, Kripke and others, and the question of propositional attitudes, via the notion of belief revision. The principal logical tool used for setting up the solution is the notion an individual theory. I shall suppose that each individual has his or her own theory consisting of the sentences the individual accepts. It is this theory, which together with the individual's preferences leads to a choice of some actions over others; it is this theory out of which the individual communicates; and it is this theory which is revised when a communication, in the form of a sentence uttered by another, is heard or otherwise received. The theory may also be revised as a result of an experience with the "world" as might happen when someone without an umbrella gets wet and adds "It is raining" to his theory.

There is also a community theory T_c which consists of sentences which we all accept and which, in the varous puzzles, has a different notion of possibility from that of the individual's theory. Thus there is conflict between these two theories not only about what is true or what is known, but even about what is possible. It turns out that most of the puzzles considered in the literature are puzzles about reconciling these two theories.

A crucial aspect of this way of looking at the puzzles is that the central role of truth conditions as the determinants of meaning is rejected and more flexible, purely syntactic tools are used. It will be argued that the truth theoretic view of meaning is not an adquate tool for semantics, and abandoning it makes the problems much more amenable. This need not imply abandoning truth, but merely dethroning it from its central role as a foundation for semantics.

1

There is a tradition going back to Frege and continuing through Tarski and Davidson, which bases the notion of meaning on that of truth, and assumes that understanding a sentence is the same as understanding its truth conditions. There are slight differences between the Fregean and Millian traditions respectively in the way they deal with proper names. But both accept a notion of a linguistic act as being meaningful, just standing by itself, and the role of society and the world in determining *meaning* is ignored. The early Wittgenstein[1] expresses this position succinctly when he says in the *Tractatus* that whether one proposition has sense cannot be allowed to depend on whether another one is true.

A different view of meaning can be associated with the later Wittgenstein. There does not seem to be complete agreement as to the views expressed in the *Philosophical Investigations*, but it is clear that the notion of truth as the basis for meaning is discarded and the proposal that "meaning is use" is put forward, albeit with some reservations. Since use tends to be in a social context, this makes comunication central, and the role of society important if not pre-eminent. Apart from Wittgenstein, the importance of communication is also emphasized by Grice. The role of society in determining meaning is evident in Putnam's experts who play an important role in deciding what we mean by words like "water" or "gold".

The work of Ramsey on probability which predates the later phase of Wittgenstein's work can also be seen this way. In Ramsey's explication of probability, the emphasis is removed from probability as an objective, pre-existing reality. Instead we look at the choices of an agent whether to accept or reject a bet, and show that if an agent is not to be vulnerable to Dutch book made against her, her bets must follow a probability function obeying the Kolmogorov axioms. If they do not, then it is possible to place bets with her in such a way that no matter what happens, she must lose. Here there is no fact of the matter as to what the probabilities are, just whether her view is consistent. But in such accounts of subjective probability the real world does enter as an influence. Two Bayesian agents starting with different prior probabilities will tend to converge to similar beliefs as time passes. However, a subjective probability statement is not *directly* about the world.

My approach in this paper will follow the second stream and steer away from Metaphysics, taking instead a more social and utilitarian view of language and communication. I will show how various puzzles in the literature are then no longer puzzling. The view will be syntactic, i.e. sentence based, but we will not have to pay the Quinean price [Q2] of total opacity in belief contexts. E.g. in understanding "Jack believes that pigs have noses", we will be able to take some advantage of the internal structure of "Pigs have noses". Also, though the approach will be sentence based, sentences will not be the objects of belief. Thus we will be able to make sense of why an English speaking person who asserts "It is raining" and a German speaking person who asserts "Es regnet" can be said to have the same belief even though there is no sentence they would both assent to.

Some difficulties with truth: the notion that understanding a sentence is the same as knowing its truth conditions has one serious defect. This supposedly necessary requirement is almost never met in practice. For instance, before the Clinton trial got under way, I believed that I knew what impeachment was. I believed, correctly, that it happened to a

---

[1]Cf. *Tractatus Logico-Philosophicus* 2.0211

President, though rarely, that it was a bad thing to happen to him and that Congress had a major role. I also believed, falsely as it turned out, that successful impeachment resulted in the removal of the President from office. I did not know that impeachment was really an indictment by the House and that it was only the prelude to a trial by the Senate. Did I not then know what impeachment was? Surely I had a partial understanding of it.

We all know that a large mouse is smaller than a small elephant. So "small" is contextual. Suppose now that I tell you that the heffalump Kevin weighs 175 pounds and is *small*. I also tell you that heffalump Jane weighs 140 pounds, and then ask you if Jane is small. You will readily say "yes". But you cannot possibly know the truth conditions for "Jane is small" in this context since you do not know what heffalumps are. Perhaps all heffalumps who weigh less than 2,000 pounds are small, or perhaps all heffalumps who weigh more than 200 pounds are already not small. But it does not matter for the question which is the case, or even if "small" for heffalumps, as for us, is vague. Given the information you have about Kevin and Jane, your answer is clear.

For a different example, suppose a travel agent in Calcutta sells airline tickets to JFK and Newark, and imagines that both of them are cities on the East Coast of the US. He is right about Newark and wrong about JFK, but his ignorance will not be a problem for him unless he comes to the US and starts looking for the Mayor of JFK. In a situation like this, a person may utter some appropriate sentences, and yet not know all the criteria for the truth of all statements containing some term. In all these examples, there is a *partial understanding* of some word which is adquate for many situations, perhaps all that a particular agent might encounter. The notion that knowing the meaning is knowing the truth conditions does not have room for such partial understandings.

Consider now a different - non truth-theoretic - view of understanding. Suppose we think of social dialogue as an enormous game played by society with mutual benefit and occasional harm, and involving the occasional utterance of sentences by different individuals. This is simply the block world of Wittgenstein but on a larger scale. In such a situation, someone may be able to make a few right moves, or at least moves that are not too wrong, without fully knowing all the rules. My understanding of impeachment is enough for me. It would not be enough for Mr. Clinton's lawyers. In Wittgenstein's block world, a builder issues commands like 'slab!', 'block!' etc, and his assistant brings him the corresponding stone. In such a world, with such a primitive language, it is not necessary that the builder's helper know how to utter the word "slab", nor is it necessary for the builder to recognize the word when uttered by someone else. It is enough that the builder knows to say "slab" when he needs one and that the helper brings a slab when he hears the word "slab".

The notion that understanding consists in knowing truth conditions is implausible for yet another reason. As we know, the sorites paradoxes present a problem for the existence of any semantics for vague predicates. Several, not wholly satisfactory semantics have been proposed (see e.g. [KS]), but it is not plausible that ordinary people are acquainted with them. Yet we understand perfectly well a sentence like "Mary believes that Bill is bald". We can make inferences from this like "Mary believes something about Bill" and "If Bill has a head full of hair, then Mary is wrong about Bill", but probably not something like

"Mary believes that Bill has no more than 1,743 hairs". We do not know Mary's criterion for baldness and do not know that she applies it consistently.

Finally, reducing meaning to truth runs up against another roadblock. Suppose that it is raining outside and I say "It is raining outside". Then what I have said is true provided that I am speaking English. So I should really say that I am speaking English and that it is raining outside. But this is clearly no solution, for if I am not speaking English, then "I am speaking English" may not have its usual meaning. An obvious, and not contrived example is where I am actually speaking Pidgin and think that I am speaking English.

Such an infinite regress is avoided in evolutionary approaches to language. In [Sk] a person A who receives an input from the set $\{a, b, c\}$ can send a signal 1 or 2 or 3 to another person B who then chooses among three actions a', b', c'. If they together choose one of the pairs $(a, a'), (b, b'), (c, c')$ then they each get, say a dollar. Otherwise they get nothing. If they play this little game repeatedly, then under appropriate assumptions, their behaviour will converge to some "language", where perhaps 1 will mean b, 2 will mean c and 3 will mean a. I.e. when A recieves input a then she sends the signal 3 to B who will then choose action a'. Such a convergence presupposes no prior understanding between A and B or even their ever meeting face to face, and yet a simple language can evolve.

I shall give my account of the semantic puzzles by concentrating on what *sentences* someone is willing to assert or assent to. These will be called that person's *s-beliefs*. T_i will be the set of i's s-beliefs. What *propositions* may be will be an issue we will come to later. For simplicity I shall assume that our sentences are not indexical, though allowing indexical sentences only introduces complications and no serious new problems.

Propositions and Possible Worlds: Consider Kripke's Pierre before he ever arrived in England.[2] At that time he has in his vocabulary a constant *Londres* and a predicate *jolie*. For him (ignoring other constants and predicates) there are two possible worlds, w_1, in which *jolie(Londres)* is true; and w_2, in which it is false. This allows him to choose among three personal theories, T_1, generated by *jolie(Londres)*; T_2 generated by $\neg$*jolie(Londres)*; and T_3 which consists only of tautologies. These three theories correspond to the S-propositions[3] $\{w_1\}$, $\{w_2\}$ and $\{w_1, w_2\}$ respectively. If he starts out in T_3, knowing nothing about *Londres* then his state of mind corresponds to $\{w_1, w_2\}$. When he learns, or comes to believe, *jolie(Londres)*, then he eliminates the world w_2 and moves to the state $\{w_1\}$. [4]

When he arrives in England, he adds a new constant *London* to his vocabulary. To simplify matters, let us not use the new predicate *pretty* and continue to use the word *jolie*. Also, let us use the word *scenario* to mean a complete theory in Pierre's new language, consistent

---

[2]In Kripke's example, Pierre grows up in France believing "Londres est jolie." Later he arrives in a not very pleasant part of London, learns English by exposure, and would agree to "London is not pretty". Kripke's puzzle is about the apparent inconsistency in Pierre's beliefs, even though there is no logical deficiency which we can blame Pierre for.

[3]Following Stalnaker, I shall take an S-proposition to be a set of possible worlds. I have not defined *proposition* yet.

[4]In many knowledge theoretic accounts one learns by eliminating possible worlds which were epistemically possible. There is a great deal of literature on this, but see, for instance [P5].

with his beliefs, but which might not be a possible world in our sense. A possible world will be a scenario from the point of view of *our* theory or the *community* theory T_c. So now he has four scenarios, w_3, w_4, w_5, w_6 such that *Londres* is *jolie* in the first two and *London* is *jolie* in the middle two, i.e. in w_4, w_5. Since he claims that Londres is jolie but London is not, he believes that the real world is the scenario w_3 in which *Londres* is *jolie* but *London* is not.

However, for us, knowing (or believing) that *Londres = London*, only the w_4 in which both are *jolie* and w_6 in which neither is, are possible worlds. Of Pierre's four scenarios, only two are possible worlds (from our viewpoint) and the one he believes he is in is not one of them.

We are defining a *possible world* as a scenario from the point of view of our theory T_c which is the community theory. Let $T_c^\square = \{A | \square A \in T_c\}$. Then $T_c^\square$ is the set of all formulae which are considered in T_c to be necessary. A scenario is a possible world iff it contains all of $T_c^\square$.

Why do we call *our* scenarios possible worlds? The reason is that we are in a Moore's paradox situation. It makes no sense for us to say that something is true but that we do not believe it, nor does it make sense to say that we (fully) believe something but that it is not true. Thus for example, believing that Hesperus is Phosphorus, we also believe that it is true that they are the same and therefore that it is necessary that they are the same. And yet it could turn out that the ancients were right after all and there actually are two objects there in the sky.

So our beliefs have a special status. When we discuss Pierre we discuss him from inside our theory and we have no way to escape this condition. Thus it makes sense to refer to *our* scenarios as possible worlds. However, the notion of possible world *is* somewhat contextual. If in some paper, a philosopher introduces two names a and b and tells us that they name the same person, then if we follow Kripke, we will take the equality $a = b$ to be necessary and this decision will affect our notion of possible world. But of course a, b may not be real names at all. The position I am taking is that there is no absolute notion of possible world, but only a contextual one and that is good enough to discuss all the issues involving necessity.

An individual theory T_i is 1-inconsistent if there are no scenarios among its complete extensions. This will be the case if *all* formulae are among the theorems of T_i. T_i is 2-inconsistent if there are no possible worlds among its extensions, i.e. iff $T_i \cup T_c^\square$ is inconsistent. It can happen that theory is 2-inconsistent but 1-consistent, and then a puzzle like Pierre's will arise.

Indeed we can see that Pierre's theory is 1-consistent and 2-inconsistent. Distinguishing the two notions resolves the tension between the intuition that he is inconsistent, and that there is no logical error which he is to blame for.

We can also define two different notions of belief. These are similar to Quine's two notions, but perhaps more articulated. We can say that Pierre believes$_1$ A if A is in his theory T_p. Pierre believes$_2$ A if A can be proved in T_p *together with* formulae in $T_c^\square$. Because of the axiom $a = b \rightarrow \square(a = b)$, which we shall include, identities $a = b$ between names, if they are in T_c, are necessary. These can then be used to justify beliefs$_2$.

It is belief$_1$ which explains behaviour. For example, when Pierre is in London, he believes$_1$ that he in London but does not believe$_1$ that he is in Londres. Thus if he believes$_1$ that a landmark, say the Buckingham Palace is in Londres, he would make no attempt to see it merely because he believes$_1$ he is in London. Belief$_1$ also controls which sentences Pierre would utter and assent to. He will not, without further information, assent to an A which he believes$_2$ but does not believe$_1$.

If it is belief$_1$ which explains behaviour, why should we consider belief$_2$ at all? Its relevance comes from the fact that *we* communicate with Pierre and need to know what to make of his utterances.

Suppose we met Pierre in France, and ourselves knew both English and French but were not aware that Pierre had this confusion about London-Londres. If he told us in French that Londres had a population of 7 million and we relied on him, we *would* add the statement "London has a population of 7 million" to our own store of beliefs. We might even report to someone else, "Pierre told me that London has a population of 7 million." But Pierre does not believe$_1$ that London has a population of 7 million. He only believes$_2$ it. However, from *our* point of view, the two names denote the same place so we are free to make the substitution. We might make that substitution even when reporting *his* beliefs as long as we do not expect any harm to come of that. Suppose that *Londres = London* were compatible with T_p and we think that he would accept our word for it that they are the same place. Then we could easily say "Pierre believes that London has a population of 7 million" to someone who has 'London' but not 'Londres' in her vocabulary.

However, since his actual theory contains both *jolie(Londres)* and $\neg jolie(London)$, *London = Londres* is incompatible with T_p and belief revision will be needed to deal with his problem. One of these two sentences will have to be deleted before T_p can accomodate *London = Londres*.

Since Pierre has to delete some of his current beliefs in order to accomodate the formula *Londres = London*, belief revision enters the picture. Belief revision along the direction started by [AGM] deals with the issue of how an agent revises her beliefs when she receives information incompatible with her current beliefs.

It is now time for a formal development.

The Formalism:

L is a first order language with equality, augmented by the modalities $\square$ and $\lozenge$. The operators $\square$ and $\lozenge$ are assumed to be S5. However we will generally only have occasion to consider modal formulas whose proper subformulae are non-modal. The formulae $a = b \rightarrow \square(a = b)$ for constants a, b will be among the axioms. Thus the theories will contain a complete set of axioms and rules for first order logic with equality, and also the axiom schemata

$$\square(A \rightarrow B) \rightarrow (\square A \rightarrow \square B)$$

$$\square A \rightarrow A$$

$$\Box A \rightarrow \Box\Box A$$

$$\Diamond A \rightarrow \Box\Diamond A$$

$$a = b \rightarrow \Box(a = b)$$

Where $\Diamond A$ is an abbreviation for $\neg\Box\neg A$. In addition to modus ponens and generalization, we will also have the rule which allows us to infer $\Box A$ from A provided that only logical axioms have been used to prove A.

A *theory* is a set of formulas which contains all axioms and is closed under generalization, modus ponens and $\Box$ generalization for those formulae A which have been proved purely logically. A *complete* theory will be one which, for every closed formula A, contains A or its negation.

We assume a community of speakers i each with his own theory T_i. We say that i accepts some sentence A, or A is an s-belief of i, just in case $A \in T_i$. It is clear and relatively unproblematic to say that i accepts the sentence A just in case i would assent to A. Since the set of i's s-beliefs is a theory, it is closed under logical inference. This is the theory out of which i acts and communicates.

Thus if I am thirsty and there is a sentence in my theory that there is cold water in the fridge, then I will walk towards the fridge and open the door. It is also the reason I utter the words "There is cold water in the fridge" if *you* are thirsty. Finally, when you are thirsty and I say "There is cold water in the fridge", then you will revise your theory so as to contain this sentence and then being thirsty, you will walk to the fridge, open the door and look inside.

I am neglecting problems of logical omniscience here by assuming that T_i is logically closed. However, such problems are much more easily tackled on this syntactic approach. We could, in a more sophisticated treatment, allow T_i not to be logically closed, but to acquire new formulas whenever an individual notices a new consequence of his old beliefs.

The community theory T_c contains sentences which are accepted for the purpose of the dialogue. For example, I suspect that the the entire story of Pierre is fictitious and there is no such person just as there is no twin earth. But when we read Kripke's paper about Pierre, or one of Putnam's papers, we are *inside* the theory that they have invented for us and for purposes of reading the paper we act as if Pierre actually exists and that there is actually a twin earth. In addition to fictions invented for the purpose of some paper, T_c also contains sentences which most members of the community accept. For instance, "The sun rises in the East" as well as sentences which experts accept, like "Gold has atomic number 79".

We also assume a revision operator $*$ on theories which satisfies at least the seven axioms given below. Intuitively, if T is a theory and A is a sentence, then $T * A$ indicates the new state of s-belief if the initial state of s-belief was T and one is informed of a formula A. $*$ may be non-deterministic in that two individuals, both with the same current s-beliefs and hearing the same sentence A may have different s-beliefs afterwards.

Notation: $A \Leftrightarrow B$ means that A and B are logically equivalent, i.e. that $A \leftrightarrow B$ is provable from the logical axioms. Similarly, $A \Rightarrow B$ means that $A \rightarrow B$ is logically true. If X is a set of formulae then $Con(X)$ is the logical closure of X. In particular, X is a theory iff $X = Con(X)$. $T * A$ is the revision of T by A. $T \overset{\bullet}{+} A$ is $Con(T \cup \{A\})$, i.e. the result of a brute addition of A to T without considering the need for consistency. Finally, T_0 is the smallest theory, consisting of the logically true formulas only.

Axioms 1–6 are due to [AGM], but have been slightly modified.

1. $T * A$ is a theory.
2. $A \in T * A$ if A is consistent.
3. If $A \Leftrightarrow B$, then $T * A = T * B$.
4. $T * A \subseteq T \overset{\bullet}{+} A$
5. If A is consistent with T, i.e. it is not the case that $T \vdash \neg A$, then $T * A = T \overset{\bullet}{+} A$.
6. $T * A$ is consistent.

We add axiom 7:
7. If $s = t$ is in both T and $T * A(s)$, then $T * A(s) = T * A(t)$.

In 1–7 I shall confine myself to theories T and new s-beliefs A which are individually consistent.

Definition: A *scenario* for individual i is a complete theory T such that if $\Box A \in T_i$ then $A \in T$. A scenario T is *epistemically possible* for i if $T_i \subseteq T$. A scenario relative to the community theory T_c will be called a *possible world*. We define a *proposition* as an equivalence class of sentences under the equivalence relation $A \approx B$ iff $\Box(A \leftrightarrow B) \in T_c$. $[A]$ is the equivalence class of A.

For every proposition $[A]$, consider the set of possible worlds T such that $A \in T$. This set does not depend on which A we choose from $[A]$, i.e. if $A \approx B$ then for all possible worlds T, $A \in T$ iff $B \in T$. This correspondence is one-one, for if $\neg(A \approx B)$, then there is a possible world which contains one of A, B and not the other. We shall say the individual i believes the proposition $[A]$ iff he believes$_2$ the formula A. It then becomes possible for an individual to believe inconsistent propositions even though his s-beliefs are consistent. But this is no longer a paradox, for the notion of proposition (as I have defined it) is parasitical on *our* beliefs T_c which we cannot expect everyone to share.

For individual i we can define i-propositions analogously but as we saw, they may not be actual propositions because of differences between T_i and T_c. Thus $jolie(Londres)$ and $jolie(London)$ belong to the same proposition but to different p-propositions and that is why Pierre can entertain the conjunction $jolie(Londres) \wedge \neg jolie(London)$, which is a consistent formula for T_p but is inconsistent with $T_c^\Box$. While there is a p-proposition corresponding to it, there is no *proposition*.

An individual's theory T_i may be 'troubled' in various ways. It might simply be 1-inconsistent. If not, it still might be (a) 2-inconsistent. Or (b) be consistent with $T_c^\Box$ but not contain it or (c) it might contain $T_c^\Box$ but conflict with T_c. In case (a) it will have scenarios, but no possible worlds. In case (b) it will have both possible worlds and scenarios which are not

possible worlds. In case (c), all its scenarios will be possible worlds but not (according to us) the actual world.

Pierre's theory is of kind (a). But The ancient astronomer who does not *know* that Hesperus and Phosphorus are the same will be in category (b). If he has a single predicate V for "visible", then he would entertain four scenarios in which, respectively,

$$(V(h), V(p)), \ (\neg V(h), V(p)), \ (V(h), \neg V(p)), \ (\neg V(h), \neg V(p))$$

hold. he thinks that only the middle two are ever realized but does not exclude the first and the last, which are the two possible worlds.

Now we can give sense to how two people speaking different languages can share a belief. Suppose a German speaker A assents to "Es regnet" and an English speaker B assents to "It is raining". Now $\Box(Es\ regnet \ \leftrightarrow \ It\ is\ raining)$ is in our theory T_c provided that we have adopted the usual translation from German to English. Hence, *given* the choice of translation we have made, $[It\ is\ raining]$ is the same proposition as $[Es\ regnet]$. Thus they believe the same proposition.

To see that our role and the choice of T_c matters, consider the Pali word *metta* which occurs often in Buddhist literature. It is variously translated in English as friendship, love or compassion. Now suppose that Sariputta says that he feels metta towards Bill, and Jane says that she feels compassion for him. Then whether we regard Sariputta and Jane as sharing an attitude will depend on how we choose to translate the Pali word 'metta'. As Quine [Q] has pointed out, there may not be a fact of the matter here if the cultures are different enough.

Partial Opacity: If Pierre believes$_2$ that $jolie(Londres)$, then he *will* believe$_2$ $jolie(London)$ whereas if he believes$_1$ $jolie(Londres)$, he need not believe$_1$ $jolie(London)$. However, belief$_1$ is not totally opaque. We just need to know a bit more about Pierre's other beliefs.[5] If he believes$_1$ $A(s)$ and also believes$_1$ $s = t$, he will believe$_1$ $A(t)$ provided that A has no modal operators. If A does have modal operators, then it would be necessary that $\Box(s = t)$ be in his theory. Moreover, if he believes$_1$ $A(s)$, $\ s = t \in T_c$ and he does not dis-believe$_1$ $A(t)$, then generally, by telling him that $s = t$ we can get him to believe$_1$ $A(t)$. This is why the difference between the two notions of belief is not usually a problem.

De Re, De Dicto: Quine's Ralph believes that the man in the gray hat (mgh) is a spy [Q2] but does not believe that the man on the beach (mob) is a spy. However, they are one and the same man, Bernard J. Ortcutt. Since $mgh = Ortcutt$ is in T_c, Ralph does believe$_2$ that Bernard J. Ortcutt is a spy, and we can say that Ralph believes *of* Bernard J. Ortcutt that he is a spy. Here we are not relying on the fact *about the world* that the man on the beach is Bernard J. Ortcutt but on the fact that this equality is *in our theory* T_c, and that only because we are going along with Quine's story. The world does enter because it plays a role in how T_i or T_c are modified. But the formulae in all these theories are not *about* the world. Rather whether they are in T_c is affected by the world.

---

[5]For belief$_2$ we only needed to know about *our* beliefs, which is much easier.

Travis and brill-bream: In this example [Tr], Jones has heard that certain fish called brill are delicious. She has also seen bream in a display somewhere and taken them to be brill. Today she has decided to have fish for dinner and decides on brill, or so she thinks. She now goes to a grocer and asks him for a brill. Travis wants now to know if Jones wants brill, or bream, or fish that are brill but look like bream, or none of these. But let us see what might happen if Jones does go shopping for a brill. Let us suppose that bream are yellow and brill are blue. She believes, falsely, that brill are yellow.

Here are two possible dialogues:

Dialogue 1:

Grocer: Here you are, madam.
Jones: No, I don't like the look of that, do you have any tofu?

Dialogue 2:

Grocer: Here you are, madam.
Jones: But surely that isn't brill. Aren't brill yellow?
Grocer: No, ma'am, it is bream which are yellow. But they taste something awful and I do not stock them. I know many people do not like the blue color of brill, but they are delicious.
Jones: All right, I will trust you. Please tell me what I owe you.

Neither of these dialogues is strange. Do they answer Travis' questions? The fact that brill are actually blue is inconsistent with Jones' beliefs. On learning it, she must revise her theory and that will involve a non-deterministic choice, which of her current beliefs to drop. We might say that in the first dialogue, Jones wanted fish that were yellow and had mistakenly thought that brill were yellow. In the second case she did really want brill even though she, as it turned out, did not know what they looked like. In dialogue one she drops her s-belief that she wants brill. In the second she drops her s-belief that she wants yellow fish. But what matters is what Jones does and not how Jones' behaviour fits into some semantic theory which works well only with people who know both what brill look and taste like. After Jones finds out from the grocer what brill really look like, she too will be one of these people.

Narrow content, wide content, and belief revision: The references below are to [St], page 134. I want to show how the notion of s-belief revision can help us understand the distinction.

In Stalnaker's example Bert lives in world a where the word "water" is used to refer to H_2O and water is indeed the best liquid for quenching thirst. In world b, the word "water" refers to XYZ which is the best liquid for quenching thirst in that world. In world c, "water" refers again to H_2O, but it is gatorade which is best for quenching thirst. Finally, in world d, "water" refers to XYZ but twin-gatorade is the best liquid for quenching thirst in d.

Stalnaker says that according to wide content, Bert's statement "Water is the best drink for quenching thirst" is true in a and false in b, c, d. By narrow content, as Stalnaker seems to

suggest, it is true in a and b and false in c and d.

How does the s-belief revision model handle this distinction?

Earlier, we referred to two ways a theory may be revised. It may be revised by contact with the world. E.g, when I stick my hand out the window, it gets wet, and I add "It is raining" to my theory. The other way is when an English speaker says "It is raining", and I add that sentence to my theory. My hand can be quite dry because, say, I am indoors.

Let us assume that XYZ is not poisonous to Bert and that he visits the world b. He sees a lake of XYZ and says "Wow! Look at that beautiful lake!". He drinks some XYZ and says "I always said that water is the best thing for quenching thirst". That is narrow content.

Now suppose in the same scenario, that Bert is accompanied by a chemist from world a, i.e. an expert from Bert's community. When Bert says, "I always said that water is the best thing for quenching thirst", the chemist will say, "That isn't water, that is XYZ!" "You mean water isn't the best thing for quenching thirst around here?". "No, there *is* no water here, but XYZ is fine, and when you want some XYZ, just say that you want some water. The waiter will bring you XYZ", says the chemist. So with this input from the chemist, Bert does not add "Water is the best ... in world b" to his theory. This is wide content. And it arises here because the chemist is bringing in elements of (Bert's) T_c. Twin-Bert would have a different T_c'.

So narrow content can be seen as coming from a relationship between Bert and the world. In wide content, his community participates in defining *his* terms.

Arthritis in the thigh: Burge's Bert asserts that arthritis is a disease which can strike the thigh as well as the knee. He is not aware that arthritis is a disease of the joints only. Does Bert have a false belief about arthritis or a true belief about some other disease? Clearly, Bert will have to drop some old s-beliefs when he learns facts which are in T_c, for the latter contains the statement "arthritis can afflict only the joints". (I assume that Bert does know that the thigh is not a joint.) If Bert actually feels pain in the thigh, he might then drop his s-belief that he has arthritis in the thigh but still insist that he has *some* disease in the thigh, only not arthritis. If, however, he had purely theoretical s-beliefs about arthritis, he might say instead, "I guess I was wrong to think that arthritis can afflict the thigh". In the latter case (but not in the former) we could say that Bert's original belief was *about* arthritis.

This is the general situation when some theory T_i is inconsistent with $T_c^\square$. In such a case the theory is true in no possible world and would need to be revised in order to accomodate at least $T_c^\square$. If we have good reason to think that i will accept some A after revision, then we could say already "i believes A". So we say in dialogue 2 that Jones really wants brill, but in dialogue 1 that she wants yellow fish. We may even know enough about Jones in advance to know how the dialogue *would* go and then we can say to her, "What you really want is ..."

Dealing with Vagueness: One of the characteristics of vague predicates is that many of them are observational in a rather crude sense. I.e. we expect to tell just by looking at a

man whether he is bald or tall. The counting of hairs or measuring his height is excluded.[6] This is an area where truth based theories go completely off the mark because they try to precisify something which is intended to be imprecise.

However, note that what we observe depends on neuro-chemical processes which go on in our body and brain. For example, when we look at a red object for a while and then look at a white wall, we will see a green spot. This is because seeing the red object has caused some chemical changes in our retina. Someone else watching the same object as I may have a different experience if she has a cataract or is color blind or has just been watching TV. This means that we cannot, as a community, respond exactly the same way to the same stimuli, nor can the same person respond the same way on two different occasions.

A truth theoretic approach does not know what to do with this fact. But it is shown in [P2,P3] how in spite of the unreliability of perceptions, we can competently deal with *some* situations involving vague predicates. For instance we can competently carry out the directive to stop when a traffic light is red and to proceed when it is green. We can do this reliably, even though we do not have a community wide notion of what green is.

Usually, given an object of color somewhere between purple and blue, different people will classify it differently, some as purple, and some as blue. So there is no community consensus on what counts as blue [P2,P3]. Thus different people tend to have different individual theories in the same language. If one person classifies a particular object o as blue and another as purple, then one will have the formula $B(o) \land \neg P(o)$ and other will have $P(o) \land \neg B(o)$ in their theories. This is unavoidable. But these theories are close enough that we do not get into serious trouble. With the traffic lights, which are only red, yellow and green. we use only those parts of the theory on which agreement is nearly perfect. It almost never happens that different people classify the same light as green and red respectively. Thus we do not *solve* the problem of which objects are which color, but get around the issue.

Conclusion: In this paper I have outlined an approach to various linguistic puzzles in terms of the problem of reconciling two different theories: an individual's own theory, and some sort of reference theory. The latter might be the theory of a community, or a theory of a particular dialogue. In this latter case it might make assumptions which we know are false, but which are taken to be true for the sake of the example. Once we see that the problem is one of reconciling two views, the paradoxes disappear. They arose only because we gave one of the theories, T_c an exalted status and called it *the World*.

References:

[AGM] Carlos Alchourron, Peter Gärdenfors and David Makinson, "On the logic of theory change: partial meet contraction and revision functions", *J. Symbolic Logic* 50 (1985) 510–530

[Bi] Akeel Bilgrami, *Belief and Meaning*, Oxford, Basil Blackwell, 1994.

[D] Donald Davidson, *Inquiries into Truth and Interpretation*, Oxford University Press, 1984.

---

[6]Just imagine the dialogue: Policeman: "Was the perpetrator tall?" You: "I don't know. I did not have a measuring tape with me." You would be lucky to escape with just a frown.

[Fr] Gottlob Frege, "On sense and reference" *Zeit. f. Philsophie u. Philsophische Kritik* vol. 100 (1892) pp. 25-50, republished in various sources.

[Gr] Paul Grice, "Logic and conversation", in *Studies in the Way of Words*, Harvard University Press, 1989, pp. 22-40. Appeared originally in *Syntax and Semantics*, vol 3, Academic Press, 1975.

[G] Adam Grove, "Two modellings for theory change", *J. Phil. Logic* 17 (1988) 157–170.

[Kr] Saul Kripke, "A Puzzle about belief" in A. Margalit (ed) *Meaning and Use* Reidel 1979, pp. 239-283.

[KS] Rosanna Keefe and Peter Smith, *Vagueness: a Reader*, MIT press, 1996.

[P] Rohit Parikh, "Beliefs, Belief Revision, and Splitting Languages", in *Proc. 2nd conference on Information Theoretic Approaches to Logic, Language, and Computation*, to appear 1998, CSLI publications.

[P2] Rohit Parikh, "Vagueness and Utility: the Semantics of Common Nouns" in *Linguistics and Philosophy* **17** 1994, 521-35.

[P3] Rohit Parikh, "Vague predicates and language games", *Theoria* Spain, vol XI, no. 27, Sep 1996, pp. 97-107.

[P4] Rohit Parikh, "Language as Social Software, a Contribution from Computer Science to Philosophy", to appear in volume dedicated to Burton Dreben, eds Floyd and Shieh, Harvard 1999.

[P5] Rohit Parikh, "Finite and Infinite Dialogues", in the *Proceedings of a Workshop on Logic from Computer Science*, Ed. Moschovakis, MSRI publications, Springer 1991 pp. 481-498.

[Pu] Hilary Putnam, "Meaning and reference", *J. of Philosophy* **70**, (1973) 699-711.

[Q] W. V. Quine, *Word and Object*, New York and London, 1960.

[Q2] W. V. Quine, "Qantifiers and propositional attitudes", in *Ways of Paradox*, Ramdom House, 1966.

[R] F. P. Ramsey, *The Foundations of Mathematics and other Logical Essays*, ed. Richard Braithwaite, London 1931.

[St] Robert Stalnaker, "Narrow content" in *Propositional Attitudes*, Ed, Anderson and Owens.

[St2] Robert Stalnaker, *Inquiry*, MIT press, 1984.

[Sc] Stephen Schiffer, *Remnants of Meaning*, MIT press, 1987.

[Sk] Brian Skyrms, "Evolution of Inference" *Dynamics in Human and Primate Societies: Agent Based Modeling of Social and Spatial Processes*, ed. Timothy Kohler and George Gumerman, Oxford U. Press, to appear, 1999.

[Tar] Alfred Tarski, "The Concept of truth in formalized languages", in *Logic,Semantics, Metamathematics*, Clarendon press, Oxford 1956.

[Tas] William Taschek, "Frege's puzzle, sense and information content", *Mind*, 1992, pp. 767-791.

[Tr] Charles Travis, "Order our of Messes", review of [Bi], *Mind*, vol 104, January 1995, pp. 133-144.

[W1] Ludwig Wittgenstein *Tractatus Logico-Philosophicus*, London, 1922.

[W2] Ludwig Wittgenstein *Philosophical Investigations*, (Trans. G.E.M. Anscombe) Basil Blackwell, 1953.