Sentences, Propositions and Logical Omniscience, or What does Deduction tell us?

Rohit Parikh
City University of New York
365 Fifth Avenue
New York, NY 10016
rparikh@gc.cuny.edu

February 28, 2007 3:15 PM

Abstract: In deductive reasoning, if ϕ is deduced from some set Γ, then ϕ is already implicit in Γ. But then how do we learn anything from deduction? That we do not learn anything is the (unsatisfying) answer suggested by Socrates in Plato's *Meno*. This problem is echoed in the problem of logical omniscience prominent in epistemic logic according to which an agent knows all the consequences of his/her knowledge. An absurd consequence is that someone who knows the axioms of Peano Arithmetic knows all its theorems.

Since knowledge presumes belief, the lack of closure of (actual) beliefs under deduction remains an important issue.

The post-Gettier literature has concentrated on the gap between justified true belief and knowledge, but has not concerned itself with what *belief* is. This question, or at least an analysis of sentences of the form *Jack believes that frogs have ears*, **has** been prominent in the philosophy of language. But even there, less attention has been paid to *how* we know what someone believes. This turns out to matter.

The question *How we know what someone believes* has, however, been addressed by Ramsey and de Finetti as well as by Savage in the context of decision theory and the foundations of subjective probability. Beliefs are revealed by the choices we make, the bets we accept and the bets we refuse. And among these choices are the choices of what to say and what to assent to. Of course the second kind of choice can only be made by humans, or at least by creatures possessing language. But the first kind of choice is perfectly open to animals and to pre-lingual children.[1]

[1]In [17, 18] Ruth Marcus describes a man and his dog in a desert, deprived of water and thirsty. When they both react the same way to a mirage, it is hard to deny that they both have the same false belief that they are seeing water. The fact that animals can experience mirages would seem to be substantiated by the fact that the Sanskrit word for a mirage is *mrigajal* which literally means 'deer water,' faux water which deer pursue to their deaths. Frans de Waal makes a much more detailed case for animal intentionality in [42].

We argue that this way of thinking about beliefs not only allows us to address the issue of logical omniscience and to offer formal models of "inconsistent" beliefs, it also allows us to say something useful about Frege's problem, whether it is sentences or something else that we believe, and whether *believes* can be taken to be a binary relation at all.

1 Introduction

We start with a dialogue between Socrates, Meno, and Meno's slave boy from Plato's *Meno* [35]. In this dialogue, Socrates carries on a conversation with the boy and asks him the area (space) of a square whose side is two. The boy correctly answers that the area is four. And then Socrates raises the question of *doubling* the area to eight. He wants to know what the side should be. The boy makes two conjectures, first that the side of the larger square should also be double (i.e., four) – but that yields an area of sixteen, twice what is wanted. The boy's second guess is that the side should be three – but that yields an area of nine, still too large.

Socrates: Do you see, Meno, what advances he has made in his power of recollection? He did not know at first, and he does not know now, what is the side of a figure of eight feet: but then he thought that he knew, and answered confidently as if he knew, and had no difficulty; now he has a difficulty, and neither knows nor fancies that he knows.

Meno: True.

Socrates: Is he not better off in knowing his ignorance?

Meno: I think that he is.

Socrates: If we have made him doubt, and given him the "torpedo's shock," have we done him any harm?

Meno: I think not.

Socrates: We have certainly, as would seem, assisted him in some degree to the discovery of the truth; and now he will wish to remedy his ignorance, but then he would have been ready to tell all the world again and again that the double space should have a double side.

Now Socrates suggests that the diagonal of the smaller square would work as the side of the larger square and the boy agrees to this. We continue with the quotation:

Socrates: And that is the line which the learned call the diagonal. And if this is the proper name, then you, Meno's slave, are prepared to affirm that the double space is the square of the diagonal?

Boy: Certainly, Socrates.

Socrates: What do you say of him, Meno? Were not all these answers given out of his own head?

Meno: Yes, they were all his own.

Socrates: And yet, as we were just now saying, he did not know?

Meno: True.

Socrates: But still he had in him those notions of his-had he not?

Meno: Yes.

Socrates: Then he who does not know may still have true notions of that which he does not know?

Meno: He has.

—

Here Socrates appears to be arguing that the beliefs which Meno's slave could be brought to have via a discussion were beliefs which he *already* had.[2] There is of course a tension in this line of argument, for if this is so, then deduction would appear to be dispensable, and yet leading the boy to make deductions was Socrates' own method in bringing him to a new state of mind.

A conclusion similar to that of Socrates also follows from the Kripke Semantics which has become a popular tool for formalizing logics of knowledge. The semantics for the logic of knowledge uses Kripke structures with an accessibility relation R, typically assumed to be reflexive, symmetric, and transitive. If we are talking about belief rather than knowledge, then R would be serial, transitive, and euclidean. Then some formula ϕ is said to be believed (known) at state s iff ϕ is true at all states R-accessible from s. Formally,

$$s \models B(\phi) \quad \text{iff} \quad (\forall t)(sRt \rightarrow t \models \phi)$$

[2]This impression is surely strengthened by the remarks which Socrates makes elsewhere in *Meno* to the effect that if knowledge was always present then the soul must be eternal. *The soul, then, as being immortal, and having been born again many times, and having seen all things that exist, whether in this world or in the world below, has knowledge of them all; and it is no wonder that she should be able to call to remembrance all that she ever knew about virtue, and about everything.*

If follows immediately that if a formula is logically valid then it is true at all states and hence it is both known and believed. Moreover, if ϕ and $\phi \to \psi$ are believed at s then both are true at all t such that sRt, and hence ψ is true at all such t. Thus ψ is also believed at s. Moreover, a logically inconsistent formula can be neither known nor believed. Aumann's semantics [3] uses partitions rather than Kripke structures, but is known to be equivalent [8] and suffers from the same difficulty.

There have been suggestions, [8, 21] that these 'problems' can be dealt with by using tools like *awareness*, *impossible possible worlds*, or by referring to *computational complexity.* In our view, these methods do not address the fundamental problem, namely *what is it that we are trying to model?* In order to have a theory of knowledge, we need to have some criterion for "measuring" what someone knows. Just to start with a logic which yields logical omniscience, and then fiddling with it by various methods, leaves aside the question of what our goal is. Unless we have a clear sight of the goal we are not likely to reach it.

For instance we do not know if Goldbach conjecture is true, but if true it is necessarily true. Surely we cannot argue that we do not know which only because we are not *aware* of it. That is surely not the problem. Nor can computational complexity help us, because *if* it is true, then there is a sound formal system which has it as an axiom, and then a proof of the Goldbach conjecture is going to be very fast in that system. If it is false, then of course the falsity is provable in arithmetic as it only requires us to take notice of a particular even number which is not the sum of two (smaller) primes.

The issue of computational complexity can only make sense for an infinite family of questions, whose answers may be undecidable or at least not in polytime. But for *individual* questions whose answers we do not know, the appeal to computational complexity misses the issue.

Moreover, beliefs are not always consistent, so one could say, *thank heaven for a lack of logical omniscience!* A person, some of whose beliefs conflict with others, and who believes all logical consequences of her belief, would end up believing *everything!*

The following two examples are from Daniel Kahneman's Nobel lecture, 2002.

> *A bat and a ball cost \$1.10 in total. The bat costs \$1 more than the ball. How much does the ball cost? Almost everyone reports an initial tendency to answer 10 cents because the sum \$1.10 separates naturally into \$1 and 10 cents, and 10 cents is about the right magnitude. Frederick found that many intelligent people yield to this immediate impulse: 50% (47/93) of Princeton students, and 56% (164/293) of students at the University of Michigan gave the wrong answer. Clearly, these respondents offered a response without checking it.*

———–

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations.

#6 Linda is a bank teller

#8 Linda is a bank teller and active in the feminist movement

89% of respondents rated item #8 higher in probability than item #6.

But the set of bank tellers who are active in the feminist movement is a proper subset (perhaps even a rather small subset) of the set of all bank tellers, so #8 *cannot* have higher probability than #6.

Clearly human states of belief are not consistent. Nor are they usually closed under logical inference. Various researchers including ourselves have looked into this issue [23, 26, 8, 41, 11]. There is also more recent work by Artemov and Nogina [2], and by Fitting [10], aimed towards the logic of proofs.

And yet we do operate in the world without getting into trouble, and when we discover an inconsistency or incompleteness in our thinking, we remove it, either by adding some beliefs which we did not have before, or be deleting some which we had. What we need is a formal account of this activity and a more general representation of beliefs than is afforded by the current Kripke semantics. In particular, logically omniscient belief states need to be represented as a proper subset of *all* belief states.

The proposal we make here draws on the work of Ramsey [37], de Finetti [9], Hayek [12], and Savage [38] with echoes from [17, 18, 20, 43]. According to this view, an agent's beliefs are revealed by the *choices* which an agent makes, and while 'incoherent' choices are unwise, they are not inconsistent, for if they were, they would be *impossible*. For an agent to make such unwise, incoherent, choices is perfectly possible and is done all the time.

Imagine that Carol assigns probabilities of .3, .3, and .8 respectively to events $X, Y, X \cup Y$. One could say that these probabilities are *inconsistent*. But in fact nothing prevents Carol from *accepting* bets based on these probabilities. What makes them incoherent is that we can make Dutch book against Carol – i.e., place bets in such a way that no matter what happens, she will end up losing money.[3] Thus incoherent beliefs, on this account, are *possible*, (and hence consistent) but unwise. It will be important to keep in mind this distinction between inconsistency and incoherence.

[3] For instance we can bet \$3 on X, \$3 on Y, and \$2 against $X \cup Y$. If either X or Y happens, we earn \$7 (at least), and lose (at most) \$5. If neither happens, we gain \$8 and lose \$6, so that we again make a profit – and Carol makes a loss.

We now look at some previous work which contains strong indications of the direction in which to proceed.

Hayek [12] considers an isolated person acting over a period according to a preconceived plan. The plan

> *may, of course, be based on wrong assumptions concerning external facts and on this account may have to be changed. But there will always be a conceivable set of external events which would make it possible to execute the plan as originally conceived.*

Beliefs are related here to an agent's plans, and it is implicit in Hayek's view that the agent believes that the world is such that his plans will work out (or have a good chance of doing so).

But note that the belief states which are implicit in plans are more general than the belief states which correspond to Kripke structures, and the first may be incoherent and/or logically incomplete. An agent may plan to buy high and sell low, and expect to make money. It is not possible for the agent to actually do so, but it is perfectly possible for the agent to have such a plan.

Note that having a plan does not require that the plan be formulated explicitly in language, or even that the planner *has* a language. It is perfectly possible for an animal to have a plan and to a smaller extent, it is also possible for a pre-lingual child to engage in deliberate behaviour which is plan-like.

This is an important point made explicitly by Marcus [18] and also fairly clear in [20] that we cannot limit states like belief and desire to language using creatures, i.e., to adult humans and older children. Frans de Waal is even more emphatic on this point.[4] We must also attribute some parts of intentional states to higher animals and to pre-lingual children. Going back further, Hume [13] is quite explicit on this point:

> *Next to the ridicule of denying an evident truth is that of taking much pains to defend it; and no truth appears to me more evident, than that beasts are endow'd with thought and reason as well as men.*

[4] "It wasn't until an ape saved a member of our own species that there was public awakening to the possibility of nonhuman kindness. This happened on August 16, 1996 when an eight-year old female gorilla named Binti Jua helped a three-year-old boy who had fallen eighteen feet into the primate exhibit at Chicago's Brookfield Zoo. Reacting immediately, Binti scooped up the boy and carried him to safety." De Waal is quite disdainful of Katherine Hepburn's remark in *The African Queen*: "Nature, Mr. Allnut, is what we are put in this world to rise above."

Ramsey [36] has a related comment.[5]

> *It is for instance possible to say that a chicken believes a caterpillar of a certain sort to be poisonous,and mean by that merely that it abstains from eating such caterpillars on account of unpleasant experiences connected with them.*

And so does Millikan [20].

> *Reasoning is just trial and error in thought. Dennett (1996) calls animals capable of trial and error in thought "Popperian." The reference is to Popper's remark that* it is better to let one's hypotheses die in one's stead. *The Popperian animal is capable of thinking hypothetically, of considering possibilities without yet fully believing or intending them. The Popperian animal discovers means by which to fulfill its purposes by trial and error with inner representations. It tries things out in its head, which is, of course, quicker and safer than trying them out in the world. It is quicker and safer than either operant conditioning or natural selection. One of many reasonable interpretations of what it is to be rational is that being rational is being a Popperian animal. The question whether any non-human animals are rational would then be the question whether any of them are Popperian.*

Finally Whyte [43] suggests that we can even define truth in this way. He appeals to Ramsey's principle (R):

> (R) *A belief's truth condition is that which guarantees the fulfilment of any desire by the action which that belief and desire would combine to cause.*

However, we need not address the issue of truth here. Brandom [5] subjects Whyte to some criticism, but it only has to do with whether the criterion of truth is adequate. We are only interested here in a representation of belief. Such a representation must accomodate all the following groups: language possessing, logically omniscient humans;[6] language possessing

[5] Of course we need not and should not attribute to the chicken the specific belief that such caterpillars are *poisonous.* But we *can* attribute to it the belief that eating them will lead to bad consequences.

[6] Of course I do not personally know any logically omniscient humans, but in a limited context it is possible. Suppose that p stands for *Pandas live in Washington DC*, q stands for *Quine was born in Ohio*, and r stands for *Rabbits are called* gavagai *at Harvard.* Suppose that Jill believes that p is true and that q and r have the same truth values. Then she is allowing two truth valuations, $v = (t, t, t)$, and $v' = (t, f, f)$. Given a formula ϕ on p, q, r in disjunctive normal form, she can evaluate $v(\phi)$ and $v'(\phi)$. If both are t she can say that she believes ϕ. If both are f, she disbelieves ϕ, and otherwise she is suspending judgment. Then Jill *will* be logically ominiscient in this domain. But note that she will actually have to *make the calculations* rather than just sit back and say, "Now *do* I believe ϕ?"

but fallible humans; and non-lingual creatures like animals and very young children.[7]

However, if beliefs are not (always) expressed by sentences, or coincide with propositions expressed by sentences, then we need another way of representing belief states, and then relate language-oriented belief states to some specific species of such belief states.

We shall consider two kinds of beliefs. Non-linguistic beliefs which may also be possessed by animals, and linguistic beliefs which can only be possessed by humans; adults and older children. Of course the last two groups will also have non-linguistic beliefs which must be somehow correlated with their linguistic beliefs.

Let $\mathcal{B}$ be the space (so far unspecified) of belief states of some agent. Then the elements of $\mathcal{B}$ will be identified with the choices which the agent makes. Roughly speaking, if I believe that it is raining, I will take my umbrella, and if I believe that it is not raining, then I won't. But clearly the choice of whether to take my umbrella or not is correlated with my belief only if I don't want to get wet. So my preferences enter *in addition* to my beliefs.

Conventional theories like that of Savage [38] take the (actual and potential) choices of an agent, assume that these choices satisfy certain axioms, and simultaneously derive both the agent's beliefs (her subjective probability) and the agent's preferences (expressed by a utility function). But it is known that agents which are rational in Savage's sense are not very common and so we would like to retain the *rough framework* without the inappropriate precision of Savage's theory.

So we will just assume that the agent has *some* space $\mathcal{P}$ of preferences, and that the choices are governed by the beliefs as well as the preferences. We will use $\mathcal{S}$ for the set of choice situations, and C for the set of choices. Thus the set $\{U, \neg U\}$ could be a choice situation (with U standing for *take the umbrella*) and both U and $\neg U$ are elements of C. We could say very roughly that an agent who prefers not to get wet will choose U from $\{U, \neg U\}$ iff she believes that it is raining.

An agent who does have language can also be subjected to a purely linguistic choice. If asked *Do you think it is raining?* the agent may choose from the set {*Yes,No,Not sure*}. And it is going to be *usually* the case that the agent will choose U in the situation, $\{U, \neg U\}$ iff she chooses *Yes* in the situation where she hears *Do you think that it is raining?* But this is not a *logical* requirement, only a pragmatic one.

We will say that an agent *endorses* (agrees with) a sentence ϕ iff she chooses *Yes* when asked, *Do you think* ϕ*?*, and *denies* (or disagrees with) ϕ iff she chooses *No.* She may also

[7] It is plausible that when a vervet monkey utters a leopard call, then it is saying, '*Ware leopard!*, but surely nothing in the behaviour of such monkeys would justify us to think that they might utter *If there were no leopards here, then this would be a wonderful spot to picnic.*

choose *Not sure*, in which case of course she neither endorses nor denies.

Note that nothing prevents an agent from endorsing ϕ as well as $\neg\phi$, but few agents are that 'out of it'. But of course an agent may endorse ϕ, $\phi \rightarrow \psi$, and either deny ψ or at least fail to endorse ψ. We will say in the first case that the agent is logically incoherent, and in the second that the agent is an incomplete reasoner – or simply incomplete.

Given the fact that $(\phi \wedge \neg\phi) \rightarrow \psi$ for arbitrary ψ is a tautology, it is obvious that an agent who is logically incoherent, but *not* incomplete, and endorses both ϕ and $\neg\phi$, will end up endorsing everything. Fortunately, most of us, though we *are* logically incoherent, tend also to be incomplete. If we endorse the statements that all men are equal, that Gandhi and Hitler are men, and that Gandhi and Hitler are not equal, we are still not likely to agree that pigs fly – which we would if we were also *complete.*

As we have made clear, elements of $\mathcal{B}$ cannot be identified with propositions, for an agent may agree to one sentence expressing a proposition and disagree (or not agree with) another sentence expressing the same proposition. An agent in some state $b \in \mathcal{B}$ may agree with *Hesperus is bright this evening*, while disagreeing with *Phosphorus is bright this evening.* But surely we *may* think of an agent speaking some language $\mathcal{L}$ as agreeing with or endorsing a particular *sentence* ϕ.

Definition 1.1 *A belief state b is* incomplete *when there are sentences $\phi_1, ..., \phi_n$ which are endorsed in b, ψ follows logically from $\phi_1, ..., \phi_n$, and ψ is not endorsed in b.*

A belief state b is incoherent *when there are sentences $\phi_1, ..., \phi_n$ which are endorsed in b, ψ follows logically from $\phi_1, ..., \phi_n$, and ψ is denied in b.*

There is of course an entirely different sort of incoherence which arises when an agent's linguistic behaviour does not comport with his choices. If an agent prefers not to get wet (which we knew somehow), says that it is raining, and does not take her umbrella, she may well be quite coherent in her linguistic behaviour, but her linguistic behaviour and her non-linguistic choices have failed to match. Whether we then conclude that the agent is irrational, or is being deceptive, or does not know the language, will depend on the weight of the evidence.

We can easily see now how deduction changes things. Meno's slave was initially both incoherent and incomplete. He believed that a square of side four had an area of eight. Since he knew some arithmetic and some geometry, his belief state was not actually coherent. At the end of the conversation with Socrates, he came to endorse the sentence, *The square whose side is the diagonal of a square of side two, has an area of eight.* It is not quite clear what practical application this information had for him in this case, but surely carpenters carrying out deductions of the same kind will manage to make furniture which does not collapse and bears the load of people and objects.

Beliefs can change not only as a result of deductions, as those Meno's slave did, they can also change as a result of experience, e.g., raindrops falling on your head, or as the result of hearing something, like *It is raining.*

2 Some technical details

We assume given a space $\mathcal{B}$ for some agent whose beliefs we are considering. The elements of $\mathcal{B}$ are the belief states of that agent, and these are not assumed to be sentences in Mentalese although for some restricted purposes they could be. There are three important update operations on $\mathcal{B}$ coming about as a result of (i) events observed, (ii) sentences heard, and (iii) deductions made. Elements of $\mathcal{B}$ are also used to make *choices*. Thus in certain states of belief an agent may make the choice to take his umbrella and we could then say that the agent believes it is raining. Many human agents are also likely to make the choice to *say*, "I think it is raining and so I am taking my umbrella" but clearly only if the agent is English speaking. Thus two agents speaking different languages, both of whom are taking their umbrellas, but making different noises, have the same belief in one sense but a different one in another. Both of these will matter. Later on we shall look into the connection.

Deduction is an update which does not require an input from the outside. But it *can* result in a change in $\mathcal{B}$. Suppose for instance that Jane thinks it is clear, and is about to leave the building without an umbrella. She might say, "Wait, didn't I just see Jack coming in with a wet umbrella? It must be raining." The sight of Jack with a wet umbrella might not have caused her to believe that it was raining, perhaps she was busy with a phone call. But the memory of that wet umbrella may later cause a deduction to take place of the fact that it is raining.

Thus our three update operations are:

$\mathcal{B} \times \mathcal{E} \rightarrow_e \mathcal{B}$

A belief state gets revised by witnessing an event.

$\mathcal{B} \times \mathcal{L} \rightarrow_s \mathcal{B}$

A belief state gets revised through hearing a sentence.[8] And hearing two logically equivalent sentences s, s' need not result in the same change occurring, although they may, in case the

[8] Even a dog may revise its state of belief on hearing *Sit!*

agent knows they are equivalent.[9]

$\mathcal{B} \rightarrow_d \mathcal{B}$

A deduction causes a change in the belief state (which we may sometimes represent as an *addition*).

Here E is the set of events which an agent may witness and $\mathcal{L}$ is some language which an agent understands (i.e., uses successfully). The last map $\rightarrow_d$ for deduction is non-deterministic as an agent in the same state of belief may make one deduction or another.

Finally, we also have a space $\mathcal{S}$ of *choice sets* where an agent makes a particular choice among various alternatives. This gives us the map

$\mathcal{B} \times \mathcal{S} \rightarrow_{ch} \mathcal{B} \times C$

An agent with a certain belief makes a choice among various alternatives.

If we want to explicitly include preferences, we could write,

$\mathcal{B} \times \mathcal{P} \times \mathcal{S} \rightarrow_{ch} \mathcal{B} \times C$

While $\mathcal{S}$ is the family of choice sets, C is the set of possible choices and $\mathcal{P}$ is *some* representation of the agent's preferences. Thus {*take umbrella, don't take umbrella*} is a choice set and an element of $\mathcal{S}$ but *take umbrella* is a choice and an element of C.

Among the choices that agents make are choices to assent to or dissent from sentences. But there is no *logical* reason why an agent who assents to "It is raining" must take an umbrella or a raincoat. It is just the more pragmatic choice to take the umbrella when one says that it is raining, because otherwise either one gets wet, or one is suspected of insincerity. We could say that the agent believes the sentence "It is raining", and dis-believes the proposition that it is raining. But we tend to be uncomfortable with such an account and prefer to say that the agent is either lying or confused.

Of course an agent may prefer to get wet and in that case, saying "It is raining," and not taking an umbrella are perfectly compatible choices. This shows that an agent's preferences need to be taken into account when correlating the actions which an agent takes and what an agent believes. But we usually do not want to get wet and to make such choices, and usually we do not say what we do not believe. It does not work for us.

Thus our theory of an agent presupposes such a belief set $\mathcal{B}$, and appropriate functions

[9] Clearly Lois Lane will react differently to the sentences *Superman flew over the Empire State Building*, and *Clark Kent flew over the Empire State Buildling.* Similarly, Kripke's Pierre will react differently to the questions, *Would you like a free trip to Londra?* and *Would you like a free trip to London?* Indeed in the second case he might *offer to pay* in order not to go to London!

$\rightarrow_e, \rightarrow_s, \rightarrow_d, \rightarrow_{ch}$. We can understand an agent (with some caveats) if what we *see* as the effects of these maps conforms to some theory of what an agent wants and what the agent thinks. And we succeed pretty well. *Contra* Wittgenstein, we not only have a theory of what a lion wants, and what it means when it growls, we even have theories for bees and bats. Naturally these theories do not have the map $\rightarrow_s$ except with creatures like dogs or cats and some parrots (who not only "parrot" our words but understand them to some extent [31]).

So far we have only talked about belief states as being indicated by choices. But of course many beliefs are expressed (or so we think) by sentences. We may suggest, with Stalnaker, that what we believe are propositions. Though I do not think such accounts can work, the fact that we find them appealing requires some explanation.

We shall now introduce two notions of belief, e-belief (which is choice-based) and i-belief (which is sentence-based and only appropriate for language users). These two notions will be used to sort out various belief puzzles.

3 The Setting

In our setting we imagine an observer o who is pondering on what some agent i believes. We assume (for convenience) that o thinks of a proposition expressed by a sentence as a set of possible worlds where that sentence is true, but that the observee i need not even have a language or a notion of truth. However, it is assumed that i does have some plans. Even if i is just a dog digging for a bone, o understands that i has a plan and roughly what that plan is. And we shall use this plan to make it possible for o to attribute beliefs to i.

We also assume that there is a context C which is the set of *relevant* possible worlds, and that worlds outside C, even though they are there, are not considered in deliberating about i's belief or beliefs. It is not assumed that i knows that there are worlds outside C; in some sense i lives inside C, but we will assume that o does know that there are possible worlds outside C. The purpose of the context is to avoid considering cases which are possible but strange, like the laws of physics failing, or Bush suddenly joining the Green party. A plan is assumed to work in *normal* circumstances, and an agent i is only required to have the sort of belief which would be enough for the plan to be carried out in normal circumstances.

For instance, chickens, when given a choice between drinking water and drinking mercury tend to prefer the latter, presumably because it looks more like water than water itself does. We certainly do not want to attribute to the chickens a belief that mercury is good to drink, merely that if you are thirsty, you should go for a shiny liquid.

So let P be i's plan at the moment, and let $\pi(P)$ be the set of worlds w in C such that the plan is *possible* at w.

There are two senses in which the plan may be possible. One is that the plan can actually be carried out. For instance for the dog digging for a bone, that possibility means that the ground is soft enough. The other sense is that the plan actually yields i the benefit which i wants; in this case, the actual presence of a bone.

So formally,
$\pi(P) = \{w|\ w \in C \wedge w\ enables\, P\}$.

Let ϕ be a sentence. Then $||\phi|| = \{w|w \models \phi\}$, the set of worlds where ϕ is true, is the *proposition* corresponding to the sentence ϕ. Note that if ϕ and ψ are logically equivalent, then $||\phi|| = ||\psi||$.

Definition 3.1 *We will say that i* e-believes *ϕ, $B_e^i(\phi)$ if $\pi(P) \subseteq ||\phi||$. We will suppress the superscript i when it is clear from context.*

It is obvious in terms of the semantics which we just gave that the statements "The dog e-believes that there is a bone where he is digging".

It is easy to see that if an agent e-believes ϕ and ψ then the agent also e-believes $\phi \wedge \psi$ and that if the agent e-believes ϕ and $\phi \rightarrow \psi$ then the agent e-believes ψ.[10] Oddly enough, creatures which do not use language do not suffer from a lack of logical omniscience!

A lot of logic goes along with e-belief, but only within the context of a single plan. For instance, one may drive a colleague to the airport for a two week vacation in Europe and then forget, and arrange a meeting three days later at which this colleague's presence is essential. But within the context of a single (short) plan, consistency and logical omniscience will tend to hold. The situation is more complex with multiple plans. And there is nothing to prevent an agent from having one e-belief in one plan and another contradicting e-belief in another plan. It is pragmatic considerations – the logical counterpart of avoiding Dutch Book – which will encourage the agent i to be consistent and to use logical closure.

Suppose someone has a plan P consisting of, "If ϕ then do α, else do β" and another plan P' consisting of "If ϕ then do γ, else do δ". Now we find him doing α and also doing δ (we are assuming that the truth value of ϕ has not changed). We could accuse him of being illogical, but there is no need to appeal to logic. For he is doing Dutch book against himself.

[10]To see this, if $\pi(P) \subseteq ||\phi||$, and $\pi(P) \subseteq ||\psi||$ then clearly $\pi(P) \subseteq ||\phi|| \cap ||\psi|| = ||\phi \wedge \psi||$. The proof for the other case is similar using the fact that $||\phi \rightarrow \psi|| = (C - ||\phi||) \cup ||\psi||$. Since $\pi(P)$ is contained in $||\phi||$, it is disjoint from $C - ||\phi||$. Hence it can be contained in $(C - ||\phi||) \cup ||\psi||$ if and only if it contained in $||\psi||$.

Presumably he assumed[11] that $u(\alpha|\phi) > u(\beta|\phi)$ but $u(\alpha|\neg\phi) < u(\beta|\neg\phi)$. Thus given ϕ, α was better than β but with $\neg\phi$ it was the other way around. Similarly, $u(\gamma|\phi) > u(\delta|\phi)$, but $u(\gamma|\neg\phi) < u(\delta|\neg\phi)$. And that is why he had these plans. But then his choice of α, δ results in a loss of utility whether ϕ is true or not. If ϕ is true then he lost out doing δ and if ϕ is false, then he lost out doing α.

For a concrete example of this, suppose that on going out I advise you to take your umbrella, but fail to take mine. If it is raining, there will be a loss of utility for I will get wet. If it is not raining, there will be a loss of utility because you will be annoyed at having to carry an umbrella for no good reason. My choice that I advise you to take your umbrella, but fail to take mine, is not *logically* impossible. It just makes no pragmatic sense. A similar argument will apply if someone endorses ϕ, endorses $\phi \rightarrow \psi$ and denies ψ. If such a person makes plans comporting with these three conditions, then he will make choices which do not maximise his utility. Of course such arguments go back to Ramsey [37] and Savage [38].

If we assume that over time, people learn to maximise their utility (they do not always but often do), then they will 'learn' a certain amount of logic and they will make certain obvious logical inferences.[12]

A little girl who was in kindergarten was in the habit of playing with her older sister in the street every day when she came home. Once, her older sister was sick. So the little girl went to visit her sick sibling in her bedroom, and then, as usual, went out into the street to play with her. Clearly the little girl had not yet learned to maximise her utility!

4 A second notion of belief – language enters

We now define a second notion of belief which does *not* imply logical omniscience. This is a more self-conscious, language-dependent notion of belief.

For agents i who do have a language (assumed to be English from now on), their plan may contain linguistic elements. At any moment of time they have a finite stock of currently believed sentences. This stock may be revised as time passes. These agents may perform atomic actions from time to time, and also make observations which may result in a revision in their stock of believed sentences.

Thus Lois seeing Superman in front of her will add the sentence "Superman is in front of

[11]The use of the letter u for utility is not meant to suggest that we have a formal notion of utility in mind; only a rough one.

[12]In [4] Bicchieri suggests that co-operation also comes about as a result of such a learning process. Such suggestions have of course also been made by many others. Since we are only considering the one-agent case here, we shall not go into this issue any further. See, however, our [24].

me", to her stock, but, since she does not know that Clark Kent is Superman, she will *not* add the sentence "Clark Kent is in front of me". Someone else may add the sentence "I see the Evening Star", but not the sentence "I see the Morning Star" at 8 PM on a summer night. A person who knows that $ES = MS$, may add the sentence, "Venus is particularly bright tonight." In any case, this stock consists of sentences and not of propositions.

The basic objects in the agents' *plans* are atomic actions and observations which may be active (one looks for something) or passive (one happens to see something). These are supplemented by the operations of concatenation (sequencing), *if then else*, and *while do*, where the tests in the *if then else* and *while do* are on *sentences.* There may also be recursive calls to the procedure: *find out if the sentence ϕ or its negation is derivable within the limits of my current resources, from my current stock of beliefs.* Thus if i's plan has currently a test on ϕ, then, to be sure, the stock of sentences will be consulted to see if ϕ or its negation is in the stock. But there may also be a recursive call to a procedure for deciding ϕ. If someone asks "Do you know the time?", we do not usually say, "I don't", but look at our watches. Thus consulting our stock of sentences is typically only the first step in deciding if some sentence or its negation can be derived with the resources we have.

This difference between sentences and propositions matters as we now show.[13]

[13]We might compare entertaining a proposition as a bit like entering a building. If Ann and Bob enter the same building through different doors, they need not be in the same spot, and indeed they might never meet. But what makes it the same building is that they *could* meet without going out into the street. Thus if Ann and Bob are apt to assent to sentences s, s' respectively where *we* know that s, s' are equivalent; then it need not follow that there is a sentence they share. But they *could* through a purely deductive process, and without appealing to any additional facts, be brought to a situation where Ann assents to s' and Bob to s (unless one of them withdraws a belief, which may also happen).

It has been suggested in this context (e.g. by Stalnaker [41]) that such issues can be addressed by using the notion of fine grain. By this account, if I understand this correctly, logical equivalence is too coarse a grain and that a finer grain may be needed. So if two sentences are in the same fine grain and an agent knows or believes one, then the agent will also know or believe the other. But if we try to flesh out this metaphor then we can see that it is not going to work.

For instance if $G(\phi, \psi)$ means that ϕ and ψ are in the same grain, then G will be an equivalence relation. But surely we cannot have transitivity in reality, because we could have a sequence ϕ_1,...,ϕ_n of sentences, any two successive ones of which are easily seen to be equivalent, whereas it is quite hard to see the equivalence of ϕ_1 and ϕ_n.

Moreover, "being in the same grain" sounds interpersonal. If two molecules are in the same rice grain for you, then they are also in the same fine grain for me – it is just a fact of the matter. But in reality people differ greatly in their ability to perceive logical equivalences.

Thus suppose some set theorist thinks of some new axiom ϕ and wonders if ϕ implies the continuum hypothesis, call it ψ. The set theorist may find it quite easy to decide this question even if we see no resemblance between ϕ and ψ. And if he does not find it easy, he may look in the literature or ask another set theorist, processes which cannot easily be built into a formal theory. And they should not be! For if they were easy to build into a formal theory, then the representation is almost certain to be wrong.

Or a chess champion may be able to see 20 moves (40 half-moves) ahead in the end game, but you and I cannot. And he too usually cannot do this during the middle game. Thus context, habit and expertise matter a lot.

Suppose for instance that Lois Lane has invited Clark Kent to dinner but he has not said yes or no. So she forms the plan, *While I do not have a definite answer one way or another,* if *I see Clark Kent, I will ask him if he is coming to dinner.* Here *seeing Clark Kent* is understood to consist of an observation followed by the addition of the sentence "I am seeing Clark Kent" to her stock.

Suppose now that she sees Superman standing on her balcony. She will *not* ask him if he is coming to dinner as the sentence "I am seeing Clark Kent" will not be in her stock of sentences. And this is the sense in which she does *not* know that when she is seeing Superman, she is also seeing Clark Kent. If she *suspects* that Clark Kent is Superman, then it may happen that her recursive call to the procedure "decide if I am seeing Clark Kent" will take the form of the question, "Are you by any chance Clark Kent, and if so, are you coming to dinner?" addressed to Superman.

Have we given up too much by using sentences and not propositions as the objects of i-belief? Suppose her plan is, "If I see Jack and Jill, I will ask them to dinner" but she sees Jill first and then Jack so that she adds the sentence "Jill and Jack are in front of me" to her stock. Will this create a problem? Not so, because if a sentence ϕ is in her stock, ϕ easily implies ψ, and she needs to know the value of ψ so she can choose, then the program *find out about* ψ which she calls will probably find the sentence she needs. If the program terminates without yielding an answer she may well have a default action which she deems safest.

So here we make use of the fact that Lois does have a reasonable amount of intelligence. Even if she does not explicitly add some sentence ψ to her stock, when she comes to a point where the choice of action depends on ψ, she *will* ask if ψ or its negation is derivable from her present stock of sentences, possibly supplemented by some actions which add to this stock.

Definition 4.1 *If an agent a comes to a point in her plan where her appropriate action is* If ϕ then do α else do β*, and she does α, then we will say that she* i-believes ϕ*. If, moreover, ϕ is true, and we believe that in a similar context she would judge it to be true only if it* is *true, then (within the context of this plan) we will say that she* i-knows ϕ*.*

A common example of such a plan is the plan to answer a question correctly. Thus if an agent is asked "Is ϕ true?", the agent will typically call the procedure "decide if ϕ is true", and then answer "yes", "no", or "I don't know" in the appropriate cases.

Now note that if an agent means to deceive, then the same procedure will be called, but the answers given will be the opposite of the ones indicated by the procedure. But if we ourselves know that the agent's plan is to deceive, then we can clearly take the statement

"ϕ is false" to indicate that the agent believes ϕ.

We no longer have the law that if the agent i-knows ϕ and ϕ implies ψ then the agent of necessity i-knows ψ. But if the agent has the resources to decide ϕ and the proof of ψ from ϕ is easy, then she might well also know ψ. But her notion of "easy" may be different from ours, and how much effort she devotes to this task will depend on her mood, how much energy she has, etc.

For instance if a customer sits down on a bar stool and the bartender sees him, we do not need to ask, "Does the bartender know he wants a drink?" Of course he knows. But suppose a woman suffering from a persistent cough calls her husband and says, "I got an appointment with the doctor for 2:30 PM", she may later think, "I wonder if he realized that I cannot pick up our daughter at 3". When i knows ϕ and ψ is deducible from ϕ, then whether we assume that i knows ψ will depend less on some objective distance between ϕ and ψ than on what we know or can assume about i's habits.

Note now that we often tend to assign knowledge and beliefs to agents even when they are not in the midst of carrying out a plan. Even when we are brushing our teeth we are regarded as knowing that the earth goes around the sun. This can be explained by a continuity assumption. Normally when we *are asked* if the earth goes around the sun or vice versa, we say that the former is the case. An agent is then justified in assuming that even in between different occasions of being so asked, if we *were* asked, we would give the same answer. This assumption, which is usually valid, is what accounts for such attributions of belief.

Acknowledgements: We thank Sergei Artemov, Samir Chopra, Horacio Arlo Costa, Juliet Floyd, Haim Gaifman, Isaac Levi, Mike Levin, Larry Moss, Eric Pacuit and Catherine Wilson for comments. The information about chess came from Danny Kopec. This research was supported by a grant from the PSC-CUNY faculty research assistance program. Earlier versions of this paper were given at *TARK-05*, *ESSLLI-2006*, at the *Jean Nicod Institute*, and at a seminar in the philosophy department at Bristol University. Some of the research for this paper was done when the author was visiting Boston University and the Netherlands Institute for Advanced Study.

References

[1] 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

[2] Artemov, S., and E. Nogina, "On epistemic logics with justifications", *Theoretical Aspecits of Rationality and Knowledge*, ed. Ron Meyden, University of Singapore press, (2005), 279–294.

[3] Aumann, R., "Agreeing to disagree", *Annals of Statistics*, **4** (1976) 1236-1239.

[4] Cristina Bicchieri, "Learning to co-operate," in *The Dynamics of Norms*, ed. Biccheri, Jeffrey, Skyrms, Cambridge1997, pp. 17-46.

[5] Brandom, Robert, "Unsuccessful seminatics", *Analysis*, **54** (1994) 175-178.

[6] Chopra, S., and L. White, "Attribution of Knowledge to Artificial Agents and their Principals", *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, Edinburgh, August 2005, forthcoming.

[7] Michael A. E Dummett *The justification of deduction*, (Henriette Hertz Trust annual philosophical lecture ; 1973).

[8] Fagin, R., Halpern, J., Moses, Y. and Vardi, M., *Reasoning about knowledge*, M.I.T. Press, 1995.

[9] de Finetti, Bruno, "Foresight: its logical laws, its subjective sources", *Annales de l'Institut Henri Poincare*, **7** (1937). Translation by Henry Kyburg in *Studies in Subjective Probability*, ed. Kyburg and Smokler, Krieger publishing company, (1980) 53-118.

[10] Fitting, M., "A logic for explicit knowledge", to appear in proceedings of *Logica 2004*.

[11] Gaifman, H., "Reasoning with limited resources and assigning probabilities to arithmetical statements", *Synthese*, **140** (2004) 97-119.

[12] Hayek, F.A., *Individualism and Economic Order*, University of Chicago Press (1936). See especially chapters II and IV.

[13] David Hume, *A Treatise of Human Nature*, ed. Selby-Brigge, Oxford U. Press 1978, 176-179.

[14] Daniel Kahneman, "Maps of Bounded Rationality," Nobel prize lecture, 2002.

[15] S. Kripke, "A Puzzle about belief," in *Meaning and Use*, ed. A. Margalit, Reidel 1979.

[16] I. Levi, *The Enterprise of Knowledge*, MIT press 1980.

[17] R. Marcus, "Some revisionary proposals about belief and believing," *Philosophy and Phenomenological Research*, **50** (1990) 133-153.

[18] R. Marcus, "The anti-naturalism of some language centere accounts of belief," *Dialectica*, **49** (1995) 112-129.

[19] M. MicKinsey, "The semantics of belief ascriptions," *Nous* **33:4**, (1999) 519-557.

[20] Ruth Millikan, "Styles of Rationality", in *Rationality in Animals*, ed. M. Nudds and S. Hurley, Oxford 2006, pp. 117-126.

[21] Y.Moses, "Resource bounded knowledge," in *Proc. Theorectical Aspects of Reasoning about Knowledge*, ed. M. Vardi, Morgan Kaufmann 1988, pp. 261-276.

[22] Nozick, R., *Philosophical Explanations*, Harvard University Press, 1981.

[23] R. Parikh, "Knowledge and the Problem of Logical Omniscience" *ISMIS- 87* (International Symp. on Methodology for IntelligentSystems), North Holland (1987) 432-439.

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

[25] R. Parikh, "Propositions, propositional attitudes and belief revision" in K. Segerberg, M. Zakharyaschev, M. de Rijke, H. Wansing, editors, *Advances in Modal Logic, Volume 2*, CSLI Publications, 2001.

[26] R. Parikh, "Logical omniscience" in *Logic and Computational Complexity*, LNCS 960, 22-29.

[27] R. Parikh, Knowledge based computation (Extended abstract), in *Proceedings of AMAST-95*, Montreal, July 1995, Edited by Alagar and Nivat, Lecture Notes ib Computer Science no. 936, 127-42.

[28] R. Parikh, "Logical omniscience", in *Logic and Computational Complexity* Ed. Leivant, Springer Lecture Notes in Computer Science no. 960, (1995) 22-29.

[29] R. Parikh, "Social Software", *Synthese*, **132**, Sep 2002, 187-211.

[30] R. Parikh, Levels of knowledge, games, and group action, in *Research in Economics*, **57**, (2003) 267-281.

[31] Irene Pepperberg, "Talking with Alex: Logic and Speech in Parrots; Exploring Intelligence," *Scientific American Mind,* August 2004.

[32] R. Parikh, and P. Krasucki, "Levels of knowledge in distributed computing", *Sadhana - Proc. Ind. Acad. Sci.* **17** (1992) 167-191.

[33] Pacuit, E., and R. Parikh, A Logic for communication graphs, to appear in the proceedings of *DALT 2004*.

[34] R. Parikh, and R. Ramanujam, A Knowledge based Semantics of Messages, in *J. Logic, Language and Information*, **12**, (2003) 453-467.

[35] *Meno*, by Plato, translation by Benjamin Jovett. Available online at http://classics.mit.edu//Plato/meno.html

[36] Ramsey, F. P., "Facts and propositions", in *Philosophical Papers*, edited by D.H. Mellor, Cambridge U. Press 1990, 34-51.

[37] Ramsey, F.P., 'Truth and probability', in *The Foundations of Mathematics*, Routledge and Kegan Paul (1931), 156-198.

[38] L. J. Savage, *The Foundations of Statistics*, Wiley 1954.

[39] S. Schiffer, "Propositional content", in *The Oxford Handbook of Philosophy of Language*, ed. E. Lepore and B. Smith, OUP 2006, 267-294.

[40] E. Schwitzgebel, "A phenomenal, dispositional account of belief," *Nous*, **36:2** (2002) 249-275.

[41] Stalnaker, R., *Context and Content*, Oxford University Press, 1999.

[42] Frans de Waal, *Our Inner Ape*, Penguin 2005.

[43] Whyte, J.T., "Success semantics", *Analysis*, **50** (1990), 149-157.

[44] Wittgenstein, L., *Philosophical Investigations*, MacMillan, 1958.