Belief, Behavior and Bisimulation

Rohit Parikh

Brooklyn College and CUNY Graduate Center
City University of New York

**Workshop on Practice-based Philosophy of
Logic and Mathematics**
August 31, 2009

The following quotes are from the Meno by Plato

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.*

And later…

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.*

But did the boy actually know Pythagoras' theorem before he was led through this argument by Socrates?

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.

The "conjunction fallacy" is committed when someone assigns higher probability to a conjunction than to one of the conjuncts.

Since our ability to deal with ours and other people's beliefs depends on having realistic models, the issues raised by Kahneman and others before him are important.

And how do our current models deal with such problems?

One popular model is the Kripke structure

Kripke Structures

A common 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,

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

Logical Omniscience

It follows from the previous definition that

- If a formula is logically valid then it is true at all states and hence it is both known and believed.
- If ϕ and $\phi \rightarrow \psi$ are believed then ψ is also believed at s.
- A logically inconsistent formula can be neither known nor believed.
- The set of formulas believed must be logically consistent

But, as we saw from the examples, beliefs can be inconsistent, and logical consequences of beliefs might not themselves be beliefs

The second part is actually **good!**. For if we have two inconsistent beliefs and our beliefs were closed under logical consequence, we would believe that pigs fly, that Sarah Palin is the president of the US and that sand is good to eat.

The three stances of Dennett's

- **The physical stance**
- **The design stance**
- **The intentional stance**

A fourth stance

- **The dispassionate agent stance**

Past logics of knowledge have tended to ignore the fact that communications (which affect knowledge and beliefs) usually have some purpose. For animals, the purpose tends to be immediate, like warning others about the presence of a predator. For humans, it can be more long term, like planning a picnic. But as Grice and others after him have emphasized, purpose tends to enter into communication.

Real Patterns

Do beliefs exist?
Three views on belief

1. **Beliefs are real,** they are written in our brains
2. **Beliefs are a** *facon de parler,* **we invent them to explain certain forms of behaving**
3. **Beliefs do not exist in any sense,** they should be explained away

Real Patterns

- Does the equator exist?
- Does the Democratic party exist?
- Does the earth have a center of gravity?
- Does the US have a center of population?
- The US lies north of Mexico and south of Canada. Did an entity lying north of Mexico and south of Canada invade Iraq?

When we go to Ecuador, we do not see any red line running through the equator. Nonetheless, it does sort of exist, it does not have the same status as the phoenix or the square cube.

Is the Democratic party a set?

A little while ago, a prominent Republican senator, Areln Specter shifted to the Democrats.

Let X = The set of Democrats without Arlen Specter. I.e., the party before he joined it.

Let Y = The set of Democrats with Arlen Specter. I.e., the party after he joined it.

By definition, Specter cannot be a member of X.

By definition, Specter does not *need* to join Y.

So which set **did** he join?

But we do not doubt that the Democratic Party exists, sort of.

Center of gravity etc.

Any object which is attracted to the earth from any direction is attracted to a specific point which is the center of gravity of the earth.

So the center of gravity does exist and has a use.

But despite what Dennett says, the notion of **center of population of the US** does not make sense.

He draws a line A such that half the Americans are to the North of A and half to the South. He draws a line B such that half the Americans are to the East of A and half to the West.

Dennett's center of population P is the intersection of A and B.

But is P actually the center of population?

For suppose we draw a third line C passing through P. Will half the population lie on one side and half on the other?

Not necessarily

100

100

A

P

C

B
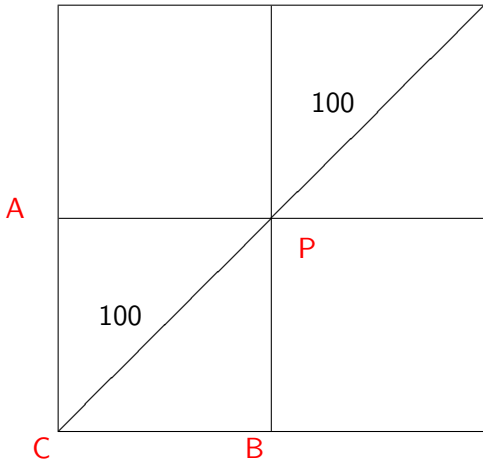
Half the population does indeed lie on one side of line A and half on the other side. Similarly for line B.

But all of the population lies on one side of line C.

So the "center of population" does not have the nice properties we would want.

When we use a noun phrase (like belief),
we have certain expectations of the properties of the
thing in question.

But the thing may fail to have all of the required
properties

Or it may be that we use a noun ambiguously,
and the various denotations of the noun may
engage in a division of labour;
some denotations meeting some of our needs
and others meeting others.

Beliefs and Preferences

Ann is going out and picks up her umbrella. From this we infer two things.

- **Ann believes it is raining or is going to rain.**
- **Ann prefers not to get wet.**

Standard accounts in decision theory (like that of Leonard Savage) impose certain axioms on an agent's behavior (including betting behavior on lotteries) and deduce two things from that behavior.

- **The agent's beliefs - specifically the agent's subjective probability.**
- **The agent's preferences - specifically the agent's utility function.**

The agent's choices are seen as the agent attempting to maximise his expected utility.

Behavioral Economics

- It is well known that axioms like those of Savage are not followed universally.
- It is also well known that agents acting 'rationally' in Savage's sense will make bad choices in settings like the Prisoner's Dilemma.
- Perhaps we can take the rough framework of Savage's theory without accepting his axioms.

That people make their choices in view of what they believe and what they prefer seems to be common sense.

Without such an assumption, we would not be able to make sense of people's (and animals') behavior at all.

Why did the mouse not go right?

Because he saw the cat over there.

Why did the mouse go left?

Because there was a piece of cheese on the left.

Do we want to assign beliefs to the mouse?

Does the BDI theory apply to the mouse?

Do animals have beliefs?

Davidson has argued that animals cannot have beliefs because they lack certain knowledge.

For instance, a dog digging for a bone cannot have a belief that there is a bone buried where he is digging, for he lacks the concept of a bone.

But, famously, Putnam did not know the difference between a beech and an elm.

We would surely not argue that Putnam lacks beliefs, or even beliefs about trees.

Tiger attacks in the Sundarbans

Fishermen and bushmen originally created masks made to look like faces to wear on the back of their heads because tigers always attack from behind. The mask induced a false belief in the tiger.

In 1987 no one wearing a mask was killed by a tiger, but 29 people without masks were killed.

Eventually the tigers realized it was a hoax, and the attacks resumed.

Nonetheless, the fact that people did not take Davidson too seriously saved quite a few lives in 1987.

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.

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.

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.

$\mathcal{B} \rightarrow_d \mathcal{B}$
A deduction causes a change in the belief state (which we may sometimes represent as an **addition**.

Elements of $\mathcal{B}$ are also used to make **choices**.

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.

States and Traits

It is customary among philosophers to talk about mental states. But when we consider a popular model, namely Kripke structures, then much of the information is contained not in states but in the accessibility relation **R**.

Moreover, even this leaves out another important property of people's psychology, namely their traits.

Thus a person can have the tendency to become angry. But that tendency need not say anything about the agent's state **right now!** Rather it tends to mean that the agent becomes angry quite easily.

If the Lion Could Speak

If the lion did speak, we *might* well understand him on certain occasions. If the lion growls at us, we know he means, *Scram!*

If he comes close to us and purrs, we know he means *Scratch my head!*

But sometimes we don't even know what *people* mean!

Bisimulation

Given two structures $\mathcal{M}$ and $\mathcal{M}'$ we can say they are *isomorphic* if there is a 1-1 function f from the domain of one onto the domain of the other which preserves all relations. For instance we will have that $R(a, b) \leftrightarrow R'(f(a), f(b))$.

Bisimulation is a weaker notion.

Bisimulation - cont'd

In theoretical computer science a bisimulation is a binary relation between state transition systems, associating systems which behave in the same way in the sense that one system simulates the other and vice-versa.
Intuitively two systems are bisimilar if they match each other's moves. In this sense, each of the systems cannot be distinguished from the other by an observer.

Bisimulation - cont'd

For instance, suppose that Jill has a son and two daughters. Ann has a daughter and two sons.

Suppose all the daughters have sons each. All the sons are childless.

Then the structures M_1 consisting of Ann and her progeny and M_2 consisting of Jill and her progeny are bisimilar but not isomorphic.

If we ask of each of Ann and Jill, Do you have a son? Then the answer is yes in both cases. Does a son have a child? Then the answer is no in both cases.

The question How many sons do you have? is not allowed to be asked in our language.

Bisimulation - cont'd

Suppose we are given two belief structures $(B, \rightarrow_e, \rightarrow_{ch})$ and $(B', \rightarrow'_e, \rightarrow'_{ch})$. Then they are bisimilar if there is a binary relation $R \subseteq B \times B'$ such that if f is an event, $R(x, y)$ and $x \rightarrow_e (f)z$ then there is a w in B' such that $R(z, w)$ and $t \rightarrow_e (f)w$. Similarly, if g is an event and $y \rightarrow_e (g)p$ then there is an $r \in B$ such that $x \rightarrow_e (g)r$ and $R(p, r)$.

Thus states in B and B' which are R-related transform to R related states when an event happens.

Two structures which are bisimilar need not satisfy the same first order formulas, but they do satisfy the same modal formulas.

We can impose similar bisimilarity conditions on the $\rightarrow_{ch}$ relations so that in corresponding states, the system B and the system B' make the same choice.

Partial Bisimulation

But two systems B and B' might simulate each other only partially.

If Jack puts on a sweater when going out, Ann may do the same thing.

On the other hand if Jack chooses vanilla between vanilla and chocolate, Ann might prefer chocolate.

Partial Bisimulation

Jack may also have a subsystem which bisimulates Ann to some extent. Jack may then say, "I know she is going to pick chocolate though God only knows why!"

Tina Fey may bisimulate Sarah Palin without actually knowing what Palin thinks.

This is a rich topic and we cannot go into the details of it.

Theory of Mind

Ever since the 1978 work of Premack and Woodruff, there has been some excitement over the issue of theory of mind (TOM). Do chimps have it? Is this a clear question?

There is some experimental evidence that chimps are aware of what other chimps (or even people) can and cannot see. There are other contexts in which chimps seem to fail certain tests.

Is it possible for us to clarify these questions so that either we are able to answer them or perhaps show that there is a gradation and how various degrees of TOM can be measured.

Is this a task which logicians can address?

Some references:

1. Daniel Dennett, *The Intentional Stance*, MIT press, 1987.

2. Rohit Parikh, Sentences, belief and logical omniscience, or what does deduction tell us?, *Review of Symboloic Logic*, **1:4**, 2008, 459-476.

3. Leonard Savage, *The Foundations of Statistics*, Wiley 1954.

4. Eric Schitzgebel's article on *Belief* in the *Stanford Encyclopedia of Philosophy*.

**Conference on Eastern and Western
Philosophical Themes**

City University of New York
Dec 4-5, 2009
http://web.cs.gc.cuny.edu/∼kgb/