> Folk Psychology, the Intentional Stance, and Theory of Mind

Jonathan Simmons

DECEMBER 11, 2009

Jonathan Simmons

3240619

December 11, 2009

PHIL 3141

Dr. Flagel

Folk Psychology, the Intentional Stance, and Theory of Mind

Introduction

Folk psychology, the intentional stance, and theory of mind are often used interchangeably, and the goal of this paper is to (1) understand their distinctions, (2) understand how they relate to one another, and (3) to explore Daniel Dennett's contributions to the philosophy of mind and a scientific study of consciousness. It should be admitted from the beginning that this paper takes a positive view of Dennett's intentional stance, and in addition to elucidating his work, it will offer a modest defense against some of his critics. With apologies to the reader, at least a portion of the

following examination will be colored by disdain for attempts to build unnecessary mystery into consciousness; consciousness is quite fascinating all by itself without the influence of wishful thinking.

In developing his intentional stance (which I will elaborate on later) Dennett was inspired by Ryle, Wittgenstein, Quine, and Putnam. Wittgenstein's influence is perhaps the most obvious when Dennett shies away from mental descriptions. From Quine, Dennett gets his relationship to the natural sciences. It is the role that Quine's influence has played that is probably the most interesting because of how Dennett rails against Wittgenstein's stronger statements about critical conceptual philosophy (the therapeutic approach) which has a nasty habit of leading to "contempt for natural science" (Miguens, 2006, p. 3). Dennett has embraced a hybrid approach to his work, blending philosophy and cognitive science and this combination has led to his characterization of mind as "incorrigible self-access to functional states" (p. 4).

Theory of Mind

Theory of mind is primarily concerned with mental attribution to others, and has been described as third-person mindreading. This paper will not attempt to explain how we go about processing these mental states beyond the acknowledgement that we use mentalistic terms, which involve some sort of interpretation. It suffices to say that the theory of mind is an innate module in the brain, the details of which are not

precisely understood outside of our taken-for-granted assumptions about our own minds.

We seem to attribute minds to the things we are interpreting, especially other beings, and this is of course an important topic for philosophers, but there is also a strong interest within the scientific community, particularly among cognitive ethologists, to find evidence of theory of mind in other animals. Theory of mind is also of concern for researchers interested in artificial intelligence, and how humans interact with machines. If, for example, we regard a machine as machine-like, we tend to want to avoid "complicated negotiation" and we do not "want to discuss ambiguous things" (Ito and Terada, 2007, p. 802). Much of Dennett's work involves getting around the potential biases inherent in the observer (such as thinking of machines as not having a theory of mind) and a criticism of representationalism, i.e., "for there to be mind there need be no such thing as real inner representations in those physical systems we take to be rational agents" (Miguens, 2006, p.1).

Heyes (1998) observes that studies which support the position that non-human primates have a theory of mind, as indicated by such behaviors as imitation, deception, and role playing, have been explained away as having "occurred by chance or as a product of non-mentalistic processes such as associative learning or inferences based on non-mental categories" (p.101). Dennett seems inconsistent when it comes to attributing a theory of mind to non-human animals, and Andrews (2000) argues that we need a more pluralistic view of behavior prediction than the intentional stance

because "we should not be able to learn facts about animals minds almost purely *a priori*" (2000, p.13).

Dennett has made efforts over the course of his career to put aside thought experiments in favor of field experiments, particularly with regard to vervet monkeys. He learned from these explorations that his methods, which work well on people, are "limited in application to animals who cannot be prepared as subjects with the help of language" (1998, p. 306). Preempting his critics, Dennett points out that it is "the intentional stances' rationality assumption that generates ("a priori" as it were) the consequent to be tested" (1998, p. 316). This of course is quite challenging in practice because of the vast number of anomalies possible in an animal's behavior (the same can be said of humans of course, but researchers have the advantage of also being human). Dennett was tempted to move the monkeys into a lab to test his more interesting hypotheses, and those experiments are of course being done by scientists, but being isolated raises a number of issues in the sense that it is not the vervet monkey's natural environment and human subjects are much better at adapting to the artificial (or so we think).

Surprisingly, Dennett thinks that wondering whether or not the vervet monkeys or other animals believe anything is a misguided approach. He is not confident that there are any *deeper facts* about human beliefs, let alone monkey beliefs (1998, p. 321). He suggests that the intentional stance does not lead to "objective truth about the contents of animals minds" (1998, p. 322). While this may seem like a weakness, the

kind of curiosity that concerns itself with deeper facts in humans, other animals, and artifacts moves beyond the realm of cognitive science and is essentially a "wild goose chase" (1998, p. 322). Regardless of whether or not this concern with deeper facts is a pointless endeavor, it is more parsimonious to accept that other animals do indeed have a theory of mind, as opposed to the alternative explanation which suggests that humans are not only unique among *Hominidae*, but among all other creatures on the planet.

Folk Psychology

Folk psychology is typically described as a “theory of human psychology implicit in our everyday talk about mental states and behavior” (Ravenscroft, 2005, p. 193), but Dennett argues that it is more like a craft than a theory, and that what we call the theory is the "ideology about the craft" (1998, p. 82). He argues that there is no controversy about the productivity of the anticipations we have regarding the behavior of other humans, and he thinks that this is to some degree innate rather than occurring as a result of a series of experiences. He compares folk psychology to folk physics, concluding that while we may not be able to enunciate what we know, our expectations regarding the behavior of liquids are fairly accurate. This is not to say that our beliefs are always correct: our intuitions are often wrong, as proven by academic physics (much of which is counter-intuitive). The same might be said for folk psychology, and "the craft itself will come to be adjusted to acknowledge the existence of perplexities and

peculiarities and contrary predictions that the craft had never before made" (1998, p. 84).

The ideology of folk psychology is different in that it consists of what others have "actually told us the craft was all about when they enunciated the lore" (1998, p. 83). This lore is for the most part Cartesian, even though contemporary philosophers and scientists are materialists. Dennett argues that it does not matter if "the folk ideology about the craft is wrong--unless you take the ideology too seriously!" (1998, p. 85).

The Intentional Stance

For Dennett, the intentional stance is the craft, though it is more technical. He approaches the intentional stance with the contention that beliefs and desires are not reducible to brain states, and that the idea that our intuitive notions are in the head somewhere is a false ideology.

As far as Dennett is concerned, anything that is predictable from the intentional stance is an intentional system, and questions that have to with whether or not a thing can be said to have a mind can be answered by analyzing the logical presuppositions of our attributions. This is a complicated way of saying that we should treat the subject of our study as if it is rational with the goal being the prediction and interpretation of behavior.

This is contrasted with the design stance and the physical stance. The latter ecompasses the methods of the physical sciences to make predictions about how any physical thing will behave, and it is often the only available strategy for things that are neither “alive nor artifacts” (2009, p.2). The former is another type of prediction, but specific to designed objects. More assumptions need to be made with the design stance, which can be dangerous in the case of badly designed objects or objects that malfunction.

Dennett describes the intentional stance as a subspecies of the design stance in that the designed object is “treated as an agent of sorts, with beliefs and desires and enough rationality to do what it ought to do given those beliefs and desires” (2009, p. 3). We use the intentional stance to explain the behavior of humans, other animals, computers and other complex artifacts; it is not difficult to see how this would provide us with certain advantages in terms of our ability to quickly make predictions about what the human or artifact under consideration is going to do without jumping into the design or physical stances. You might be wondering whether or not this is just a useful metaphor, one we have come to embrace through evolution as a way to efficiently parse our environment (it is not a good idea to stop suddenly on the Serengeti and analyze the approaching lion’s behavior using either the physical or the design stance).

Dennett argues to the contrary that the use of the intentional stance is not just metaphorical, and he raises an important question to illustrate his point: How exactly are we to distinguish between intentionality that is metaphorical and intentionality that

is intrinsic? When we refer to a computer as a *she* and anticipate *her* actions according to our own rationality, are we doing so because there is fecundity in metaphor or because she actually has intentions? Most people would agree that a computer does not have intentions, i.e., beliefs, desires, etc., but how are we to tell?

Original Intentionality

Clearly a shopping list is embedded in our practices, i.e., there is no need to make any attributions regarding a list of the number of food items we want to buy (derived meaning), but can we say the same for robots? Admittedly, our robots are not as sophisticated as we might like at this point, but as seen in robots like HRP-4C, developed by Japan's National Institute of Advanced Industrial Science and Technology (AIST), it is not hard to imagine a robot that not only resembles a human (as in the case of HRP-4C), but behaves in a human-like way, such as the Korean robot EveR-2, which has the ability to express emotions and sing.

Being able to mimic human behavior is of course not sufficient to support the notion of intrinsic intentionality in robots but is it so unreasonable to imagine robotic advances that separate the artifact's trajectories from its dependence on humans? Dennett gives an example of a robot poker player that is able to bluff a human opponent. It would seem to be "guided by internal states that function just as a human poker player's intentions do" (2009, p. 9).

Why can't the robot poker player have original intentionality? There are no easy answers to this question, but it might be helpful to examine human intentionality for a moment. We would like to think that it was implanted in us, that we are special, unique, and even miraculous, but as Dennett argues, it must have "evolved over the eons from ancestors with simpler cognitive equipment" (2009, p. 9), and if this is the case then there is a serious problem with how we approach our mostly arbitrary distinctions between the literal and metaphorical in the case of legitimate attributions.

A Pragmatic Approach

Regardless of where we stand on this issue, the intentional stance does work. Does the female robot really want to sing? It does not matter because the intentional stance explains what the robot is doing anyway. The predictive capability of the intentional stance functions just as well for simpler minds as it does for more complex human minds, and it has heuristic advantages as a result, allowing us to investigate the differences between our minds and simpler minds. The intentional stance is a theory neutral way of capturing cognitive competences. We do not need to make detailed hypotheses about the internal workings of a Crow, for example, or how a Crow's brain is different from a Chimpanzee's brain. We can evaluate their strengths and weaknesses independently of how we think their minds might work, and we have done just that, discovering for instance, that when it comes to certain kinds of problem solving, Crows

are able to use causal reasoning in a way not documented in primates (Taylor, Hunt, et al., 2008).

One possible way of interpreting crow behavior is to look at foraging. Presumably, Crows forage to eat, which allows them to continue living. This is an example of goal-directed behavior. How might a crow go about seeking such a goal? Crows might engage in rational foraging, seeking to optimize their potential at achieving their goal (eating). The second method might involve a self-regulation system, which broadly means that the organism under examination would try to optimize "any behavior" (p.181). This briefly illustrates how in the case of foraging, we can make predictions about crow behavior without a complete understanding of the Crow's reasoning process.

Admittedly, this is a simple example, and it probably does not lead to any enlightening conclusions, but using the intentional stance, we can chart the "continuities between simpler animal minds and our own minds," perhaps leading to a discovery of how simple "machines" can lead to complex behavior. There is a problem with this approach, in that "much of the ontogeny and behavior of biological organisms is not intentional" (Blommaert and Janssen, 2005, p. 180), but this implies that there is a scientific way to distinguish between intended and unintended behavior in non-humans, a position that Dennett discards at least to a certain extent.

Criticisms of Dennett's Intentional Stance

Critics of Dennett's intentional stance argue that his attribution of intentional states to artifacts or even other animals is missing something important, such as the phenomenal dimension of consciousness. It initially seems obvious that the Subject needs to be accounted for, either in terms of qualia or some other way of talking about what goes on inside our heads when we appreciate the smell of fresh coffee or lose a loved one. We certainly do not want to dismiss these aspects of our experience or as Fodor (1998) remarks "If, in short, there is a community of computers living in my head, there had also better be somebody who is in charge; and, by God, it had better be me" (p. 207).

Dennett's reply to these criticisms is that if there is to be a science of consciousness, it cannot possibly rely on first person experience. In his words, there "is no such thing as first-person science." (2001, p. 230). This is not to say that Dennett turns a blind eye to the phenomena of consciousness; he just wants to get it out of our heads, so to speak. His method, which he refers to as *heterophenomenology*, is a third-person approach, based on the descriptions that subjects give. It is understandable why this would not be satisfying to his critics, because relying on reports does not sound particularly rigorous, but Dennett disagrees, claiming that it is about as objective as we're going to get about "patterns discernible in the behavior of subjects" (p. 231). It is a method of anchoring subjectivity to something that can be "detected and confirmed in replicable experiments" (p. 231).

The question that must be asked of course is how first-person description can be confirmed and replicated in experiments. It turns out that subjective experiences are increasingly the focus of evaluation in contemporary neuroscience, and the gulf that divides self-reports from physical brains studies is not as large as one might assume. One reason for this might be that neuroscientists do not just rely on subjective statements, such as "That hurts" or "I smell burnt toast." Neuroscientists also pay attention to signs or outwardly observable facts. While it is comforting to know that there are legitimate scientific reasons for taking self-reports seriously, there is still a problem when it comes to interpretation. Dennett's simplification that you simply have to “take the vocal sounds emanating from the subjects’ mouths (and your own mouth) and interpret them” (2003, p. 2) is a little too cheeky for those of us concerned about the theory-ladenness of interpretation and the disconnect between subjective experience and what is actually going on in the brain.

Dennett (2003) thankfully does not consider speech acts and other forms of behavioral reporting to be representative of the real world, and that is fine because scientists should only be interested in the subjective world of one subject in so far as it elucidates other data about "concurrent events in the brains of subjects and in the surrounding environment" (p.2). What exactly would the alternative be? Dennett appears to be taking the first person point of view as seriously as it can be taken while still doing science. The Hard Problem is illusory, and echoing Wittgenstein, Dennett argues that this is an invitation to therapy.

Some scholars claim that a naturalized phenomenology is possible (Pessoa, Thompson, and Noe, 1998), but Dennett sees no evidence that such a method of scientific investigation is viable. But what about the zombies? Zombies are of great concern to some philosophers, and so they should be given at least a brief treatment here. Chalmers and Searle imagine a zombie that "might be able to do everything a conscious person does, passing every test, reporting every effect, without being conscious" (Dennett, 2001, p. 232). Dennett has little patience for such puzzles, arguing that one "must dismiss as a chimera the prospect of a philosopher's zombie, a being that is behaviorally, objectively indistinguishable from a conscious person but not conscious" (p. 232).

One of the traps that philosophers fall into with these zombie thought experiments is that humans can easily be "regarded as sophisticated artifacts, designed as it were, or at least selected over eons, by evolution to exhibit behavior that for perfectly understandable reasons, turns out to be interpretable from an intentional stance" (Beisecker, 2006, p. 44). Any attempt to search for an *élan vital* is nothing more than mysticism, i.e., when “vitalism died, so did content essentialism” (The Dualist, 2002, p.78). Much of this approach digs into the bowels of philosophy and the resistance of many philosophers to abandon some version of dualism. This problem is beyond the scope of this essay, but when we consider the assumptions made by those who are on a quest for primary data we come to an obvious conclusion: we are not authorities about what is happening inside of us, so even if we could access the mysterious consciousness that we identify with the first-person *I* and all of our related

beliefs and desires, would we find anything that cannot be discovered in a third-person account? A common rebuttal, that many experiences are ineffable and cannot be made verbal, returns us to the question just asked, and this is not a problem for heterophenomenologists: "What better grounds could we have for believing that you are unable to describe something than that (1) you don't describe it, and (2) confess that you cannot? Of course you might be lying, but we'll give you the benefit of the doubt" (Dennett, 1991, p. 96–7).

This approach is admittedly agnostic, but once again, the alternative is fuzzy. Velmans (2006) distinguishes between heterophenomenology and his approach, which he refers to as critical phenomenology. He argues that "unlike HP [heterophenomenology], CP [critical phenomenology] does not assume that subjects are necessarily deluded about their experiences or doubt that these experiences can have real qualities that can, in principle, be described" (n.p.). It is not clear what he means by this and he does not present any examples that add clarity to his distinction. He seems to find Dennett's approach less empathetic to the Subject, and perhaps that is what people mean when they talk about a first-person approach, such as when a social-scientist describes personal experiences using thick-description, while still claiming some critical distance (interviews with victims of domestic violence as an example), but this is problematic, especially if taken to mean that the researcher should in some fashion be intimately familiar with the experiences of the subject being studied. While this is no doubt valuable in the case of nurses and doctors experiencing a patient's discomfort or pain for purposes of hospital policy or simply improved bedside manner,

it is not clear how far it should be explored with regard to consciousness research. Should a man not investigate the consciousness of a woman because the experience of being female is not shared between the two of them? Dennett argues that "When guidance from experts is available, one should certainly avail oneself of it, but the claim that one must be an expert (an expert musician, an expert woman, an expert obese person) before conducting the research is an extravagant one" (p.10).

Conclusion

This paper has explored what distinctions there are to be made between folk psychology, the intentional stance, and theory of mind. Obviously there is a great deal of overlap between the three concepts, but the main thrust of what has been argued is that Dennett's intentional stance is a more technical way of approaching the craft of folk psychology, and that it is had heuristic and predictive advantages over other approaches to theory of mind. This paper has also considered some of the criticisms of the intentional stance by way of the theory neutral heterophenomenology, and it concludes that it is the most scientific approach, coupled with other methods, for addressing consciousness.

References

Andrews, K. (2000). Our Understanding of Other Minds: Theory of Mind and the Intentional Stance. *Journal of Consciousness Studies, 7*(7)

Beisecker, D. (2006). Dennett's Overlooked Originality. *Mind Mach, 16*

Blommaert, F., & Janssen, R. (2005). Dennett, Darwin, and Skinner Crows. *Evolutionary Psychology, 3*

Dennett, D. C. (1992). *Consciousness Explained*. Boston: Back Bay Books.

Dennett, D. C. (1998). *Brainchildren: essays on designing minds*. Cambridge, Massachusetts: The MIT Press.

Dennett, D. C. (2001). Are we explaining consciousness yet? *Cognition,* (79)

Dennett, D. C. (2001). *The Fantasy of First-Person Science*. Retrieved December 10, 2009, from http://ase.tufts.edu/cogstud/papers/chalmersdeb3dft.htm

Dennett, D. C. (2003). Who's On First? Heterophenomenology Explained. *Journal of Consciousness Studies, 10*

Dennett, D. C. (2005). *Sweet dreams : philosophical obstacles to a science of consciousness*. Cambridge, Massachusetts: The MIT Press.

Dennett, D. C. (2009). Intentional Systems Theory. In B. McLaughlin, A. Beckermann & S. Walter (Eds.), *Oxford Handbook of the Philosophy of Mind* () OUP.

Fodor, J. (1998). *Review of Steven Pinker's How the mind works, and Henry Plotkin's Evolution in mind*. London Review of Books.

Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behav Brain Sci., 21*(1)

An interview Daniel Dennett(2002). (R. Talbot, D. Yoon Trans.). Tufts University: The Dualist.

Ito, A., & Terada, K. (2007). *The Importance of Human Stance in Reading Machine's Mind (Intention)*. Yanagido, Gifu: Gifu University.

Miguens, S. (2006). D. Dennett's brand of anti-representationalism, *Protosociology, 22*

Morton, A. (1996). Folk Psychology is not a Predictive Device. *Mind, 105*

Pessoa, L., Thompson, E., & NoeÈ, A. (1998). Finding out about filling in: a guide to perceptual completion for visual science and the philosophy of perception. *Behavioral and Brain Sciences, 21*

Ravenscroft, I. (2005). *Philosophy of Mind: A Beginners Guide* OUP.

Sober, E. (2009). Parsimony and Models of Animal Minds. In R. W. Lurz (Ed.), *The philosophy of animal minds* (). Brooklyn College, City University of New York: Cambridge University Press.

Taylor, A. H., Hunt, G. R., Medina, F. S., & Gray, R. D. (2008). Do New Caledonian crows solve physical problems through causal reasoning? *Proc. R. Soc. B, 276*

Velmans, M. (2006). Heterophenomenology versus critical phenomenology. *Phenomenology and the Cognitive Sciences, 6*(1-2)