The Private Language Argument and a Second-Person Approach to Mindreading

Joshua Johnson

Abstract

I argue that if Wittgenstein's Private Language Argument is correct, then both Theory Theory and Simulation Theory are inadequate accounts of how we come to know other minds since both theories assume the reality of a private language. Further, following the work of a number of philosophers and psychologists, I defend a 'Second-Person Approach' to mindreading according to which it is possible for us to be directly aware of at least some of the mental states of others. Because it is not necessary to assume a private language within the Second-Person Approach, I argue that this account of social cognition is superior to Theory-Theory and Simulation Theory since it avoids the objections of the PLA.

INTRODUCTION

Until recently, the debate over how we are able to know the mental states of others has largely been restricted to Theory-Theory (hereafter TT) and Simulation Theory (hereafter ST). Both TT and ST, however, share common assumptions about the nature of how we come to understand mental terms, assumptions that render both theories implausible in light of Wittgenstein's Private Language Argument (hereafter PLA). In this paper I defend the claim that, if the PLA is correct, both TT and ST fail as adequate theories of mindreading.[1] Further, I argue for a 'Second-Person Approach' to mindreading that both

---

[1] The term 'mindreading' as used in the title and throughout this paper is commonly used within the theory of mind literature to denote the process of how one comes to know the mental states of others.

avoids the objections of the PLA and provides a much more intuitive account of how we come to know other minds.

I. THE PRIVATE LANGUAGE ARGUMENT, THEORY-THEORY, AND SIMULATION THEORY

According to Wittgenstein, language is essentially public and a private language is impossible. By a 'private language' Wittgenstein does not mean a language known to only one person (such as the last speaker of an otherwise dead language), nor does he mean a language that one may invent only for oneself (like a cipher for a private journal). Rather, he is describing a language that could, in principle, *only* be known by a single person; particularly, a language whose words refer exclusively to the private mental states of an individual (states such as pain, joy, etc.). The words of such a language would 'refer to what can only be known to the person speaking; to his immediate private sensations. So another person cannot understand the language' (1986: 88e-89e). Of course, in rejecting the possibility of a private language, Wittgenstein is not denying that such internal sensations exist; rather, he is saying that one cannot come to know the meaning of a term such as 'pain' by an act of internal ostension that fixes the meaning of the term by mentally 'pointing' to it.

In Wittgenstein's view, language is similar to a game in that is it rule-governed. In order for one's behavior to be meaningful within the context of a particular game, one must be open to correction by following certain rules that are publicly accepted.[2] For example, an individual's behavior in a chess game is meaningful insofar as it conforms to the rules of chess. If the first move of the game was for a player to throw her knight at the board in order to bowl over her opponent's pieces, then such an action would be meaningless -- within the context of chess -- to those playing (or observing) the game. Similarly, if a person attempts to use a word, say 'cup', without recourse to the rules for how this term functions within a particular language, then such a usage would be meaningless. Language requires rules to justify the proper use of words, and the rules themselves are public in nature. Thus, a person cannot come to understand the meaning of a word like 'pain' by merely labeling an internal sensation with a particular sign. If such were the case, then there would be no real sense in which one could correctly follow the rules for using this sign -- since following the rule would simply amount to *appearing* to oneself to be following a rule (1986: 81). That is, a term used in such a way would not be a candidate for public correction, and hence, would not be in accordance with the rule-governed nature of language.

Wittgenstein provides an interesting thought experiment involving beetles and boxes to help further illuminate his concerns with a private language:

> Now someone tells me that he knows what pain is only from his own case! -- Suppose everyone had a box with something in it: we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle. – Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. – But suppose the word "beetle" had a use in these people's language? -- If so it would not be used as the name of a thing. The thing in the

---

[2] 'To obey a rule, to make a report, to give an order, to play a game of chess, are customs (uses, institutions).

To understand a sentence means to understand a language. To understand a language means to be master of a technique.' (1986:81)

box has no place in the language-game at all; not even as a something: for the box might even be empty. -- No, one can 'divide through' by the thing in the box; it cancels out, whatever it is.

That is to say: if we construe the grammar of the expression of sensation on the model of 'object and designation' the object drops out of consideration as irrelevant. (1986: 100^e)

The analogy between the words 'beetle' and 'pain' should be clear: if our word 'pain' refers to something that is *essentially* private, then there would be no way for one individual to truly understand what another individual means by the term 'pain'. To quote Wittgenstein again,

If one has to imagine someone else's pain on the model of one's own, this is none too easy a thing to do: for I have to imagine pain which I do not feel on the model of the pain which I do feel. (1986: 101*)

Of course, Wittgenstein is not saying that internal experiences do not exist, nor is he necessarily stating that such sensations play no part in our mental language; rather he is claiming that if the meaning of certain words is completely derived from an essentially *private* experience, then the use of those words would be absurd within a public context.[3]

A number of psychologists and philosophers have argued that the PLA presents a significant problem for both TT and ST (e.g. Montgomery 1997, Carpendale and Lewis 2004, Hobson 2009, Racine 2004, and Reddy 2008), and I shall follow their interpretation of the PLA with respect to these theories. In short, both TT and ST make the common assumption that the mental states of others are private, unobservable entities that must be accessed either through theoretical inference or personal introspection (Carpendale and Lewis 2004: 83), and it is this assumption that requires that both theories assume the reality of a private language.

A common formulation of TT portrays the mindreader as a scientist who, based upon her observations of human behavior and accumulation of evidence, postulates a set of psychological laws by which she infers the mental states of other organisms based upon their behavior (e.g. Gopnik and Meltzoff 1997). Another version of TT -- the 'modular' approach -- says that humans come naturally equipped with various cognitive mechanisms that, when fully developed, enable them to appeal to an internalized folk psychology in order to ascribe mental states to others (e.g. Baron-Cohen 1995). Importantly, in both variations of TT, the person is making *inferences* about the mental states of other organisms by recourse to a theory of mind. Given that TT operates under the assumption that an individual's mental experiences are private unobservable entities, TT must also assume that the mindreader's *own* mental experiences are private (Carpendale and Lewis 2004: 83). Hence, her inferences about the states of others must

---

[3] Soren Overgaard offers an interesting interpretation of Wittgenstein's 'beetle in the box' scenario: 'The last portion strongly suggests that Wittgenstein's argument is really a *reductio*; he is trying to show that a particular assumption has absurd consequences, and the point is, on the basis of its absurd consequences, to reject that assumption. The conclusion, then, is not that pain-sensations are irrelevant to our attributions of pain to each other. Rather, since Wittgenstein takes the latter to be an absurd consequence, he can reject the assumption from which it follows. Wittgenstein is saying something like the following: If we construe sensation talk in a certain way, then the absurd consequence follows that the sensations themselves are completely irrelevant. Since they cannot be irrelevant – indeed what could be more relevant to our attributions of pain to each other than the actual pains of actual people? – we should avoid construing sensation talk in that way.' (2005: 253)

be at least partially built upon terms whose meanings are derived from a private language referencing her own internal experiences.

According to ST, people engage in mindreading by using their own cognitive equipment to run internal simulations -- thus enabling the individual to place herself in the perspective of another person via pretence (Goldman 2006). The experiences in question are 'offline' in that they are mere simulations rather than the actual experiences of the individual performing the simulation. Further, while running her simulation, the individual must quarantine her own mental states that are not simultaneously held by the target of mindreading. Upon completion of the simulation, the individual generalizes her own simulated experience to another. Note that the individual comes to know the mental states of others by *introspection* of her own mental states. Given the structure of ST, a person comes to understand the meaning of mental terms by recourse to her own subjective experiences (Montgomery 1997: 296) -- hence, ST must assume the reality of a private language.[4] Peter Hobson reaches a similar conclusion when he says that

> [It] is commonplace for contemporary developmental psychologists to espouse the view that we need to infer the nature of other people's minds, in some cases on the basis of our first-person experience of our own minds. […] Wittgenstein's attack on the very concept of a private language undermines the assumption that all by oneself and without the possibility of correction by others (already experienced as others), one would be able to identify a given mental state as the same when this recurs within one's own experience, and then go on to ascribe it to other people (2009: 84).

While there is considerable difference in how TT and ST explain the phenomenon of mindreading, it appears that *both* theories require the existence of a private language. However, if the conclusion of the PLA is correct, then neither theory as described here can provide us with a plausible account. Given the problems associated with the possibility of a private language, it appears that both TT and ST are in trouble.[5] A radically different account of mindreading is needed to provide an alternative approach that avoids the objections of the PLA. One such account which I shall refer to as the 'Second-Person Approach' (hereafter SPA) says that we have direct awareness of at least some of the mental states of others.

---

[4] This objection primarily applies to 'explicit' forms of ST (e.g. Goldman 1998) which hold that introspection is done consciously and for the purpose of mindreading. 'Implicit' forms of ST (e.g. Gallese and Goldman, 1998; Goldman, 2006), in which introspection and ascription occur at a subconscious level may not be as vulnerable to the PLA. However, there is good reason to think that implicit ST is not a version of ST at all (Gallagher 2007). I address this issue in greater detail below.

[5] This is not to say that we *never* engage in the practice of theorizing or simulation with respect to others. Indeed, there are, no doubt, many occasions when we must employ theory attribution or pretence in order to understand the behavior of another. However, if the PLA is correct, these could not be the *fundamental* way that we mindread others both developmentally and in our normal daily experience. Shaun Gallagher expresses similar sentiments when he claims that before 'we are in a position to form a theory about or to simulate what the other person believes or desires, we already have specific pre-theoretical knowledge about how people behave in particular contexts. We are able to get this kind of knowledge precisely through the various capabilities that characterize primary intersubjectivity [similar to the Second Person Approach that I am defending] including, imitation, intentionality detection, eye-tracking, the perception of intentional or goal-related movements, and the perception of meaning and emotion in movement and posture.' (2001: 90)

II. A SECOND-PERSON APPROACH TO MINDREADING

SPA takes it names from its unique approach to mindreading. Whereas TT endorses the importance of a third-person perspective in mindreading in theory attribution, and ST emphasizes the first-person perspective in introspection, SPA holds that what matters most in knowing the minds of others is the second-person experience of another person. Vasudevi Reddy has summed up the distinction nicely:

> In the first- and third-person approaches to knowing other minds, both retaining the premises of the [privacy] gap, other persons are "known" either by extension of the experiences of the self or from the outside through observation, inference, and theory. [The second-person] approach suggests that others are experienced as *others* in direct emotional engagement, and that this fundamentally undermines the "problem" in the "problem of other minds". (2008: 26)

It will help to begin by examining the *kind* of knowledge with which SPA is concerned. Eleonore Stump argues that the knowledge one gains from second-person experiences is not propositional or 'knowledge that'. Instead, Stump holds that knowledge of persons (and their mental states) is a form of 'knowledge by acquaintance' (2010: 51-53). While both TT and ST agree that an individual gains propositional knowledge about another person -- that is, one knows *that* it is the case that the target is angry, happy, or worried, SPA says that we have non-propositional knowledge directly *of* another person's mental states -- that is to say, one may simply *know* another's anger, happiness, or worry.

Stump provides a clever variation of Frank Jackson's famous 'Mary' thought experiment (Jackson 1986: 291-295) in order to reinforce her position.[6] The story is altered slightly so that that Mary is a super-psychologist who has come to know all the propositional knowledge that there is to know about the mental states of others, although she has been raised in isolation from any second-person experiences.[7] Upon being released from her solitary existence, Mary is introduced to her mother and, for the first time, experiences her mother from a second-person perspective. According to Stump, there appears to be no doubt that Mary will come to have new, non-propositional knowledge and that she

> will know things she did not know before, even if she knew everything about her mother that could be made available to her in non-narrative propositional form, including her mother's psychological states. Although Mary knew that her mother loved her before she met her, Mary will learn what it is like to be loved. (2010: 52).

---

[6] For the original thought experiment see 'What Mary Didn't Know', *The Journal of Philosophy*, Vol. 83, No. 5 (1986), pp. 291-95. Those familiar with the thought experiment will recall that Jackson used it to argue that there are facts that can be known that are non-physical. Mary, a super-scientist raised in a grayscale room, comes to know all there is to know about the science of color vision and, thus, knows all physical facts concerning the topic. One day she is released from her drab domicile into the world of color. Upon observing a ripe tomato, Mary learns something new; namely, what *it is like* to see red. Hence, there are facts that can be known that are non-physical as well as non-propositional (the qualitative 'what-it-is-like' knowledge of phenomenal experience).

[7] Of course, Stump acknowledges that this scenario is merely a thought-experiment and that it could not occur in the actual world since there would be severely debilitating psychological consequences for an individual raised in such extreme isolation from personal contact with other humans.

Stump's conclusion appears to be in agreement with much of what Wittgenstein has to say about how we know other minds. For example,

> "We *see* emotion." -- As opposed to what? -- We do not see facial contortions and *make the inference* that he is feeling joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give any other description of the features. -- Grief, one would like to say, is personified in the face. This is essential to what we call "emotion" (1980b: 100e).

Here, Wittgenstein appears to make the distinction between a third-person, propositional description *that* a person is grieved (based upon the inference from how her face is construed), and what appears to be a second-person, non-propositional experience of simply *seeing* a person's grief. Elsewhere Wittgenstein writes:

> "I see that the child wants to touch the dog, but doesn't dare." How can I see that? -- Is this description of what is seen on the same level as a description of moving shapes and colours? Is an interpretation in question? Well, remember that you may also mimic a human being who would like to touch something, but doesn't dare (1080a 186e)

Again, Wittgenstein appears to be distinguishing between a propositional description of a person's affective state based upon inference and an immediate awareness of this state. Another point of interest in the above passage is that Wittgenstein appeals to the importance of *mimicry* as a means of communicating the affective state of another. Why is this significant? In reference to the above passage, Hobson has pointed out that that if one simply tried to provide a physical description of the child's affective state *without* mimicry, 'then one would fail to understand something important about what is expressed'. That is, when attempting to describe the affective states of an individual to a third person, we often use mimicry because there is non-propositional information communicated by doing so that could not be communicated by a mere propositional description (2009: 248).[8]

Recent work in neuroscience on the mirror neurons of both monkeys and humans appears to support the above notion with respect to mimicry. Speaking of the mirror neuron system, Marco Iacaboni states that the 'functional properties of these neurons suggest that they may implement a *simple, noninferential mechanism* of action recognition based on neural identity. This mechanism may be a building block for imitative behavior' (2005). Iacaboni's description of the knowledge provided by the mirror neuron system sounds very much like non-propositional knowledge. As research into the mirror neuron system has progressed, the notion that our knowledge of at least some of the mental states of others is non-propositional and direct appears to be increasingly verified. With this point in mind, I shall now proceed to show why, unlike both TT and ST, the second-person approach to mindreading is not rendered implausible by the PLA.

---

[8] The notion of mimicry employed here should not be confused with the sort of 'offline' simulation described by ST. The latter is used in order to understand the mental state of a target of mindreading, while the former is used to express to a third person the mental state of another.

III. THE SECOND-PERSON APPROACH AND THE DIRECT AWARENESS OF MENTAL STATES

Recall that Stump's thought experiment is designed to show that the knowledge gained through second-person experience is non-propositional. Another important aspect of this thought experiment is that Mary appears to be *directly aware* of her mother's emotional state. That is, Mary apparently perceives her mother's affective state without recourse to inference or introspection. The fact that humans have the capacity to directly perceive the emotions of others appears to be supported by a test, constructed by Hobson, Moore, and Lee, that is designed to examine the proficiency of autistic children at detecting emotion in the behavior of others (2004: 52-58). Since a common distinguishing characteristic of autistic individuals is the inability to mindread others, the test is designed to determine if autistic children are able to detect emotion from stimuli that present no facial expressions (or recognizable human bodies for that matter). The team attached reflective lights to the torso and limbs of a person and then filmed the individual in a darkened room so that all that could be seen were the reflective patches of light as they matched the movements of the person's body. Hobson observes that when watching the moving light display,

> What you see *through* the moving dots is a person. There is no doubt about it –
> you are watching a person doing things. […] It is as if the displays home in on a
> brain mechanism that detects people. No need to think, no need to go through a
> conscious process of judgment – one simply sees a person (2004: 53-54).

But the test shows that this is not *all* that one sees in such displays. The team instructed the individual modeling the reflective dots to engage in various emotional behaviors (surprise, sadness, fear, anger, and happiness). After questioning the children as to what was happening in the light display, it turns out that non-autistic children overwhelmingly tend to report emotion in the movement of the lights; while those impaired by autism typically fail to report emotion in the displays (2004: 56). For our purposes, what is so interesting about this case is that the non-autistic children involved in the test appear to simply '*see*' the emotions expressed through the behavior of the moving lights. This constitutes a strong *phenomenological* argument for SPA. In most cases, when we apprehend the emotions or intentional actions of others, we are not conscious of appealing to a theory of mind or running an internal simulation. Like the non-autistic children involved in the point-light test, and as Wittgenstein emphasizes, we appear simply to *perceive* the affective and intentional states of others.

Here I should point out that an advocate of ST who endorses the notion of 'low level' or 'implicit' simulation may object that recognition of emotions in facial expressions and behavior involves simulation at a non-conscious and sub-personal level (e.g. Gallese and Goldman 1998; Goldman 2006); hence, one cannot appeal to one's own phenomenal experience in order to deny that simulation has occurred. However, I am inclined to agree with Shaun Gallagher who makes a strong case for the idea that implicit ST should not properly be thought of as a version of ST at all since it does not involve actual simulation (2007). According to Gallagher, simulation requires both instrumentality and pretence so that the mindreader makes 'use of a first-person model to form third-person "as if" or "pretend" mental states.' (360). Implicit simulation is said to occur at a pre-conscious level, but neither instrumentality nor pretence are present at this level since the person has no 'instrumental access to neuronal activation' and the sub-personal brain does not use its neurons to initiate 'pretend states'. Indeed, Gallagher claims that what is referred to as 'implicit ST' is actually something very much like *perception* of another person's mental states:

> The perception of the other person's action automatically activates in our brain the same areas that are activated when we engage in similar action. The other *person has an effect on us*. The other *elicits* this activation. This is not a simulation, but a perceptual elicitation. It is not us (or our brain) *doing* it, but the other who does this to us (2007: 360-61).

Recall that TT and ST are both vulnerable to the criticism contained in the PLA because they assume a private language and treat the mental experiences of others as hidden things that the individual knows through inference or personal introspection. However, this is precisely what SPA denies. There is no need for the assumption of a private language within SPA since the knowledge that we gain about the mental states of others through second-person experiences is both non-propositional and direct; one need not come to understand the meaning of mental terms exclusively by recourse to one's own experiences. SPA, therefore, rather than being a target of the PLA, actually *supports* the underlying notion behind it: the mental states of others are not hidden but are directly accessible to us within the second-person experience.

CONCLUSION

I have argued that if one considers the PLA to be a persuasive line of reasoning, then one must agree that TT and ST are flawed accounts of mindreading. Since SPA holds that knowledge of persons and their mental states is non-propositional and direct, it avoids the objections contained within the PLA. SPA is also superior to TT and ST because it provides what appears to be a much more intuitive account of how we come to know persons and what they are thinking. In my second-person experience of you, I do not simply know about you, or claim to know by inference or introspection what it is like to be you. Instead, much like Augustine's declaration to God recounted in his famous *Confessions*, I know *you*.[9]

Bibliography

Carpendale, Jeremy and Lewis, Charlie. 2004. "Constructing an understanding of the mind: the development of children's social understanding within social interaction", *Behavioral and Brain Sciences,* 27:1

Gallagher, Shaun. 2001. "The Practice of Mind", *Journal of Consciousness Studies*, 8, No, 5-7: 83-108

Gallagher, Shaun. 2007. "Simulation Trouble", *Social Neuroscience*, 2 (3-4): 353-365

Gallese,Victor & Goldman, Alvin. 1998. "Mirror neurons and the simulation theory of mind-reading", *Trends in Cognitive Sciences*, 2, 493-501

Goldman, Alvin. 1998. "Interpretation psychologized", *Mind and Language*, 4, 161-185

Goldman, Alvin. 2006. *Simulating Minds*, (Oxford: Oxford University Press)

Hobson, Peter. 2004. *The Cradle of Thought*, (Oxford: Oxford University Press)

Hobson, Peter. 2009. "Wittgenstein and the developmental psychopathology of autism", *New Ideas in Psychology*, 27

Iacaboni, Marco. 2005. "Understanding Others: Imitation, Language, and Empathy", in *Perspectives on Imitation, from Neuroscience to Social Science*, ed. by Susan Hurley and Nick Chater (Cambridge: MIT Press)

Jackson, Frank. 1986. *The Journal of Philosophy*, 83, No. 5: 291-95

Krueger, Joel and Overgaard, Soren. 2012. "Seeing subjectivity: Defending a perceptual account of other minds", *ProtoSociology: Consciousness and Subjectivity*, 47: 239-262

Overgaard, Soren. 2005. "Rethinking Other Minds: Wittgenstein and Levinas on Expression", *Inquiry* 48:3

Stump, Eleonore. 2010. *Wandering in Darkness*, (Oxford: Oxford University Press)

Racine, Timothy. 2004. "Wittgenstein's Internalistic Logic and Children's Theories of Mind", in *Social Interaction and the Development of Knowledge*, ed. by Jeremy Carpendale and Ulrich Muller (Mahwah: Lawrence Erlbaum Associates)

Reddy, Vasudevi. 2008. *How Infants Know Minds*, (Cambridge: Harvard University Press)

Wittgenstein, Ludwig. 1986. *Philosophical Investigations*, trans. by G. E. M. Anscombe (Oxford: Blackwell)

Wittgenstein, Ludwig. 1980. *Remarks on the Philosophy of Psychology, Vol. I,* trans. by G. E. M. Anscombe (Oxford: Basil Blackwell)