My Mind . . . or Your Mind

Adam Brandenburger*

Version 12/28/15

> "When I am getting ready to reason with a man I spend one-third of my time thinking about myself and what I am going to say, and two-thirds thinking about him and what he is going to say."
> – Abraham Lincoln

1 Introduction

It is a commonplace that good strategists try to put themselves in the shoes — better, the heads — of other players in order to predict what those other players will do. In this note we will look at what the cognitive sciences have to say about this process, and at some hints on how to improve at it.

2 Theory of Mind and Empathy

In cognitive science, two terms, **Theory of Mind** (**ToM**) and **empathy**, are used to refer to the process of connecting to other people's minds. By ToM is meant "the capacity to infer and represent another person's intentions, desires, or beliefs."[1] By empathy is meant "the ability to share the feelings of others."[2] The term "Theory of Mind" was coined in a famous 1978 paper written by primatologists David Premack and Guy Woodruff,[3] which investigated whether chimpanzees possess this capability. Since then, ToM capability and its development in infants has been studied both in humans and in a number of other species.

It is important not to confuse ToM and empathy. As the definitions we gave above indicate, the first is a cognitive process and the second is an affective process. Furthermore, the two processes appear to use different neural circuits.[4] In the remainder of this note, we will focus on the cognitive process of ToM.

3 Neuroscience of ToM

A common exercise in ToM studies is to present subjects with stories (verbal or visual) involving social situations with several characters and then to test the ToM ability of subjects by asking them to answer questions about what they think a character is thinking, and the like.[5] Investigation using **neuroimaging technology** of subjects engaged in such exercises has identified several regions of the human brain active in ToM processing:[6]

*Stern School of Business, Tandon School of Engineering, NYU Shanghai, New York University, New York, NY 10012, U.S.A., adam.brandenburger@stern.nyu.edu, adambrandenburger.com.

- temporal poles (TP)
- posterior superior temporal sulcus (STS)
- medial prefrontal cortex (mPFC)

The TP and the STS appear to involve neural processes that support ToM, but stop somewhat short, while the mPFC is where the actual representation of mental states occurs.[7] The TP are activated in exercises involving language, such as comparing sentences with random word strings, and in exercises involving memory retrieval. These regions may therefore build up what are sometimes called "scripts"[8] (somewhat like those for a play or a film) that facilitate predicting how others will behave in the future. The STS is activated on reception of inputs (motion, sound, light) about the behavior of other living agents. Underlying knowledge about the behavior of living agents, together with inputs from current behavior, may help predict subsequent behavior.

The mPFC is thought to be the locus of actual neural representation of mental states (intentions, desires, knowledge, and beliefs). There is some evidence that, at the neurological level, there is an anchoring and adjustment process that begins with one's own mental states and then operates to try to infer the mental states of others.[9] The mPFC is directly connected to the TP and the STS.[10]

4 ToM in Games

Even if subjects in experiments are not asked to engage in ToM reasoning, they may do so on their own accord. A very important finding of this kind is that when subjects play a game against another person, neuroimaging shows that the ToM regions (as above) are also activated in this setting.[11] This tells us that game-playing humans, at least in a controlled experimental setting, do indeed engage in the kind of "I think you think ..." reasoning that is the subject matter of game theory. In fact, more than just corroborating game theory, this finding tells us how game theory, as a member of the cognitive sciences, should develop.

Here is the definition of **Nash equilibrium**:[12]

> In game theory, the Nash equilibrium is a solution concept of a non-cooperative game[13] involving two or more players, in which each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing only their own strategy.

We see that the concept of Nash equilibrium 'short-circuits' the ToM process, by assuming that each player knows the strategies followed by the other players, and does not need to try to guess the mental states (intentions, desires, knowledge, and beliefs) of the other players and thereby infer what their strategies might be. A more recent approach, called **epistemic game theory**,[14] captures thinking about thinking ... directly. In this way, game theory becomes closer both to everyday intuitions about game playing and to what neuroscience tells us about game playing.

Using techniques from epistemic game theory, game theorist Terri Kneeland has looked experimentally at the question of how many levels of thinking about thinking ... players undertake.[15] Some rough definitions:[16]

- call a player **level-0** if the player chooses a strategy at random
- call a player **level-1** if the player chooses a strategy that is optimal when other players are level-0

- call a player **level-2** if the player chooses a strategy that is optimal when other players are level-1
- call a player **level-3** if the player chooses a strategy that is optimal when other players are level-2
- ...

Figure 1 summarizes Kneeland's findings,[17] which are that 7 percent of her subjects were level-0 players, 22 percent were level-1 players, 27 percent were level-2 players, 22 percent were level-3 players, and 22 percent were level-4 players. These numbers agree broadly with human cognitive limits found elsewhere.[18]

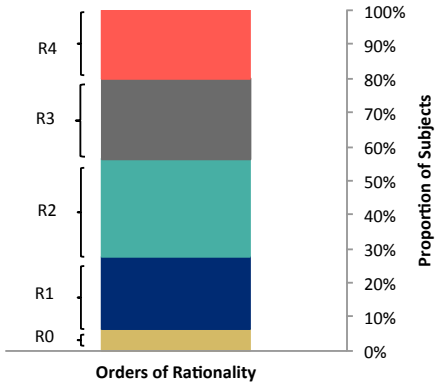

Figure 1: From Kneeland (2015)

There is an interesting historical note. Daniel Ellsberg, famous for releasing the Pentagon Papers, a secret U.S. government study of American involvement in Vietnam, to the press in 1972,[19] is also a famous decision theorist.[20] In a 1959 paper, Ellsberg expressed concern about what game theory might become as a field, if it did not include properly the kind of "I think you think ..." reasoning we have been discussing:[21]

> "These particular uncertainties — as to the other player's beliefs about oneself — are almost universal, and it would constrict the application of a game theory fatally to rule them out."

Happily, game theory today is moving in the direction of analyzing not ruling out the uncertainties in a game.

5 ToM and Language

People engage in ToM reasoning when explicitly asked to do so (Section 3) and when placed in explicitly strategic situations (Section 4). But we also engage in ToM reasoning in everyday activities, more or less automatically, although this seems likely to be to fewer levels than when we are prompted to do so. One very important such activity is communication via language.

The term **syntax** refers to the rules according to which the structures, such as phrases and sentences, are built in a language.[22] The term **semantics** refers to the meaning of words, phrases, sentences, and the like. But even if someone knows the syntax and semantics of a language, it seems there is no guarantee that the person will correctly interpret correctly everything he hears. According to what is known as **intention-based semantics**, understanding correctly what one hears depends on understanding what mental state the speaker intends to bring about in one's mind.[23] Under this view, the effectiveness of language as a communication device depends critically on possession of ToM ability by speakers and hearers.

Here is an example.[24] Suppose someone hears the sentence: "Bob is well read. He's even heard of Shakespeare." To interpret this sentence correctly, the hearer needs to understand that the speaker intends the hearer to conclude that Bob is not well read. The intended semantics are precisely opposite to the 'literal' semantics, and, the argument goes, the hearer needs to use ToM to arrive at the intended meaning.

Concerning the evolution of ToM ability, some theorists argue for a chain of causation that runs from the emergence of large cooperative groups in early humans, to the emergence of language to enable effective coordination within these groups, to the emergence of ToM ability to enable effective communication via language.[25] Relevant (though not conclusive) to this line of argument is a positive correlation in primates between size of social group and (scaled) size of the neocortex region of the brain.[26] The Social Brain Hypothesis, as this line of argument is called, is striking in that it goes against a common view that our large brains evolved in relation to challenges in the physical world (tool making and the like) and, instead, points to challenges in the social world (communication).

6 ToM and Fiction

There is some evidence that certain types of activity lead to improvement on ToM tasks. A recent study found that the bigger the fiction-reading habits of a subject, the better that subject performed on tasks testing empathic and ToM abilities, and argued (based on statistical analysis) that the causation runs from reading to improved social abilities, and not vice versa.[27] It is interesting to note that this finding refers specifically to fiction-reading habits. There does not appear to be a corresponding effect from reading non-fiction.[28]

7 Exercises

With no pretense at full scientific rigor, let us take and apply what we learned in Section 3 of this note. We start with the (tentative) understanding that there are three areas of the brain associated with ToM processing. We then convert this understanding into a three-part guide to use when consciously performing strategic reasoning, as follows. Part a: Collect historical information about the particular context being studied. (For example, in a negotiation, the context might be the prior negotiation styles of the various players.) Part b: Collect data, based on various inputs, about the current behavior of the other players. Part c: Predict the subsequent behavior of other players, by asking how one might oneself subsequently act in their positions and then adjusting these predictions to take into account how one thinks their mental states differs from one's own. (Remember that "mental state" refers to intentions, desires, knowledge, and beliefs.)

As a shorthand in these exercises, let us call someone who uses (respectively, does not use) this three-part guide a good (respectively, bad) strategist.

- Find a current or historical situation (in any domain: business, government, national, ...) and

identify a player involved whom you think is or was a good strategist as just defined. Justify your view.

- Find a current or historical situation (in any domain: business, government, national, ...) and identify a player involved whom you think is or was a bad strategist as just defined. Justify your view.

The next two exercises relate to Section 6 of this note.

- Find a work of fiction (book, movie, painting, ...) which you argue requires the audience to engage to a considerable degree in ToM activity.

- Find a work of fiction (book, movie, painting, ...) which you argue requires the audience to engage only to a small degree in ToM activity.

Notes

[1]Singer, T., and A. Tusche, "Understanding Others: Brain Mechanisms of Theory of Mind and Empathy," in Glimcher, P., and E. Fehr, *Neuroeconomics: Decision Making and the Brain*, 2nd ed, Academic Press, 2014, p.514.

[2]Loc. cit..

[3]Premack, D., and G. Woodruff, "Does the Chimpanzee Have a Theory of Mind?" *Behavioral and Brain Science*, 1, 1978, 515-526.

[4]Singer and Tusche, op. cit, p.517.

[5]For an early example of this type of exercise, see Kinderman, P., R. Dunbar, and R. Bentall, "Theory-of-Mind Deficits and Causal Attributions," *British Journal of Psychology*, 89, 1998, 191-204. A recent study is Stiller, J., and R. Dunbar, "Perspective-Taking and Memory Capacity Predict Social Network Size," *Social Networks*, 29, 2007, 93-104.

[6]Singer and Tusche, op.cit., pp.518-519.

[7]See Frith, U., and C. Frith, "Development and Neurophysiology of Mentalizing," *Phil. Trans. R. Soc. Lond.* B, 358, 2003, 459-473, from which the discussion of this paragraph comes.

[8]Schank, R., and R. Abelson, *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*, Hillsdale, 1977.

[9]Singer and Tusche, loc.cit., and Tamir, D., and J. Mitchell, "Neural Correlates of Anchoring-and-Adjustment During Mentalizing," *Proc. Natl. Acad. Sci. U.S.A.*, 107, 2010, 10827-10832.

[10]Frith and Frith, op.cit., p.466.

[11]See, inter alia, McCabe, K., D. Houser, L. Ryan, V. Smith and T. Trouard, "A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange." *PNAS*, 98, 2001,11832-11835; Gallagher, H., A. Jack, A. Roepstorff and C. Frith, "Imaging the Intentional Stance in a Competitive Game," *NeuroImage*, 16, 2002, 814-821; Rilling, J., A. Sanfey, J. Aronson, L. Nystrom and J. Cohen, "The Neural Correlates of Theory of Mind Within Interpersonal Interactions," *NeuroImage*, 22, 2004, 1694-1703.

[12]http://en.wikipedia.org/wiki/Nash_equilibrium.

[13]For the distinction between non-cooperative and cooperative game theory, see http://adambrandenburger.com/wp/wp-content/uploads/2015/01/cor-01-01-15.pdf.

[14]"Epistemic" means "Of or relating to knowledge or cognition" (http://en.wiktionary.org). For an introduction to the field, see Brandenburger, A., *The Language of Game Theory: Putting Epistemics into the Mathematics of Games*, World Scientific, 2014.

[15]Kneeland, T., "Identifying Higher-Order Rationality," *Econometrica*, 83, 2015, 2065-2079.

[16]Following the pioneering definitions in Nagel, R., "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85, 1995, 1313-1326. There are some very significant challenges in defining and measuring levels of thinking; see Kneeland, op.cit. for a full treatment.

[17]Kneeland, op.cit.. Diagram included with permission.

[18]Kinderman, Dunbar, and Bentall, op.cit., Stiller and Dunbar, op.cit..

[19]See http://en.wikipedia.org/wiki/Pentagon_Papers.

[20]His most famous paper is Ellsberg, D., "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 1961, 643-669.

[21]Ellsberg, D., "Rejoinder," *Review of Economics and Statistics*, 16, 1959, 42-43.

[22]The material in this and the next paragraph comes from Frith and Frith, op.cit., p.469.

[23]Grice, P., "Meaning," *Philosophical Review*, 66, 1957, 377-388.

[24]Frith and Frith, loc.cit..

[25]Dunbar, R., "The Social Brain Hypothesis," *Evolutionary Anthropology*, 6, 1998, 178-190.

[26]Dunbar, R., "Neocortex Size as a Constraint on Group Size in Primates," *Journal of Human Evolution*, 20, 1992, 469-493; also Dunbar, R., *Grooming, Gossip, and the Evolution of Language*, Harvard University Press, 1996.

[27]Mar, R., K. Oatley, and J. Peterson, "Exploring the Link between Fiction and Empathy: Ruling Out Individual Differences and Examining Outcomes," *Communications: The European Journal of Communication Research*, 34, 2009, 407-428; Oatley, K., "In the Minds of Others," *Scientific American Mind*, November/December 2011, 63-67.

[28]Oatley, K., op.cit..