Where is the Self in Connectionism?

QUERY SHEET

**Q1:** Au: Please provide the complete detail.
**Q2:** Au: Please provide the page, edition, and volume of the reference.
Where is the Self in Connectionism?

Frank Van Overwalle

Vrije Universiteit Brussel, Belgium

Kashima and colleagues (2007) made a bold attempt to model the self in a connectionist framework. Following the footsteps of Mead (1934) and James (1890), they argued that a significant proportion of the dynamic self can be viewed as a narrative knowledge structure unfolding in the stream of consciousness, and that this knowledge emerges form on-going social interactions with others and imitating the behaviors of others. Although interesting, I was not very thrilled by their approach. Probably most, if not all psychologists would agree that the self develops not in social isolation but rather in interaction with others. However, the manner in which some aspects of the self were brought to the foreground and applied in a connectionist framework seems debatable. Perhaps most importantly, their connectionist I-SELF model is not convincing in its attempt to model the self as a unique or special functional entity. In my opinion, it misses the point that the self is very motivational, whose cognitive function is to preserve and enhance self’s well-being even at the expense of others.

What is (Special about) the Self?

Most humans intuitively grasp the idea that knowledge about the self is special, and different from other people. This does not implies that the self occupies a different place or structure in the brain, but rather that there are functional differences in the manner we perceive, represent and cognize ourselves in comparison with other persons, even if they are familiar to us. Information about other social agents is initially filtered and processed in reference of the self. From an evolutionary perspective, using the self as an informational source and point of reference for social interaction, inference and prediction brings important advantages, not in the least that it prevents taking actions that harm the self. There is evidence that self-judgments generate distinct neuron-anatomical processes (Kelley et al., 2002). Contrary to Mead’s view of the self as an imitation of other’s behaviors or a (direct) reflection of people’s appraisals about the self, research documented that our self-perceptions are more similar to our perceptions of how others see us than how others actually see us (Krueger, 2003). Direct judgments about the self activate neurological structures involved in the first-person perspective which are different from those involved in one’s thought about how other people judge the self, or a third-person perspective (Ochsner et al., 2005). Thus, direct self-appraisals draw primarily on self-knowledge abstracted from previous personal experiences and life events, rather than on beliefs how others see us. Moreover, because of the greater familiarity with behaviors, thoughts, emotions and goals of oneself, self-knowledge is more elaborated than other social knowledge, and research attests that the self is a robust and habitual reference point, more so than other familiar others (Sedikides, 2003). Because of this central place, it has also been suggested that the self is more coherently structured and integrated than our perception of others (Novak, Vallacher, Tesser & Borkowski, 2000).

Contrary to Kashima et al.’s claim that I-SELF is a model about the self, none of their simulations captures this special status of the self. In practically all their simulations, one could exchange the node representing “I” by “you” or “s/he”, and nothing essentially would change in the results. Thus, the conclusions they draw from their model seem equally valid for any single individual, and do not seem to reflect anything particular about the self. Nevertheless, I suggest that their model can encompass a great deal of unique aspects of the self, mainly with respect to the cognitive differences between the self and other people such as its greater complexity, which comes from the greater frequency of and familiarity with behaviors and experiences of oneself than other people. Providing more information about the self than about other people would probably generate a more differentiated knowledge base (e.g., hidden layer). This could explain the greater complexity and memorability of self-descriptions, and why self-schemas have a guiding and organizational function in processing information about others.

However, not only are different cognitive processes involved in the self, but also its motivational role is radically different. The self is egocentric and ego-protective (Sedikides, 2003; Krueger, 2003). The self cares much about itself, and attempts to preserve a positive image. Self-knowledge is disproportionately positive (Gaertner, Sedikides, Vevea, & Iuzzini, 2002). This is a crucial aspect of the self that is universal and permeates all cultures. The need for a positive self-regard can also be traced back to evolutionary advantages. The subjective experience of self-esteem influences one’s standing in the social hierarchy,
was essential in reproductive success in an ancestral environment. High self-esteem is a universal defense mechanism protecting against terror, depression and life problems (Sedikides & Koole, 2004; Sedikides, Gaertner & Toguchi, 2003). This protective motive results in a host of biases in the perception of the self, that go by terms such as self-enhancement bias (seeing the self as superior to others), self-reference effect (superior memory for self descriptive judgments), self-defense memory (lower recall of negative behaviors by the self), hedonic or self-serving attribution bias (making attributions for success and failure that maintain one’s positive self view), false consensus (believing that others are similar to the self), self-anchoring (viewing one’s ingroup as similar to the self), and so on (Sedikides, 2003; Sedikides & Green, 2000, 2004).

Totally unclear is how the I-SELF model could capture the self-protective motivational nature of the self. Incorporating these self enhancing and protecting biases would require a basic change in design. Not only a new set of simulations, but also a different connectionist model altogether. To build such a model that would naturally fall out from well-known connectionist principles is not an easy endeavor, and this, I am convinced, would result in a major contribution in the field. Therefore, I would encourage the authors to pursue such a line of connectionist simulations. Some of the self protective biases go against well-know cognitive biases involved in judgments of others, such as memory neglect of negative self-inconsistencies which goes against empirical evidence demonstrating higher recall for inconsistent behaviors by others (Sedikides & Green, 2000), so that this is certainly not an obvious task.

**The Self as a Stream of Behaviors**

One of the concepts that I appreciated very much is Kashima et al.’s (2007) starting assumption that much of our learning is based on observations of other’s behaviors, and imitating them. The ability to learn a sequence of actions and the goals looming behind them is probably not only the basis of one’s own behaviors, but also of one’s own goals, aspirations and life ambitions. In this sense, the model captures an important aspect of the self that has been neglected in connectionist modeling in the social domain. However, I have two reservations.

First, although self-knowledge is generated on abstracted sequences of actions, this does not necessarily implies that self-knowledge predominantly takes a narrative form, as Kashima et al. claim. Most of what I can tell about my childhood past are short habitual events, like the way I went to school, how mother called me in after play, where I spent our vacation, but these memories do not involve an extensive chain of actions. Even special events that are very memorable such as when I nearly escaped suffocation in a trench do not have a sequential nature, but rather an instant-like characteristic. When you ask yourself: “Who am I?”, you most likely give a description that is not a sequence of behaviors, but rather a list of habits, attributes, characteristics, close others, and perhaps life goals. Obviously, people experience events the whole time, but this does not necessarily means that narratives are the most prevailing or compelling structure of one’s current self-perception. Fortunately, the I-SELF network is perfectly capable to learn associations between the self and a series of properties that are frequently present during behaviors conducted by the self, so that it can learn a list-like knowledge base of the self.

Second, Kashima et al. stress very much consciousness (see their title) as a basic aspect of the self, as well as the symbolic nature of it. This is a risky business when a connectionist model is applied. One of the most acclaimed accomplishments of connectionist learning is that no central supervisory agent or consciousness is needed. Learning can proceed without us realizing that it does. So too, I would suggest, can we learn about ourselves and act upon this knowledge, without us realizing that we do. Thus consciousness seems not to be a core aspect of a self-representation, especially in a connectionist framework (although people can bring their self to consciousness when they deliberate about it, but how exactly this happens is still a point of much speculation in research). It is more likely that in every-day social interactions, we act on a pre-conscious notion of our self, like probably our evolutionary ancestors also do.

The same goes for symbolic representation of actions. When symbols about behaviors are put together with actual behaviors in a simulation, one would expect some differences in their representation. An ecologically valid and natural solution would be to represent symbols by single nodes (localist encoding) because a major aspect of a symbol is its stable higher-level meaning, not is changing lower-level perceptual characteristics (like letters or sounds); and to represent behaviors as a whole array of lower-level perceptual features (distributed encoding) because each behavior is always somewhat different from the preceding one, in response to small differences in the context, even if its symbolic meaning is the same. The I-SELF model presented by the authors failed to incorporate such differential coding. Although this is not a major concern, the model appears quite capable to accomplish this, and it would have lent more ecological credence to it.

**A Minimal Architecture for Behavioral Associations and Sequences**

It is interesting to point out for the more technically interested reader, that the I-SELF model could have
been modeled in a more parsimonious way. Kashima et al. combined two well-known connectionist architectures, a feedforward architecture that is capable of associating symbols with behaviors and vice-versa, and a simple recurrent architecture that is capable to tie sequences of behavioral events to each other. However, the two models have much in common: They both include an input, hidden and output layer, and they both learn by back-propagation of the error. They differ only in the extent to which past events may inform the current input for sequence learning. Given this similarity, I reran Kashima et al.’s Simulations 1 and 2, either with a feedforward or a simple recurrent architecture. To learn behaviors associations, the network’s target or output was identical to the input, for learning behavioral sequences; the output was the next step of the input. Essentially then, the model was taught to associate an input with itself (for forming associations) or with the next input (for forming sequences). Both architectures by themselves were sufficient to reproduce the basic aspects of the simulations, although as one might expect, the feedforward net was slightly superior for behavior associations, and the recurrent net for behavior sequencing. It is puzzling why Kashima et al. (2007) did not choose the most powerful simple recurrent net for all their simulations, as there is no apparent need for a more complicated architecture.

Social Imitation and Interaction

As acknowledged by the authors, their model is limited in that imitating other’s behaviors presents problems for social coordination where one’s actions often follow in suit with other people’s actions, and are not necessarily an imitation, but rather complementary (e.g., give and take). This is, in my view, not a major problem. Children and adults could learn how blatant or more subtle sequences of social acts are performed by different actors, and so learn how one action complements another, and so form a natural unit. This was, in a crude sense, also modeled in their Simulation 2, where different actors performed a sequence of coordinated actions to accomplish their goals. People could learn from such observed sequences how to improve on their own social performance.

Perhaps more importantly, unlike the model suggests, one does not indiscriminately does or believes what has been told. The authors also acknowledged this limitation. Social interaction involves negotiation and persuasion. In our lab, we (Van Overwalle & Heylighen, 2006) recently modeled a social interaction process that allows for self-interpretation and negotiation, by developing a multi-agent network consisting of several individual networks or agents, each talking and listening to (some of) the other individual networks. Each individual network is a standard connectionist recurrent or auto-associative network, and the transmission of beliefs is regulated by connections between these networks, which reflected the trust in the information transmitted. Given this central role, we termed this multi-agent network the trust network. Thus, besides basic associations within in each individual, there are also ingoing connections representing how strongly the notions expressed by another agent are considered trustworthy. Kashima, Kashima and Aldridge (2001) developed a similar model in which they connected individual networks to each other, but this connection was direct as if the individuals’ brains are linked in one overarching cognitive system, and not mediated by information transmission and the trust in the information provided.

Although one might dispute the existence of separate trust connections as assumed in the trust network (Van Overwalle & Heylighen, 2006), it is undeniable that trust plays a major role in social interaction and belief updating. What is interesting is that trust in the model is learned and depends primarily on the fit between the notions expressed by another agent and the self’s own ideas. This is very similar to the popular delta connectionist algorithm, where learning depends on the fit between one’s internal representation of the environment and one’s direct perception of it. The trust model by Van Overwalle and Heylighen (2006) was therefore capable to model a number of key phenomena in meaning formation and persuasion. In the domain of social referencing, it simulated how agents agree to assign a referential symbol to a novel object, and how they resolve synonyms and ambiguities. In the persuasion domain, it simulated how minorities are over run by majority ideas unless their informational exchange is blocked, how information is biased by the self’s background during transmission and expression to other agents, and why people typically start a conversion with common beliefs (such as the weather) as this information is most trusted and thus survives initially. The trust model was an attempt to simulate social interaction more naturally, from the perspective of a social group context with multiple selves interacting with each other.

However, one important limitation that this trust model shares with the present I-SELF model is that it does not allow the agents to select their own actions or express their beliefs for their own profit (see Eiser, Fazio, Stafford & Prescott, 2003, for a connectionist model including some sort of self-agency) so that each agents’ self-serving behaviors and intentions could not be modeled. Social connectionism during the last decade has been predominantly cognitive, and perhaps the time has come to move to other, more motivational issues, involving people’s emotions, goals and behaviors to reap their personal benefits.
COMMENTARIES

Note

Address correspondence to Frank Van Overwalle, Department of Psychology, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussel, Belgium. E-mail: Frank.VanOverwalle@vub.ac.be.

References


