Psychological Inquiry
2007, Vol. 18, No. 2, 129–137

REPLY

Agency, Sociality, and Time

Yoshihisa Kashima, Aparna Kanakatte Gurumurthy, and Trevor Chong
Department of Psychology, The University of Melbourne, Melbourne, Victoria, Australia

Jason Mattingley
University of Queensland, Australia

To participate in the kind of intellectual discourse exemplified in *Psychological Inquiry* is at once a humbling and exhilarating experience. The researchers that one respects most devote their considerable capacities, efforts, and wisdom to analyze, interpret, and comment on, often critically, but always constructively, one's best effort at the time of preparing the target article. It is particularly rewarding when one finds—as we do—common themes, which may point to a new goal, encourage the discourse further, and present the possibility of a new conceptual horizon in *The Neverending Story* (to borrow the title of Michael Ende's fantasy and to echo Sedikides, Green, and Gregg, this issue) of scientific investigation. In this response, we construct three main threads that we identified in the commentaries, and provide responses to each. They are concerned with *agency, sociality,* and *time*. For each theme, we will first provide a general characterization of the issue and our general response; we will then give our responses to the comments on more specific aspects of the I-SELF and simulations, which are likely of greater relevance for those who are interested in connectionism and formal and simulation modeling of social psychological processes.

Agency

Can connectionism in general and the I-SELF in particular handle agency? This is the most intriguing metatheoretical question that emerged in a number of commentaries. Morf and Horvath (this issue) wrote, "agency, or goal-directed behavior, ... is not a component of the current model. (p. 112)" Hannover and Kühnen (this issue) noted, "as yet I-SELF does not capture the *self as an executive agent*. (emphasis, theirs; p. 103)" In a somewhat different vein, Rameson and Lieberman (this issue) made a similar point: connectionism may be best suited for the modeling of what they call the X-system, but not for the C-system implicated in the controlled and reflective process. Somewhat more indirectly, Sedikides, Green, and Gregg (this issue) too raised a closely linked question. How can the I-SELF handle the process of motivated self-protection, a goal-directed process that results in the maintenance of self-esteem. Most explicitly, Van Overwalle (this issue) asked, "Where is the self in connectionism?" citing Sedikides and his colleagues' work on self-protection as a critical point that he says the I-SELF is missing.

By agency, the commentators meant two aspects of this ill-defined, and yet generative concept. One is executive control (Hannover & Kühnen), and the other, goal-directedness and motivation (Morf & Horvath, Sedikides et al., and Van Overwalle). When the commentators referred to motivation, rather than goal-directedness, we surmise that they meant to emphasize the phenomenal experience of wishing, wanting, and desiring that often accompanies the goal directed activities – the experience that may reflect the embodied nature of goal-directedness as we will discuss later. These two components of agency are conceptually related. To executively control a process implies that a to-be-controlled process is influenced by another controlling process, so that the former is carried out to match a certain standard. This standard is the goal for the controlling process; feedback about the discrepancy between the standard and the to-be-controlled process then activates the controlling process further, and the discrepancy is reduced to approach and eventually match the standard. Hence, agency in the sense of executive control implies that (1) there exists a controlling process that can influence the to-be-controlled process, that (2) the controlling process is goal-directed, and that (3) there exists a cybernetic mechanism to approach the goal (or perhaps to avoid undesired ends). In the following, we use the term agency to mean the broad sense that includes these three subcomponents.

Connectionism can handle agency in this sense, if not entirely, at least to some extent. At the most abstract level, addressing the connectionist metaphor of personality, Morf and Horvath (p. 109, this issue) noted that "[a]gency is solely in the organization and activation of the emergent network." This is, in a way, true about most connectionist networks. Most connectionist algorithms can be construed as an optimization mechanism, in which they reduce the discrepancy between a desired goal and a current state. In particular, the backpropagation (generalized delta, or error-driven learning) algorithm used in our and many other simulations in social psychology does exactly that. In this sense, it learns to move the system towards a certain goal. More concretely, some connectionist mechanisms could model reinforcement learning and other types of goal-directed behaviors, as well as cognitive control processes such as performances on the Stroop task (O'Reilly & Munakata, 2000). Connectionism should be construed as a general approach to modeling psychological processes within which specific models can be constructed. Viewed this way, we would argue that it does have a potential to model agency.

More specifically about connectionism's capacity to model executive control based on explicit and symbolic (or language-based) representations, Hannover and Kühnen (this issue) noted that "we do not believe that [connectionism] can substitute theories that rely on symbolic self-models. . . . . Symbolic self-models provide analytic differentiations between varying qualities of self-knowledge and processes operating on knowledge representations. (p. 104)" Likewise, echoing McClelland, McNaughton, and O'Reilly (1995; also see Smith & DeCoster, 2000), Rameson and Lieberman (this issue) too suggested the reflective and controlled process that the C-system carries out may be fundamentally different, and may be more explicit and symbolic. The basic contention is that connectionism may be a poor metaphor for explicit and symbolic representations, and a symbolic model may be useful for characterizing the more explicit process. This contention has an intuitive appeal. It is possible to construct a hybrid model that combines the strength of connectionist and symbolic models. Nonetheless, it is debatable whether connectionism cannot model symbolic processes in principle. At the very least, there exist some connectionist models that are capable of representing language-based processes, and more sophisticated attempts within the connectionist framework are continuing (O'Reilly & Munakata, 2000). It remains to be seen whether these connectionist models of symbolic processes are capable of modeling explicit executive control (e.g., Now that my thoughts are going in circles, I should stop and take a break.).

Nonetheless, it is important to consider the above issue within the context of the on-going debate about the automatic and controlled, or implicit and explicit, processes as two distinctive modes of information processing. In the Volume 17, Issue 3 of *Psychological Inquiry*, Kruglanski, Erb, Pierro, Mannetti, and Chun (2006), Deutsch and Strack (2006), and Sherman (2006) made their cases for one, two, and four distinct processes. No fewer than ten commentaries follow the three target articles. A parallel debate occurred in Psychological Bulletin surrounding Gawronski and Bodenhausen's (2006) associative-propositional evaluation (APE) model, accompanied by three commentaries. Gawronski and Bodenhausen drew a distinction between the associative and propositional processes, whereas others did not necessarily agree. The issue is too complex to explore in this short response; however, what this shows is the inconclusiveness of the assumed distinction between the two processes. In this juncture, perhaps most intriguing is Albarracín, Noguchi, and Earl's (2006) exploration of Joyce's *Ulysses* and Virginia Woolf's *Jacob's Room*, two of the best known examples of the literary exemplars of stream of consciousness. In *Ulysses*'s 267,198 word corpus, they counted the total of 18 instances (astounding less than .007%) of words that imply conscious control (try/tried, intend/t, attempt, goal) and most of them pertained to someone else's perceived mental states; in *Jacob's Room*, eight out of 55,094 words (less than .015%) had the root inten-. Obviously, this count underestimates the frequency of mental events of intending because the unit of counting for events is better approximated by sentences than by words. Still, this raises an intriguing question. In enculturated streams of consciousness, how evident is the phenomenal experience of conscious control?

Can the I-SELF model agency? The simulations reported in the target article do not directly address this question. However, in principle, we believe that the I-SELF architecture is capable of simulating some aspects of agency. Its two components speak to this. First, the I-SELF can learn a narrative of goal-directed activities. When the protagonist's goal is appropriated by the system (see some aspects of simulation experiment 2), the system pursues the protagonist's goal as its own goal. If a learned narrative contains a sequence of actions that can be executed to regulate the approach towards the goal, the I-SELF should in principle be able to follow this sequence as well. As Green (this issue) noted, appropriation of goals is a critical issue in this process. Furthermore, the I-SELF could in principle learn a narrative that describes a sequence of actions to follow when a first attempt to approach the goal fails. Thus, the I-SELF can in principle learn to enact a variety of coping strategies in the case of a failed attempt. So, if events fail to unfold as expected in a scripted course of an event sequence, the I-SELF may be trained to take a different course of action by providing it with different contingencies of narrative events. Contrary to Morf and Horvath (this issue), we contend that the I-SELF is in principle capable of simulating agency in the sense of goal-directedness. The choice of the word *I* in the I-SELF is deliberate to imply that it is about the process of *I* in the James-Mead sense, and therefore it is about agency, although we are the first to concede (as we did in the target article) that the current simulations need to be scaled up to make a stronger claim.

Second, the I-SELF architecture can in principle simulate the embodiment, and we believe this component is critical for a psychological theory of agency. In simulation experiment 1—admittedly a simple demonstration—we tried to show that feedforward networks in general and the I-SELF in particular can learn to associate language-based symbol uses with embodied behaviors. We will examine Read and Monroe's (this issue) critique of this aspect somewhat later. The phenomenal experience of goal-directed behavior—well captured by such words as wishing, wanting, and desiring—may reflect the embodiment of the goal-directed behavior. Antonio Damasio (1994/2006), a well known cognitive neurologist, hypothesized that the phenomenal experience of emotion and feeling associated with one's goal-directed behavior may reflect what the body has learned in performing the behavior (also see Bechera, Damasio, Tranel, & Damasio, 1997), and he called it a somatic marker. He further speculated that these emotions and feelings are critical for one's experience of the self. Likewise, Leung and Cohen's (in press) investigation point to the significance of embodiment in an enculturated experience of oneself in time and space. Contrary to Van Overwalle's (this issue) charge, it is not just the word *I* or *We*, that marks the self process, but one's feelings of one's own singular body as reflected in the symbol-body association, that marks the phenomenal experience of the self in the I-SELF.

These considerations help us to address Sedikides et al.'s challenge about self-protection and Morf and Horvath's point about narcissism. To focus our discussion, let us concentrate on the intriguing mnemonic neglect (MN) effect. In the I-SELF, strong motivations to protect one's self-esteem may be understood as a well-learned (and possibly even genetically predisposed) tendency to enact a self-narrative whose goal (or *Object* of desire in Greimas's scheme) is to maintain the image of oneself that is true, good, and beautiful (mostly "good" and secondarily "true and beautiful," perhaps). Such a narrative may have a number of different plots and subplots with myriad contingencies and strategies to protect one's self-esteem. The MN effect is hypothesized to result from strategies reflected in plots and subplots of self-protection. Again, we contend that in principle connectionism and the I-SELF may be able to simulate it though this reply is not a place to go into the details of just how it may be done. More detailed simulation experiments are needed in a separate occasion. At this point, however, several interesting possibilities come to mind. Green, Sedikides, & Gregg (2006, cited in Sedikides et al., this issue) showed that self-threatening words are less likely recalled than, but as equally likely recognized as, self-affirming words. This suggests that self-threatening information is encoded and stored in memory, but not retrieved. This is corroborated by Sedikides and Green (2000); who showed that the MN effect emerged when there was ample time to encode self-threatening information. There are at least two (maybe more) ways to explain this finding. One is a possibility of self-censorship. A connectionist architecture capable of selecting certain types of information for communication needs to be constructed. This is a challenge in and of itself because it involves an executive function, though we contend that in principle it is possible. The other is a process of constructing a retrieval cue. In this case, in retrieving self-relevant information, different types of retrieval cues may be constructed depending on the state of the memory system. When retrieval cues are positive in valence, positive self-affirming information may be more likely retrieved; if retrieval cues include negative representations, negative self-threatening information may also be retrieved. This latter possibility seems to be congruent with Green et al.'s finding that ego-inflating feedback eliminated the MN effect.

Finally, the question of agency brings us to the nature of *I* as conceptualized by James and conceived in the I-SELF. As we explicated in the target article, according to James, the current thought is *I*, and our memory of the experience of *I* is about *Me*. In this metatheory, thought and *I* are coextensive. James would say researching *I* is about researching thought; researching thought is about researching *I*. So, Morf and Horvath's and Van Overwalle's observation that the I-SELF is about thinking in general is in fact right. It is designed to be about thinking in general. The whole point was to show that general connectionist architecture like the I-SELF can simulate some rich and varied phenomena. In support of his claim that self-processes are special, and not general, Van Overwalle (this issue) raised the knotty issue of neural representations of the self, which he says are special. In contrast, Rameson and Lieberman (this issue) suggested that the issue of where the self is "localized in the brain" is not so clearly resolved yet, and that it may depend, in a way, on time. According to them, medial prefrontal cortex (MPFC), which is often implicated in self-referential processes, may not be solely concerned about the self, but more generally about social objects for both self and others. However, depending on how much one knows about oneself (i.e., how long one has experienced oneself engaging in a particular domain of activity; both age and experience would matter), dorsal or ventral MPFC may be implicated for self-referential processes relative to other-referential processes. The research is on-going, and the issue of neural representations of the self seems to elude a clear answer. Rameson and Lieberman noted "the multifaceted nature (p. 119)" of the self and suggested that "there is probably no one specific "self" area of the brain. Rather, many component processes work together and under different circumstances to generate our sense of self. (p. 119)"

Connectionism and Agency

To address more technical aspects of connectionism and agency, we will first discuss our modeling principle, which will help us address general criticisms about connectionist or other formal modeling approaches, and then examine more specific issues pertaining to our simulation work.

Modeling Principle. We believe that formal modeling of psychological processes must be guided by a broader *metatheoretical goal*. The metatheoretical goal, namely, what it is that the model is designed to deal with, determines the modeler's choice of the level of analysis, and its *ontology*, that is, what the model assumes to be its basic components and structure. We also believe that formal modeling of psychological processes, and connectionist models in particular, must be *principled*: a modeler adopts a general principle and chooses to be constrained by it. To the extent that the principle is well reasoned and justified, it serves as a guide and a constraint. Otherwise, a modeling attempt can easily become a *post hoc* adhockery. In the case of connectionist models, this problem can be particularly acute because connectionism is a general framework that can have a large number of parameters on a par with something like natural language in which to construct specific models. In our case, the metatheoretical goal was to construct a physically possible and plausible realization of the James-Mead dynamic self at the level of psychological functions, which are nonetheless underpinned by their neural correlates (we called it a Functional Artificial Neural Network System). As we will see, this metatheoretical goal helps us to address the commentators' concerns.

General criticisms of formal and connectionist modeling. The worry of post hockery is reflected in Morf and Horvath's (this issue) general criticism of connectionist and other modeling approaches: they can only show what's already known. To this worry, however, we can simply suggest that this is not the case in general. In one of the early attempts to use connectionism to model social cognitive processes, one of us (Kashima & Kelekes, 1994; Kashima, Woolcock, & Kashima, 2000) used a connectionist model to make novel predictions about the effect of judgment timing on impression judgments; this was tested and supported by empirical data. Clearly, this was not a *post hoc* modeling of existing data. In relation to the I-SELF, it did in fact make significant novel predictions when we were conducting simulation experiment 2. The I-SELF showed that reading a story with a well-learned narrative structure may result in the *automatic* appropriation of the narrative goal by the reader. This is because there is nothing in the I-SELF that requires its "conscious" adoption of the narrativized goal. We began to investigate this possibility empirically in our lab; we also began to review the literature on automatic goal activation. To our delight, we found a growing literature on what Aarts, Gollwitzer, and Hassin (2004) called goal contagion: reading a story about someone else's goal pursuit can automatically activate the corresponding goal in the reader. This is another example of non *post hoc* nature of simulation research. It is true that many simulation studies are conducted to explain existing data after the fact; however, this is not always the case.

Furthermore, principled simulation research is not *ad hoc* either. Far from it, principled modeling brings rigor and coherence to one's research. Read and Monroe's (this issue) critical question about the relevance of the neuroscience data in our discussion of the I-SELF architecture serves as an illustration. As we stated at the outset, we thought it important for a model of psychological processes to be linked to the neural level in some way. We chose to do this at the *functional* level: this principle is reflected in what we called the Functional Artificial Neural Network System. The general modeling principle can be translated to this particular case as follows: If there are two psychological functions that are thought to be carried out by two distinctive neural substrates, these psychological functions should be modeled by two components within a model. The two components of the I-SELF (the imitative feedforward network and the sequence-learning simple recurrent network) themselves reflect recent empirical findings and models in cognitive neuroscience. Specifically, these data suggest that the abilities to imitate a novel action and to sequence a series of actions involve overlapping, but largely separate, neural substrates. Several neuroimaging studies suggest that the capacity to imitate is subserved by areas of the mirror neuron system – namely, the inferior frontal gyrus, the inferior parietal cortex and the superior temporal sulcus (e.g., Iacoboni et al., 1999; Koski et al., 2002; Grezes et al., 2003; Nishitani & Hari, 2000, 2002). Once an action is encoded and represented, it can then be incorporated into a sequence that may be subsequently executed. The organization of movement sequences involves several areas that are quite separate from the mirror system, and include cortical areas, such as the prefrontal cortex, supplementary motor area (SMA) and pre-SMA, and subcortical areas, such as the basal ganglia. In particular, the SMA and pre-SMA are thought to be critical in encoding the serial relations of movements in a sequence. The involvement of these areas has been demonstrated directly, through electrophysiological recordings in non-human primates (Tanji, 2001; Tanji & Shima, 1994), and through disruption of the human SMA with transcranial magnetic stimulation, which interferes with the organization of subsequent movements in a sequence (Gerloff, Corwell, Chen, Hallett, & Cohen, 1997). The evidence for the involvement of different neural circuits underpins and justifies the *functional* structure of the I-SELF. Although Van Overwalle (this issue) is right in suggesting that either the feedforward or simple recurrent network alone is sufficient to simulate the results of some simulation experiments, our decision to include two components to the I-SELF gives it a greater fidelity with the neuroscience literature. We believe that the principle of linking psychological and neural data through the analysis and modeling of cognitive *functions* is useful in grounding models of self and identity; in particular, those of the James-Mead dynamic self.

Finally, both Morf and Horvath (this issue) and Green (this issue) raised the perennial issue associated with modeling or simulating psychological processes—that they are too simple and the reality is far more complex. Yes, a model simplifies the infinitely complex reality, but herein lies its strength and its weakness. With simplification, a model may bring new insights; with simplification, a model loses the richness of the social psychological reality. Indeed, any modeling attempt involves a trade-off. On the one hand, a modeler wishes to capture what he or she regards is theoretically most significant and parsimonious. On the other hand, a modeler wishes to capture the richness of reality; a model should have high fidelity. However, there is a dilemma between abstraction and fidelity; you can't have both. How does a modeler resolve this dilemma? In our view, this must be guided by a metatheoretical goal and the existing knowledge about the domain that the modeler wishes to model. As we will see below, this dilemma reverberates throughout our dialogue with Read and Monroe (this issue) and Van Overwalle (this issue).

Simulating embodiment with the I-SELF. Read and Monroe (this issue) charged that our simulation of embodiment (simulation experiment 1) is too simple and its results were obvious. Indeed, it is simple and obvious for connectionists whose objective is to model neural processes. However, we believe the point is worth making in modeling the James-Mead dynamic self. Read and Monroe's assessment reflects a difference in metatheoretical goal. Our simulation assumes that there exists an allocentric representation of another agent's body and its movements, and there exists an egocentric representation of one's own body and its movements. It further assumes that there exists an input representation of symbolic codes (perhaps in auditory or visual codes) and an output representation of symbolic codes (perhaps in vocal or motor, e.g., hand movements). As we noted from the outset, we wished to construct a *social psychological* model of the James-Mead dynamic self. Therefore, "[p]rocesses such as modality specific perceptions, motor behaviors, and language comprehension using syntactic, semantic, and pragmatic knowledge are assumed to occur, rather than modeled explicitly (Kashima et al., 2007, p. 77)." Our point was simply that the I-SELF connectionist architecture is capable of associating the processing of behavioral information and symbolic information. Although Read and Monroe (2007, p. 123) argued that "we do not see in what *interesting* sense 14 nodes can represent such detailed things as the agent's or self's egocentric spatial and movement information. (our emphasis)" What is "interesting" depends on the reader's frame of reference and tacitly adopted approach to theorizing and modeling. We surmise that Read and Monroe wish to model embodiment (and imitation, as we will see) at a lower-level of specific visual processes and motor movements, as well as at the level of specific language processing. Indeed, it would be "interesting" to model embodiment using a simulator such as PDP++ (O'Reilly & Munakata, 2000), which is designed to model at that level of fidelity. However, it was not our objective; we were not *interested* in that. As we explicitly stated in the target article and discussed earlier, our objective was to model at the *functional* (algorithmic) level, rather than at the *neural* (implementational) level, although we maintain some degree of mapping between the functional and neural levels of analysis. Our objective was simply to point out the importance of embodiment in the James-Mead dynamic self and to suggest that a feedforward network is capable of modeling it, rather than to simulate the details of the phenomena of embodiment. Every psychological modeling involves a trade-off as we noted: in the present case, one may attempt to model a high-level function or to model with greater fidelity its ontological specifications. Read and Monroe's tacit metatheoretical goal emphasizes the former less and the latter more than ours.

Sociality

What kinds of sociality can connectionism and the I-SELF deal with? The comments have two centers of gravity. One is about the process—the I-SELF is supposed to imitate, but how does it do that, and how can it handle more complex and sophisticated social interaction that humans clearly engage in. Adler and McAdams (this issue), Green (this issue), and Van Overwalle (this issue) suggested that the process by which cultural information is transmitted and learned is likely more complex than mere imitation. As we implied in the target article, a culture provides narrative types that are appropriated by individuals for the construction of their narrative identities (Adler & McAdams; Green). There may be individual differences in the extent to which narratives impact on their construction of identity (Green). Like Adler and McAdams (this issue) and Green (this issue), one of us (Kashima, Klein, & Clark, 2007) argued that social communication is a result of an active collaboration between the communicators, and particularly for narrative communication, a co-construction of a narrative understanding between the storyteller and listener. The active roles of the listener and the speaker, and also the dynamic interaction between the two, are clearly an issue.

The second center of gravity is about the content—the I-SELF processes a narrative, but whether and how it can handle social information contained in narratives. Generally speaking, narrative is a powerful representational device that can contain significant information about social relationships and how to handle them. A narrative contains not only the protagonist (Greimas's *Subject*), but also other characters (e.g., Greimas's Helper, Opponent, Sender, or Receiver), and the main body of a story is often about the relationships among them, be it a friendship or a conflict. It conveys useful social information. In line with this reasoning, Mesoudi, Whiten, and Dunbar's (2006) findings suggest that stories that contain social, especially emotive, information are likely to be transmitted in communication chains. Mar, Oatley, Hirsh, dela Paz, and Peterson (2006) showed that those who have been exposed to stories in their life long reading of narratives in print were more socially able than those who prefer non-narrative non-fictions. As Green (this issue) correctly noted, narrative readers can form a variety of relationships with the characters in the narratively simulated social world. Although computer simulations showed that the I-SELF appropriates the story, and may empathize and identify with the protagonist—a strong form of relationship between the self and the character—there are other forms of self-character relationships: a role model, partner, friend, or even a negative role model—someone one tries not to emulate. Can the I-SELF handle such nuanced relationships with a character?

These are significant challenges. We will attempt only a sketch of what may need to be done. To begin with, to model a nuanced interaction between the self and other, and more specifically to simulate communication between them, the self-other differentiation and coordination must be learned in the I-SELF. At one level, the current model is capable of doing this (see simulation experiment 4); however, the self-other coordination is a significant challenge as we noted in the original article. Furthermore, the I-SELF must be able to enact interpersonal and communication behaviors and respond to an interaction partner's behaviors in different types of relationships. In principle, a *scripted* form of social interaction sequences in a type of social relationship in a type of social context can be simulated even in the current I-SELF; this type of information can be learned from stories. Perhaps this is what Van Overwalle (this issue) had in mind when he said the self-other coordination would not be a major problem. Nonetheless, a flexible handling of disruptions to a scripted sequence of behaviors is a challenge that needs to be addressed more concretely in the future although we believe it is possible in principle. Finally, even if a connectionist model can simulate a certain social relationship, the model will then need to be able to use it to construct a relationship with a narrative character.

Connectionism and Sociality

Van Overwalle and Heylighen's (2006) *trust net* is an innovative solution to some of the problems of social communication as outlined above. In this model, one connectionist agent is connected to another agent, such that the strength of this inter-agent connection is modified as a function of the discrepancy between the two connectionist networks' inputs and outputs. The connection strength is conceptualized as reflecting the receiving agent's trust of the sending agent. This model's strength lies in its ability to simulate a number of phenomena in social communication based on a set of simple assumptions about the social cognition and communication embodied in their learning and trust modification algorithms. It is a groundbreaking step in connectionist modeling combined with the multiagent modeling perspective.

There are some significant differences between the trust net and the I-SELF. First of all, there is a significant difference in metatheoretical goal, and hence the model's ontology. The trust net *assumes* that the mechanism of social communication operates, and then models one of its (clearly significant) aspects in terms of trust connections. In contrast, the I-SELF approaches sociality from ground up. We start from what we regard is a most basic process of sociality, namely, imitation, and to set up an architecture with greater fidelity in terms of psychological functions. This is reflected in a difference in the ontological assumption about sociality. As Van Overwalle (this issue) noted, the ontological assumption of sociality in trust nets may be disputed. That is, a trust connection between connectionist networks can be construed as a convenient fiction from a psychological perspective; in a typical psychological ontology, trust exists not as a connection between agents as the trust net assumes, but as one agent's belief about another agent. In contrast, sociality in I-SELF does not have this problem; the current I-SELF is designed to handle only a limited aspect of sociality, namely, imitation. Trust nets purchase its power to model sophisticated (though still limited) social communication and meaning negotiation at the expense of its low-level fidelity. This is a trade-off that every modeling attempt faces; it is a matter of what each model is designed to do. The I-SELF tries to model sociality from a low-level consideration; the trust net starts from a higher-level consideration of trust.

In this respect, Read and Monroe (this issue) raised a significant question about the way in which the I-SELF modeled imitation. However, their critique again partly reflects differences in metatheoretical goal. They seem to be interested in modeling imitation with greater fidelity; our objective was to show that the general learning architecture of the feedforward network, which is a component of the I-SELF, can model *in principle* the process of imitation. Our metatheoretical goal was not to give detailed account of imitation learning, but to show that the I-SELF is capable of handling imitation that Mead assumed to underlie human sociality. Nonetheless, we agree with their assessment that the mechanism of imitative learning is a contentious issue. It is true that the proponents of mirror neurons often imply that the mirror neurons constitute a specially evolved modular mechanism, whereas some others (e.g., Brass & Heyes, 2005) suggest that imitation is a result of general purpose learning mechanism. They essentially suggest that mirror neurons might *do* imitation, without being *for* imitation. For instance, Heyes's Associative Sequence Learning (ASL) model (Heyes & Ray, 2000; Heyes, 2001) suggests that imitation does not rely on some innate special module, but depends more on learning. Through a process of general learning, the associations between a sensory percept of an action and our motor representation of it may be strengthened. The effect of this experience is then to reconfigure general-purpose cognitive mechanisms through an associative learning process. Our attempt to model imitation is generally congruent with this model.

In retrospect, our modeling ontology is pitched at the metatheoretical level between that aspired by Van Overwalle's trust net and what Read and Monroe seem to assume. It is a consequence of adopting the James-Mead dynamic self as a metatheoretical framework in our research. Although it is possible that one day these ontological levels connect with each other, there is clearly a long way to go before then. These differences in opinion notwithstanding, we share with them a belief that a formal modeling (and connectionist modeling in particular) of sociality can help us advance our understanding of human sociality.

Parenthetically, there is one minor issue. Van Overwalle (this issue) cited a connectionist multiagent model that one of us was involved in (Kashima, Kashima, & Aldridge, 2001), and commented on its ontological assumption, suggesting that its multiple agents were connected as if their brains were directly connected. This is a misunderstanding. In Kashima et al. (2001), multiple agents were connected to each other via output and input layers, which reflected observable objects and behaviors.

Simulating transportation and priming with the I-SELF. Read and Monroe (this issue) raised issues about our simulations. First with regard to transportation, they argued that transportation as Green and Brock (2000) originally examined was about the experience of reading a narrative for the first time (i.e., learning phase), rather than the experience of re-reading it (i.e., testing phase). What we tried to examine in our simulation was the situation in which one learns a certain *generic* narrative structure (as characterized by Greimas's *Subject*), and then reads a *specific* story with the same narrative structure while identifying with the narrative protagonist. We first ascertained that, if the generic narrative structure is well learned, the identification with the protagonist (i.e., activation of *I* rather than *Subject*) results in the appropriation of the story (i.e., being able to reproduce the sequence correctly). We then tried to examine whether the appropriation effect was strengthened when the reading of the specific story resembled transportation (i.e., taking its own output replaced with I as an input for the reproduction of the sequence), relative to when it was not like transportation (i.e., taking the original replaced with *I* as an input). Our results showed that the reproduction was more accurate under what we regarded as a transported reading experience. Our tacit assumption was that any new reading of a transporting or non-transporting story is done against the context of previous readings of similar narratives. We attempted to simulate *this* kind of transportation in our work.

Read and Monroe (this issue) suggested that a simulation experiment may be more convincing if transportation is manipulated at the learning phase, so that the I-SELF learns a story with *I* as the protagonist in the transported condition, but the same story with *Subject* as the protagonist in the non-transported condition, and is tested with *I* in the test phase. We conducted just this simulation experiment. For each replication, a set of random starting connections was generated for a network, and it was used for the training (75 times) with the *I* story and for the *Subject* story. We then tested the network with the *I* story and examined the number of mistakes it made and the amount of errors (sum or squared differences between the desired output vectors and the observed output vectors) it generated. We replicated this 10 times using different random starting connections. In the *Subject* story (non-transported) condition, the network made mistakes in four replications and the average amount of errors was 2.23; in the *I* story (transported) condition, the network made no mistake and the average amount of errors was 1.35. The two conditions differed in the average errors,

$t(9) = 7.18$, $p < 0.001$. Thus, the transported condition as Read and Monroe defined it did produce a better performance than in the non-transported condition. Thus, either way, the I-SELF is capable of simulating transportation.

Second, Read and Monroe wished to see the priming simulation done somewhat differently. They suggested that the I-SELF should be trained with the *I*-story in the individual condition and with the *We*-story in the collective condition; *I* is associated with Agent 1 only, but *We* is associated with Agent 1, 2, and 3; and the activation levels for Agent 2 and 3 should be tested when the trained network is tested by activating Agent 1. We followed their suggestion by training the network with the two versions of the same story using random starting connections for each replication, and conducting 10 replications with different random starting connections. As expected, in the individual relative to the collective condition, Agent 1 was more, but Agent 2 and 3 were less activated. The mean activation levels for Agent 1, 2, and 3 in the individual and collective conditions were as follows: $M_{\text{I}-\text{story Agent 1}} = 0.90$ vs. $M_{\text{We}-\text{story Agent 1}} = 0.80$, $t(9) = 8.79$, $p < 0.001$; $M_{\text{I}-\text{story Agent 2}} = 0.06$ vs. $M_{\text{We}-\text{story Agent 2}} = 0.79$, $t(9) = -\ 0.65.75$, $p. < 0.001$; and $M_{\text{I}-\text{story Agent 2}} = 0.06$ vs. $M_{\text{We}-\text{story Agent 2}} = 0.80$, $t(9) = -66.46$, $p. < 0.001$. Again, both our simulation as well as the simulation Read and Monroe suggested show that the I-SELF can simulate something like the priming effect found in the empirical literature.

We believe these additional simulations strengthened our case for the I-SELF.

Time, Culture, and Self

Agency and sociality unfold in time; so does the self. In the James-Mead model, temporality is critical in self-processes. We argued that currently connectionism offers the best chance of capturing this temporal dynamism. Morf and Horvath (this issue) agreed; however, they suggested that the existing connectionist metaphors of personality capture the temporal dynamics, and wanted to know what constitutes the I-SELF's contribution beyond them. We take their point that the existing connectionist inspired treatment of personality captures some aspects of temporality. Nonetheless, a clarification is in order. We did not mean to suggest that it does not address temporality generally, but meant to say that they do not capture *William James's* temporality as envisaged in his notion of stream of consciousness as a flow of partly (though not completely) coherent chain of thoughts. Clearly, none of the existing simulations in personality that Morf and Horvath cited models sequence learning and reproduction or the kind of narrative self we tried to simulate in the target article. So, these are our new contributions; other commentators seem to agree with our assessment.

Second, although they clearly acknowledged the dynamic and evolving nature of self as we do, Morf and Horvath (this issue) emphasized the *synchronic* integration of self via *if . . . then self-signatures*. Agents "come to respond to particular types of trigger conditions that are perceived as relevant to their self-goals with characteristic thought, emotion, and behavior patterns. (p. 109)" Such condition-action contingencies (action conceptualized very broadly to include not only overt behaviors, but also thoughts and feelings) learned by an agent over time become the agent's self-signature. In a way, the I-SELF's feedforward network can capture this aspect of condition-action contingencies. Basically, an activation pattern in the input layer of the feedforward network can represent a certain condition-action contingency; when a given condition is activated, this network will generate the associated action. (Parenthetically, Van Overwalle, this issue, suggested that not all self-relevant memories are narrative; we agree – the I-SELF represents such synchronic information within the feedforward network, and therefore recognizes this aspect of self-relevant memories). What is missing in self-signatures is the conception of temporality that Adler and McAdams (this issue) called *diachronic* integration of self over the life span. That is, one may construe oneself as going from one stage of one's life to another stage; this may be a redemptive sequence or a fall from grace. How one construes this sequence is clearly linked to an aspect of one's personality and self. We argue that it is this kind of narrative temporality that we try to capture in line with the James-Mead dynamic self-metatheory.

Finally, Morf and Horvath pointed out an intriguing analogy between culture and personality. Both are dynamic and situated. We agree. As they observed, our simulation experiment 4 can easily be construed as modeling individual differences. Their general point is deep, and has a far-reaching metatheoretical implication. To put it simply, when one conceptualizes culture as *if . . . then* condition-action contingencies shared among a group of people, cultural differences become a subset of individual differences, as Morf and Horvath seem to think. This is in fact one of the metatheoretical challenges facing the concept of culture generally. However, one of our metatheoretical goals is to explore under what circumstances what types of *if . . . then* contingencies are generated and is shared among a group of agents as a result of the dynamic *social* interaction among them. Construed this way, culture is no longer reducible to individual differences or personality. Although we did not report any of the simulations that examined a consequence of transmitting some cultural patterns from one agent to another by imitation, we are in the process of conducting simulations and writing up their results. What the target article attempted to do was to provide groundwork and to signal the possibility of such dynamic conceptions of culture based on microprocesses of individual social interaction. We believe the James-Mead dynamic self-metatheory in general and something like the I-SELF in particular can move us closer towards this goal. Clearly, however, we have a long way to go.

Conclusion

The commentaries appraise, challenge, and urge on. In response, the research program is further consolidated, refined, and enriched. Just as the metatheory of the James-Mead dynamic self would suggest, thinking is not a solitary activity, but a social one enabled by the rich cultural resources provided by the intellectual tradition and community of researchers. Thus, the stream of enculturated consciousness continues, and so does the story.

Note

Address correspondence to Yoshihisa Kashima, Department of Psychology, The University of Melbourne, Parkville, Victoria 3010, Australia. E-mail: ykashima@unimelb.edu.au

References

Albarracín, D., Noguchi, K., & Earl, A. N. (2006). Joyce's *Ulysses* and Woolf's *Jacob's Room* as the phenomenology of reasoning: Intentions and control as emergent of language and social interaction. *Psychological Inquiry, 17*, 236–245.

Aarts, H., Gollwitzer, P. M., & Hassin, R. R. (2004). Goal contagion: perceiving is for pursuing. *Journal of Personality and Social Psychology, 87*, 23–37.

Bachera, A., Damasio, H., Tranel., D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275*, 1293–1295.

Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Science, 9*, 489–495.

Damasio, A. (1994/2006). *Descartes' error*. London, UK: Vintage Books.

Deutch, R., & Strack, F. (2006). Duality models in social psychology: From dual processes to interacting systems. *Psychological Inquiry, 17*, 166–172.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692–731.

Gerloff, C., Corwell, B., Chen, R., Hallett, M., & Cohen, L. G. (1997). Stimulation over the human supplementary motor area interferes with the organization of future elements in complex motor sequences. *Brain, 120*, 1587–1602.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology, 79*, 401–421.

Green, J. D., Sedikides, C., & Gregg, A. P. (2006). *Forgotten but not gone: The recall and recognition of self-threatening memories*. Manuscript under review. Virginia Commonwealth University. Cited in Sedikides, Green, & Gregg (this issue).

Heyes, C. (2001) Causes and consequences of imitation. *Trends in Cognitive Science, 5*, 253–261.

Heyes, C. M., & Ray, E. D. (2000). What is the significance of imitation in animals? *Advances in the Study of Behavior, 29*, 215–245.

Kashima, Y., & Kerekes, A. R. Z. (1994). A distributed memory model of averaging phenomena in person impression formation. *Journal of Experimental Social Psychology, 30*, 407–455.

Kashima, Y., Woolcock, J., & Kashima, E. S. (2000). Group impressions as dynamic configurations: The tensor product model of group impression formation and change. *Psychological Review, 107*, 914–942.

Kruglanski, A. W., Erb, H.-P., Pierro, A., Mannetti, L., & Chun, W. Y. (2006). On parametric continuities in the world of binary either ors. *Psychological Inquiry, 17*, 153–165.

Leung, A. K.-Y. & Cohen, D. (in press). The soft embodiment of culture: Cambera angles and motion through time and space. *Psychological Science*.

Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., & Peterson, J. B. (2006). Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *Journal of Research in Personality, 40*, 694–712.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary earning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419–457.

Mesoudi, A., Whiten, A., & Dunbar, R. (2006). A bias for social informaiton in human cultural transmission. *British Journal of Psychology, 97*, 405–423.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.

Sherman, J. W. (2006). On building a better process model: It's not only how many, but which ones and by which means. *Psychological Inquiry, 17*, 173–184.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108–131.

Tanji, J. (2001). Sequential organization of multiple movements: involvement of cortical motor areas. *Annual Review of Neuroscience, 24*, 631–651.

Tanji, J., & Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. *Nature, 371*, 413–416.