

what is presented with the SCM has convergences with a social cognition model we advanced (Semin & Cacioppo, in press a; in press b), although the social cognition model we have advanced is cast in a different mold, in particular with respect to the distributed processes taking place between two or more individuals.

However, the SCM relies on *reception*, namely, the construction of inner neural representations based on observed behavior. Second, social cognition is restricted to a *reproduction metaphor* (e.g., empathy, resonance, imitation, shared representations, or mindreading). Finally, the model attempts to provide an answer as to how intersubjectivity is achieved. However, the model remains at a purely representational level, neglecting the reciprocal nature and co-regulation of social behavior. The three R's (reception, reproduction, and representation) are conceptual consequences of relying on an individual-centered paradigm. There is a problem when social cognition, especially social cognition that emphasizes imitation, is centered on the individual, however. Cognition evolved for the control of adaptive *action*, and social cognition evolved for the control of adaptive interaction in response to evolutionary demands for the organism's survival and reproduction, which for humans always takes place in a social context (Caporael 1997; Fiske 1992), and involves the *co-regulation of action*. Imitation of a parent by an infant is not a solitary event in the service of social cognition. Instead, the infant's imitative behavior elicits an imitative or nurturant response by the parent, which not only reinforces the infant's imitative response but also establishes a connection and constitutes a co-regulation of action by the parent and infant. Any depiction of the social cognition of imitation that ignores the interaction and emergent information between individuals is incomplete.

Thus, social cognition is not driven entirely by inner processes and representations as the SCM suggests, but relies on resources that are distributed across neural, bodily, and environmental features (e.g., Agre 1997; Brooks 1999; Hutchins 1995; Kirsch 1995) with the social and physical environment supporting social action and interaction (Smith & Semin 2004). As this example illustrates, two or more individuals are capable of (a) joint work to perform a feat that supersedes their individual capabilities, and (b) co-cognition and co-regulation to achieve this joint feat. Co-regulation encompasses qualitatively different forms of co-action. The first is *entrainment* and is exemplified by periodic co-action and occurs in cycles. This can be illustrated with the example of rhythmic clapping (e.g., Neda et al. 2000). The second form is non-periodic co-action illustrated by *mimicry* or imitation (e.g., Chartrand & Bargh 1999). The third case is exemplified when people have to perform a complex task requiring interfacing each other's actions (as in open-heart surgery or playing tennis). The third case entails the execution of complementary actions, namely *coordination*, in the pursuit of accomplishing a task (e.g., successful surgery, winning in tennis).

Entrainment, mimicry, and coordination can obviously all occur simultaneously and to different degrees during social interaction. Take, for instance, a dialogue. Any dialogue features a variety of instances of multimodal coordination, entrainment, and mimicry. A dialogue can simultaneously manifest *coordination* as in the case of turn taking in a conversation (e.g., Sachs et al. 1974), or introducing a new topic, at a syntactic level (e.g., syntactic priming; Bock 1986; 1989; Bock & Loebell 1990) or at an affective level (e.g., mood contagion; Neumann & Strack 2000). Simultaneously, it is possible to see cyclically occurring instances of affective facial expressions (e.g., Dimberg et al. 2000) and breathing movements (e.g., Furuyama et al. 2005). Coordination and entrainment can converge when joint behavior is goal driven (e.g., playing tennis versus choral singing): It can be consciously accessible or escape conscious access (two people moving a heavy object versus emotional contagion), or a combination of both.

If the aim of the SCM is to fully understand the bases of emergent social cognition and behavior, then it has to incorporate a level of analysis of interacting dyads and beyond.

## Goals are not implied by actions, but inferred from actions and contexts

doi: 10.1017/S0140525X07003305

Iris van Rooij, Willem Haselager, and Harold Bekkering

Nijmegen Institute for Cognition and Information, Radboud University Nijmegen, 6500 HE Nijmegen, The Netherlands.

i.vanrooij@nici.ru.nl w.haselager@nici.ru.nl

h.bekkering@nici.ru.nl

**Abstract:** People cannot understand intentions behind observed actions by direct simulation, because goal inference is highly context dependent. Context dependency is a major source of computational intractability in traditional information-processing models. An embodied embedded view of cognition may be able to overcome this problem, but then the problem needs recognition and explication within the context of the new, layered cognitive architecture.

Susan Hurley proposes a layered cognitive architecture to model, among other things, the human capacity for understanding people's actions. We applaud the effort because we believe cognitive science can benefit from pursuing alternatives to the traditional cognitive-sandwich account, especially when it comes to higher cognition (Haselager et al. 2003; van Rooij et al. 2002). We do see one potential problem with Hurley's conception of how layers 3 and 4 of the shared circuits model (SCM) implement our ability to understand the goals that drive people's actions.

According to the SCM, people understand why people act by "mirroring" the "means/ends structure of observed actions" (sect. 4, para. 5 [*layer 3*]). From reading the target article, it is less than clear what mechanism underlies the activity of mirroring, but Hurley seems to have in mind a non-inferential mechanism in which goals and actions are directly coupled. According to Hurley, this is made possible by the fact that humans can reverse the direction of the goal – action associations generated by their own goal-directed actions. As a result, Hurley argues, "observing movements generates motor signals in the observer that tend to cause similar movements" (sect. 4, para. 5 [*layer 3*]). When the motor outputs are inhibited to prevent overt copying, then the system is able to engage in a form of "mirroring [that] simulates in the observer the causes of observed action" (sect. 3.4, para. 5, *layer 4* of the SCM).

This conception of inferred goals and their relationship to observed actions is not unproblematic. It seems implausible that a simple one-to-one association between action and goal can account for the intelligent ways in which humans infer goals from observed actions. Research shows that the goals that people infer depend in complex ways on the context in which the actions are observed. For example, the action "pushing a button with one's head" can suggest the goal "that the button be pushed" (e.g., when the person's hands are occupied holding a towel), or the goal "that the button be pushed with the head" (when the hands are free to do the pushing as well). Even infants are sensitive to such contextual factors, leading them to push the button with their hands after seeing an adult push it with her head while holding a towel in her hands, but pushing the button with their heads when the adult's hands were free during the action (Gergely et al. 2002). These observations underscore the problematic nature of Hurley's idea that "observing movements generates motor signals in the observer that tend to cause similar movements" (sect. 4, para. 5 [*layer 3*]). From the perspective of motor plans, after all, pushing a

button with the hand is very dissimilar from pushing it with the head, yet infants will “copy” observed actions of adults in dissimilar ways if appropriate given the context.

Two defenses of the SCM could be formulated at this point: First, one could propose that the action-goal associations in the SCM are not necessarily one to one. That is, multiple goals could become associated with one and the same action (e.g., picking up a pen could be associated with writing, pointing, giving, etc.), and multiple actions could become associated with one and the same goal (the goal to go to work can be associated with walking, biking, driving, etc.). By “mirroring” one could then retrieve multiple (hypothetical) goals for any given observed action. Although it is conceivable that our brains build complexes of action-goal associations, the question remains how it selects which of the – potentially very many – possible goals is the most plausible or likely goal in the current context. It is known that context sensitivity of such abductive inferences can lead traditional information-processing models into the problem of computational intractability, be they logicist (Bylander et al. 1991), connectionist (Thagard 2000), or Bayesian models (Cooper 1990). It remains a challenge for the SCM, or other layered architectures, to incorporate abductive inference processes that can circumvent this classical intractability problem (e.g., Cuijpers et al. 2006).

Second, one could argue that from the perspective of the observer, two actions do not constitute one and the same observed action if the context of the actions differs. The argument could go as follows: the notion of “observed action” is to be understood to include relevant parts of the context (in our foregoing example: whether the hands are occupied or not); then a unique mapping from action-context pairs to goals can possibly be achieved by a mere “mirroring.” Note, however, that such a proposal serves only to move the problem from understanding the role of context in goal inference to the problem of understanding how people decide which aspects of the context are relevant parts of the current action. This is one of the many disguises in which the infamous frame-problem shows itself (Ford & Pylyshyn 1996; Haselager 1997; Pylyshyn 1987): Figuring out the proper demarcation of what constitutes an “action” is computationally no less challenging than finding the most likely goal in a set of possible goals.

By claiming that goal understanding involves in part an inferential process, we do not mean to suggest that the process is necessarily conscious, controlled, or reasoned in any way. The mechanism can be highly automatic, unconscious, and even build on associative principles. Its implementation may involve the so-called mirror neuron system (Newman-Norlund et al. 2007), but it may also draw upon different neural systems, depending on the nature or complexity of the inferential task (e.g., de Lange et al., submitted). We see it as a challenge for future research to reconcile functional, computational, and neural explanations of goal inference in a way that explains how people can effectively and efficiently make plausible inferences about other people’s goals and intentions in contexts of real-world complexity. So far, traditional information-processing models have failed in this pursuit, due to the apparently insurmountable problem of computational intractability. This is not the place for a full sketch of our views, but we would like to suggest that an embodied embedded view of cognition may prove useful in addressing this problem. First of all, Hurley’s layered (rather than “sandwiched”) view of the cognitive architecture may invite an alternative, nontraditional conception of the inferential task posed to the brain (e.g., van Dijk et al., in press). Secondly, properties of world and body can serve as cognitive resources that may reduce the computational complexity of the inferential task (van Rooij & Wareham, in press).

In sum, Hurley’s model is to be welcomed as a nontraditional model of action understanding, but the mechanisms behind layers 3 and 4 need clarification in view of the computational problems they are supposed to be solving. Embodiment and

embeddedness may help to provide clues for such clarification, although currently this is more a way to formulate the challenge than to answer it.

## Imitation, emulation, and the transmission of culture

doi: 10.1017/S0140525X07003317

Andrew Whiten

Centre for Social Learning and Cognitive Evolution, and Scottish Primate Research Group, School of Psychology, University of St Andrews, St Andrews KY16 9JP, United Kingdom.

aw2@st-andrews.ac.uk <http://psy.st-andrews.ac.uk/people/lect/aw2>

**Abstract:** Three related issues are addressed. First, Hurley treats emulation and imitation as a straightforward dichotomy with emulation emerging first. Recent conceptual analyses and “ghost” chimpanzee experiments challenge this. Second, other recent chimpanzee experiments reveal high-fidelity social transmission, questioning whether copying fidelity is the brake on cumulative culture. Finally, other cognitive processes such as pretence need to be integrated.

Susan Hurley’s target article is an extraordinary achievement. I cannot judge her contribution from the perspective of philosophers, but she has done those of us working on the empirical sciences side of the topics she reviews a great service by pulling together potentially mutually relevant discoveries and developing specific models of how they might form evolutionary or developmental cascades. Having said this, my three sets of comments are rather along the lines of: “here are yet more empirical findings with which any modelling needs to be consistent.”

**1. Imitation and emulation.** Hurley presents imitation versus emulation as a clear and straightforward dichotomy, with imitation emerging at a higher level in the shared circuits model (SCM). However, recent analyses in animal social learning studies have dissected emulation into several different categories, some of which overlap with imitation under the general heading of “copying” (Whiten 2006). For example, when a chimpanzee matches whichever of two forms of tool use they see (Whiten et al. 2005a), some level of copying is implicated, and even if the matching to the model is in how the tool moves rather than the bodily actions involved, it is not so clear why this should not be regarded as a form of imitation (copying what the model does with their tool) rather than emulation (copying only the model’s effects in the world – how the tool operates). Turning to the empirical evidence, I would argue there is far less direct evidence for emulation in nonhuman animals than Hurley implies. Rather, emulation is often ascribed to an animal as a default explanation when the animal evidences social learning that goes beyond mere stimulus enhancement, but shows little or no imitative fidelity. Positive evidence for emulation is harder to come by. If one sees emulation as learning specifically about the environmental results of actions alone, then one experimental design recently developed becomes particularly significant in this regard. This is the “ghost” experiment in which we discover what observers learn from seeing only an action’s environmental effects, which is what an emulation hypothesis suggests they are learning about. In a recent experiment of this kind we obtained the striking result that chimpanzees failed to learn the complex tool technique involved (Hopper et al. 2007), despite having demonstrated earlier that they would copy the particular one of two such techniques they see performed by a chimpanzee (Whiten et al. 2005a). In short, for learning the most complex technologies they can handle, chimpanzees seem to need to see another chimpanzee perform the act. This finding and the handful of other similar “ghost” studies completed with children and nonhuman species (for a review, see Hopper