Bootstrapping the mind

Citation for published version:
Kiverstein, J & Clark, A 2008, 'Bootstrapping the mind' Behavioral and Brain Sciences, vol. 31, no. 01, pp. 41-58. DOI: 10.1017/S0140525X07003330

Digital Object Identifier (DOI):
10.1017/S0140525X07003330

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Behavioral and Brain Sciences

Publisher Rights Statement:

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Even at a liberal threshold, there was no greater activity for the alternative action compared to the imitation condition. This demonstrated no evidence of inhibition. In contrast, there was much greater brain activity during imitation. Clusters of activity were identified in rostral anterior cingulate cortex and lateral orbitofrontal cortex. These findings can be understood in terms of current models of motor control (e.g., Wolpert et al., 2001). Imitation is another motor skill which relies on actively developing motor control in an incremental fashion. The effects of any action plan being executed are experienced as consequences for sensory feedback, which serve to modify motor planning functions (cf. the target article). The role of mirror neurons within our model (Williams et al., 2007) is to respond to the fidelity of the enacted action compared to the perceived action. Error detection results in conflict-related activity in rostral anterior cingulate and a drive to alter behaviour emanating from lateral orbitofrontal cortex. Imitation is therefore an active process of comparison between self and other, requiring continuous modification of action planning with an aim of achieving greater fidelity between self and others. Imitation is intentional, operates within a social context, draws on a capacity for social judgment, and involves new learning.

One form of pathological “automatic” imitation occurs as echopraxia, described in people with autism or frontal lobe lesions (Lhermitte et al., 1986) or those institutionalized with echopraxia, described in people with autism or frontal lobe and involves new learning.

Within a social context, drawing on a capacity for social judgment, and involves new learning.

One form of pathological “automatic” imitation occurs as echopraxia, described in people with autism or frontal lobe lesions (Lhermitte et al., 1986) or those institutionalized with echopraxia, described in people with autism or frontal lobe and involves new learning. Such individuals may have increased suggestibility, and impaired capacity for social judgment and flexible rule learning. Williams et al. (2004) suggested that echopraxia in children may reflect delayed rather than deviant learning of imitation skills, because echopraxia places lower demands on new learning.

Finally, to return to the central theme of Hurley’s article, what are the implications of this for the relationship between imitation and mentalizing? The demands placed by self-other matching on our model (Williams et al., 2007) is to respond to the fidelity of the shared circuits model (SCM) has elicited. One special source of delight would have been the sheer interdisciplinary diversity of the responses. For Susan believed very strongly that a proper understanding of minds, persons, and reasons would emerge only from tough, cooperative, interdisciplinary work drawing on psychology, philosophy, neuroscience, social science, and cognitive science. One measure of the success of SCM is thus its capacity to stir that larger scientific pot. In that, the treatment (as we see) succeeds wonderfully. Moreover, there seems to be significant agreement concerning many of the finer details of the story. In our response, we try to do three things: First, we briefly clarify the nature of the story on offer; second, we highlight (and where possible respond to) the main critical issues raised by the commentators; and third, we showcase the exciting range of new suggestions (and additional mechanisms) that the commentary phase has uncovered. In suggesting the responses that follow, we are acutely aware of our own shortcomings as surrogate respondents. Some of the issues raised simply exceeded our grasp of the subject area, the target material, or both. Where this has arisen, we have simply remained silent, and beg the readers’ (and the commentators’) forbearance.

ACKNOWLEDGMENT
I am grateful to Nina Williams for comments upon an earlier draft.

Response

Bootstrapping the mind

doi: 10.1017/S0140525X07003330

Julian Kiverstein and Andy Clark

School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Edinburgh, EH8 9JX, Scotland, United Kingdom.

j.kiverstein@ed.ac.uk andy.clark@ed.ac.uk

Abstract: After offering a brief account of how we understand the shared circuits model (SCM), we divide our response into four sections. First, in section R1, we assess to what extent SCM is committed to an account of the ontogeny and phylogeny of shared circuits. In section R2, we examine doubts raised by several commentators as to whether SCM might be expanded so as to accommodate the mirroring of emotions, sensations, and intransitive actions more generally. Section R3 responds to various criticisms that relate to the account of social-learning Hurley proposes in the target article. We conclude in section R4 by responding to a number of commentators who argued for the limitation of control theory as a framework for studying social cognition.

Introduction: Stirring the science

“Shared circuits” was Susan Hurley’s last grand project. It set the agenda that might, in another possible world, have allowed Susan and her commentators to begin to converge on a unified and integrated understanding of something quite fundamental to human thought and reason: our ability to know the minds of others, and of ourselves. Susan’s goal was to show one possible way in which a shared basic information space for perception and action might be bootstrapped, function by function, into a grip on self and other, thus providing the essential entrance ticket to the rich realm of social cognition. This project would close the circle, showing that the issues concerning embodiment and dynamics (Hurley 1998) were never that far away from those concerning social cognition, policy, and the possibility of responsibility (Hurley 1989; 2003; 2006a).

Susan died before she could see this project into print. But had she been able to do so, she would have been truly delighted by the wide array of thoughtful, challenging, and constructive commentaries that her shared circuits model (SCM) has elicited. One special source of delight would have been the sheer interdisciplinary diversity of the responses. For Susan believed very strongly that a proper understanding of minds, persons, and reasons would emerge only from tough, cooperative, interdisciplinary work drawing on psychology, philosophy, neuroscience, social science, and cognitive science. One measure of the success of SCM is thus its capacity to stir that larger scientific pot. In that, the treatment (as we see) succeeds wonderfully. Moreover, there seems to be significant agreement concerning many of the finer details of the story. In our response, we try to do three things: First, we briefly clarify the nature of the story on offer; second, we highlight (and where possible respond to) the main critical issues raised by the commentators; and third, we showcase the exciting range of new suggestions (and additional mechanisms) that the commentary phase has uncovered. In suggesting the responses that follow, we are acutely aware of our own shortcomings as surrogate respondents. Some of the issues raised simply exceeded our grasp of the subject area, the target material, or both. Where this has arisen, we have simply remained silent, and beg the readers’ (and the commentators’) forbearance.

R1. The nature of the beast

One of the challenges facing an embodied and situated approach to cognitive science is to map out a path that might have taken humans from the basic kinds of capacities for on-line adaptive response we share with robots and nonhuman animals to distinctively human cognitive abilities, such as the capacity for rational deliberation and the ability to make sense of the purposeful behaviour of other agents. The shared circuits model...
(SCM) contributes to meeting this challenge. It describes a set of mechanisms that might have taken humans and their evolutionary ancestors from active perception to imitation, mindreading, and deliberative, strategic thinking.

Each of the model’s five layers is given a functional description, which deliberately abstracts away from details of neural implementation. This is not to say that the model is silent about implementation issues – it predicts a common coding for perception and action implemented by the mirror system, for instance (see sect. 2.2 of the target article). However, it leaves the details of how each layer might be neurally implemented as open questions for further investigation. At least one commentator (Preston) took the lack of detail at the neuroanatomical/functional level to be a weakness of the model. Notice, however, that any explanation of how the layers of the model are implemented in the brain will itself most likely be a functional explanation. Preston concedes as much in referring to a neuroanatomical/functional level of description. Consider, for instance, a putative explanation of how layer 3 could be implemented in the mirror system. Such an explanation will identify a widely distributed neural system that includes amongst other regions the temporal lobe, the rostral inferior parietal lobule, and the ventral premotor cortex. In virtue of what do these separate neural regions form parts of a single neural system that implements a mirror system? Arguably, disparate neural regions form a part of a single distributed system because of what they do – because the cells at these regions have activation profiles that make a contribution to realising a particular task. It is true that SCM doesn’t tell us which parts of the brain coalesce to form the different layers of the model, but it does purport to describe what sub-tasks these parts of the brain must perform if they are to contribute to realising a capacity for action understanding.

In this spirit, we propose interpreting SCM as having two objectives. First, it offers a task-level description of action understanding. Second, it identifies possible mechanisms that could do the work of accomplishing each of these tasks. Action understanding is a multi-faceted capacity, including the abilities to learn novel behaviours by copying, to predict and explain other’s behaviour, and to think strategically about one’s social interactions. Hurley decomposes each of these complex capacities into more basic sub-tasks. The different layers of SCM describe possible mechanisms that could perform these sub-tasks.

Consider SCM’s layer 3 as an example. Layer 3 identifies a mechanism for mirroring – the kind of behavioural priming that occurs in us when we observe an action performed by another person. Our observing the other’s action makes us more likely to perform an action of the same type, ourselves. Hurley argues that our capacity for action understanding is facilitated by this tendency to copy behaviours. As the behaviours we can copy become more complex, so also does the repertoire of actions we can potentially understand. Copying behaviour is therefore a sub-task one has to be capable of performing if one is to get into the business of understanding the goal-directed behaviour of others.

Now consider the mechanism SCM introduces to explain an animal’s ability to copy behaviour. The mechanism is not new to layer 3 but has already been introduced at the previous layer to explain amongst other things the motor system’s execution of fast, fluent sensorimotor behaviour. Thus, an explanation is being given of one aspect of action understanding in terms of the same mechanisms used to control sensorimotor behaviour. Often, sensorimotor behaviour will require sensory feedback to be made available faster than the sensory systems can supply. A way around this problem would be for the motor system to employ its learned associations between motor outputs and the sensory consequences of those outputs to make predictions about what will happen when a given motor command is executed. In control theory, these predictive models are called forward models. At layer 3, forward models are run in reverse, so that a system can, instead of predicting the sensory consequence of an action, work out from an observed action the motor commands that were the cause of this action. Running a forward model in reverse thus produces a motor command. Whether the system then carries out this motor command is a further question. Most of the time, the execution of motor commands arrived at exogenously from observing the actions of others will be inhibited. This makes good evolutionary sense, as Makino (and indeed Hurley) notes: A creature that copied the behaviour of an approaching predator would not be long of this world.

How does this mechanism of running forward models in reverse explain an animal’s ability to copy behaviour? When a forward model is run in reverse, the outcome of this process will be the production of a motor command. This explains why, when we observe another acting, we are automatically primed to perform the same or similar action ourselves. We have this standing disposition because running a forward model backwards produces in us the same motor plan that initiated the action in the other person.

We can also see now what Hurley means when she claims that a shared information space for perception and action can also function as a shared information space for self and other. Perception and action share a common information space in part because perceiving another agent acting causes our motor system to produce the same or similar motor commands that were the causes of that agent’s actions. This information space can also be deployed in action understanding. Say I perceive another person reach for and answer their ringing handset. This motor plan is now available to be used by other subpersonal systems (layer 4 of SCM) to make sense of the other person’s action and its causes. However, the information my motor system makes available doesn’t distinguish between motor plans that are my own and have been initiated endogenously, and the motor plans of another person that have been initiated in me exogenously. In this sense, the information space that perception and action share is also an information space that self and other share. This sharing of information makes possible a kind of direct, non-inferential understanding of the action of others. There is no difficulty at this level of processing about how we acquire access to information about the causes of another person’s actions. At this level of processing, the
information that is used to make sense of the actions of others is *intersubjective* – it doesn’t distinguish between self and other.

So far, we have proposed a way of thinking about each of SCM’s layers, but of course a good deal of the model’s explanatory work is accomplished by exploring possible interactions between the layers. We have already seen something of how one layer can borrow mechanisms from an earlier layer and expat this mechanism for a new function. Hurley also describes some ways in which layers might function together to achieve complex tasks that each layer cannot accomplish on its own. Thus, the forward models of layer 2 can be combined with a mechanism for monitored output inhibition introduced at layer 4 with the result that a system can begin to model possible courses of action and assess the consequences of these actions in advance of executing them. A system with layers 2 and 4 can begin to engage in trial-and-error learning “in the head.” By combining the mirroring functions made possible by layer 3 with the monitored output inhibition of layer 4, we get a system that can distinguish actions that are its own from actions that belong to another. One difference between motor plans that are endogenously produced and those that are exogenously produced by observing the action of another is that the latter tend to be inhibited. By monitoring inhibited output, a system could thus acquire information that could be used to distinguish its own endogenously produced motor plans from motor plans it finds itself with as a consequence of layer 3. Moreover, such a system could use the information made available by layer 3 to begin to make sense of the other’s behaviour.

Depending on the systems own repertoire of behaviour and the associations between means and ends that fuel its forward models, a system combining layers 3 and 4 could use its learned associations to understand the means/end structure of the other’s behaviour. Layer 5 introduces a capacity for the monitored simulation of inputs. This can be combined with layers 3 and 4 to generate information about the possible actions of others and the causes and effects of such possible actions. A system with this combination of layers can begin to engage in strategic thinking and game-theoretic deliberation about the action of others. (For more details on the role of mindreading in strategic social intelligence, see Hurley 2005a.)

### R2. Linking layers

Many of Hurley’s commentators raised questions that bear on the interaction among layers we have sketched, and it is to these questions that we now turn. Chakrabarti & Baron-Cohen’s commentary raises some challenging questions about the developmental progression between layers. Oberman & Ramachandran also wonder how shared circuits might develop, but they raise questions about both the ontogeny and the phylogeny of shared circuits. Are shared circuits hard-wired, learned, or a combination of both?

Hurley states that she doesn’t take SCM to imply a single account of the development of capacities for imitation, mindreading, and deliberative thinking. The numbering of the layers, she writes, “does not necessarily represent the order of evolution or development” (sect. 4, para. 9). Rather, the model is intended to provoke hypotheses that map the layers onto “specific phylogenetic or ontogenetic progressions” (sect. 4, para. 9). In other words, Hurley is not committed to a particular answer about how the different layers might feed into a story about the development of mindreading capacities. Nor does she take a firm stand on the question of whether a capacity for imitative learning is culturally acquired or forms a part of our innate inheritance. Hurley believed that questions of this kind would be settled through close collaboration between science and philosophy. Hence, she would have very much welcomed Oberman & Ramachandran’s constructive suggestions about possible experiments that might answer the nature/nurture question as it arises for mirroring.

This is not to say that Hurley had nothing to say either in her target article or elsewhere about these questions. Indeed, she makes a concrete proposal about how mirroring might have arisen. She begins by telling a Hebb-inspired story about how cells might come to fire both for others’ actions that are observed and for actions of one’s own that are executed. (The story doesn’t originate with Hurley, but can also be found in Goldman [2006, Ch. 6] Heyes [2002; 2005], and Keysers & Perrett [2004]) Suppose the action is one of grasping. Superior temporal sulcus (STS) neurons that respond to observations of grasping behaviour might overlap in time with activity in areas (e.g., PF and F5) that are involved in initiating the grasping behaviour. As a result of Hebbian learning, the connections between STS and the motor areas will be reinforced. The effect of this reinforcement will be that cells in motor areas will fire both when the agent observes his own movements and when he observes the movements of others. Clearly, this sort of account is going to work only for movements that the agent can see himself performing. In order to account for the copying of facial expressions – we can assume that an agent will often be able to copy many facial expression without being able to see his own face – Hurley introduces a number of different factors. She concedes that there could be a role for innate supramodal correspondence between observed acts and an observer’s similar acts, of the kind suggested by Meltzoff and Moore’s (1997) active intermodal mapping (AIM) hypothesis. Hurley also considers a number of other possible explanations for mirroring when one cannot observe one’s own behaviour, including one in terms of stimulus enhancement (sect. 3.3, para. 7, 8). We won’t repeat the hypothesis. Suffice it to say that Hurley saw a role both for learning and for innate capacities in explaining the emergence of mirroring.

Could a creature understand the action of others but lack a capacity for mirroring? Conversely, could a creature have an unimpaired capacity for mirroring but be incapable of making sense of the behaviour of others? Answering these questions promises to have ramifications for how we think about the relation between mirroring and layers 4 and 5 that do the work of explaining mindreading. Chakrabarti & Baron-Cohen suggest that psychopaths may have intact mindreading abilities but deficits in affective empathy. Affective empathy is arguably explained by the mirroring capacities introduced at layer 3, more on which in a moment. Thus, psychopaths may present a case in which we have intact layers 4 and 5 but a
compromised layer 3. Subjects with autism spectrum disorders (ASDs), on the other hand, exhibit impairments in mindreading and affective empathy. Perhaps they provide an example of what can go wrong when layers 3, 4, and 5 are compromised. Chakrabarti & Baron-Cohen present these two cases as an example of double dissociation of mindreading and affective empathy capacities. However, subjects with ASDs do not display the opposite profile to that of psychopaths: they do not have intact capacities for affective empathy but impaired mindreading skills. Baron-Cohen and Wheelwright (2004), for instance, found that subjects with Asperger Syndrome scored significantly lower than normals in a questionnaire testing for empathic skills. Psychopathy certainly establishes the possibility of mindreading without affective empathy, but ASD does not, as far as we can see, establish the possibility of affective empathy without mindreading. This casts doubt on the suggestions that we here confront a clear double dissociation. But leaving this issue to one side, we want to focus on Chakrabarti & Baron-Cohen’s interesting claim that this body of evidence challenges the claim that layers 3 and 4 are required for layer 5.

The first response we would make in Hurley’s defence is that SCM makes no hypotheses about the development of mindreading abilities, and in particular it does not explicitly claim that layers 3 and 4 are necessary for the emergence of mindreading abilities at layer 5. We have just seen how Hurley left an open question how the layers of SCM relate to the development and acquisition of mindreading abilities. Nevertheless, it is true that Hurley does offer an explanation as to how a capacity for mindreading might get started, and perhaps it is this story that Chakrabarti & Baron-Cohen mean to dispute. We will first consider whether the explanation Hurley offers commits her to the claim that layers 3 and 4 are required for mindreading abilities. Second, we will assess whether the latter hypothesis is really challenged by the disorders discussed by Chakrabarti & Baron-Cohen.

Is Hurley committed to the claim that layers 3 and 4 are necessary if a person is to acquire the sorts of mindreading skills made possible by layer 5? We have already explained how Hurley took mindreading to begin at layer 4. Prior to layer 4, the information a creature has available for making sense of the behaviour of others does not distinguish between self and other. This information can be used to (implicitly) recognise and identify agents that behave in ways similar to me. This recognition of the fundamental similarity between self and other forms the basis for empathy. However, mindreading – the interpretation and prediction of other’s actions – begins with the acquisition of a grasp of the self/other distinction. Once this distinction is understood, a creature can begin to attribute mental states to the other – to interpret or “read” the other’s mind. According to SCM, one acquires an ability to distinguish self and other by acquiring a mechanism for monitoring the inhibition of mirroring. The monitoring of inhibited mirroring allows a creature to identify motor plans that are not its own. With this understanding in place, the creature can begin to populate the world with other perspectives and decentre from its own situation in the here and now to entertain other possible points of view. This capacity becomes more powerful in creatures that possess the representational capacities introduced at layer 5, and can model not just possible courses of action but can in addition model possible mappings from sensory input to motor output. At first glance, then, it would seem correct to attribute to Hurley the hypothesis that layers 3 and 4 are required for layer 5. Layer 4 looks to be required for layer 5 since the former supplies models of the outputs, which are used at layer 5 in the simulations of complete mappings from inputs to outputs. Layer 3 seems to be required by layer 4 since it makes available the bi-directional simulations that are taken off-line at layer 4. Thus, we might conclude on these grounds that Hurley is indeed committed to the hypothesis attacked by Chakrabarti & Baron-Cohen.

Although there are strong grounds for attributing this hypothesis to Hurley, it doesn’t seem to us to be strictly entailed by SCM. SCM suggests an explanation of how an animal might come to be able to distinguish itself from others. If we accept the idea that mirroring provides information that does not differentiate self from other, some such explanation is required. However, the cases discussed by Chakrabarti & Baron-Cohen involve individuals whose capacity for mirroring is impaired. Such individuals will not need layer 4 to distinguish themselves from others, since they do not have information at their disposal for which the self/other distinction fails. They precisely do not identify with others empathically, nor do they recognise themselves to be similar to others. Although they do not need layer 4 to distinguish themselves from others, layer 4 could nevertheless continue to work in conjunction with layer 2 to provide information about alternative possible courses of action. Layer 4 could continue to supply simulations of possible actions to layer 5. Therefore, it doesn’t seem out of the question that a system could exhibit the sort of mindreading abilities made possible by layers 4 and 5 despite having an impaired layer 3.

Suppose we nevertheless concede that a fully intact layer 3 is required for layers 4 and 5. Could an individual with an impaired layer 3 nevertheless exhibit intact mindreading abilities? Possibly. Hurley concedes that even a creature equipped with the sorts of sophisticated representational capacities ushered in by layer 5 might not have what it takes for full-fledged mature mindreading. Mature mindreaders can track many different agents, identifying them in a wide range of different situations. Hurley suggests that language might well be required for an “understanding of multiple others with multiple alternatives and varying beliefs” (sect. 3.5, para. 2). Suppose a person could acquire mastery of a language without the use of layer 3 (a possibility challenged by the claim that imitation is required for the acquisition of language; but for a defence of this hypothesis, see, e.g., Arbib & Rizzolatti 1997; Iacoboni 2005). It would then be possible for such a person to exhibit high-level mindreading skills despite lacking a capacity for mirroring. Perhaps such an individual could acquire a theory of mind by learning generalisations relating behaviour, the environment, and mental states in much the same way as scientists generate theories based on observations (for an account of mindreading abilities along these lines, see, e.g., Gopnik & Wellman 1992, 1994). Gallagher (2005) suggests that a high-functioning autistic like Temple-Grandin might deploy exactly this type of theorising strategy to understand the intentions and emotions of others. Gallagher writes of Temple-Grandin that she “reads
about people, and observes them, in an attempt to arrive at the various principles that would explain and predict their actions in what she describes as “a strictly logical process” (Gallagher 2005, p. 236). We can imagine a psychopath acquiring mindreading abilities in much the same way. We conclude, then, that even if we were to suppose that SCM entails the hypothesis that layer 3 is required for layers 4 and 5 (something we are inclined to dispute), SCM can still handle the case of psychopathy.

What about subjects with ASD? These subjects have difficulties taking up perspectives that are not their own, and there is some evidence that this leads to difficulties in imitating (for a balanced assessment of this evidence, see Goldman 2005). Hobson and Lee (1999), for instance, showed that autistic subjects failed to copy behaviours that required perspective switching. In one task, the experimenter took a wooden pipe rack in his left hand and held it against the upper part of his left shoulder. With his right hand, the experimenter took a wooden stick and strummed across the ridges and slots of the pipe rack three times, making a staccato sound. Of the 16 autistic subjects, 15 ran the stick over the pipe rack, but only 2 of the 16 held the pipe rack against their shoulder as the experimenter had demonstrated. In order to copy the action, the autistic subjects had to perform the same action in relation to a different body, their own. First, they had to recognise the relation between the action and the experimenter’s body. This required them to switch from their own perspective to adopt that of the experimenter. Having recognised the relation of the action to the experimenter’s body, they then had to switch back and re-enact this relation from their own perspective. SCM would predict that subjects with impaired mirroring and mindreading abilities would find this sort of perspective switching difficult. Subjects with impaired mirroring abilities will not identify with and recognise others as similar to themselves. Furthermore, when layers 4 and 5 are damaged, subjects will find it difficult to detach from their own perspective. This is exactly what we find in autistic subjects. We conclude that the deficits we find in subjects with ASD may also be consistent with SCM.

Hurley tells us that the non-negotiable parts of the model concern (1) the explanation of mirroring as “an expactive reversal of online prediction” and (2) “the way the actual/possible and self/other distinctions arise as online processes are overlain by monitored inhibition” (sect. 4, para. 9). We consider next the commentaries that challenge each of these claims beginning with the first of SCM’s non-negotiable claims. Whereas the set of issues we have just considered relate to the interaction between layers, the commentaries to which we now turn question the use that is made of forward models and sensory feedback introduced at layers 1 and 2 to explain the capacities for mindreading that come on the scene with layers 3 to 5.

Goldman worries about the attempt to explain the mirroring of emotion, pain, and other sensations by appeal to lower-level mechanisms of adaptive feedback control and forward models introduced at layers 1 and 2. He challenges a core claim of SCM that there are systematic relationships between the mechanisms used in the control of sensorimotor behaviour and those that underpin our mindreading abilities. Similar concerns are voiced in the commentaries of Heyes, Preston, and Chakrabarti & Baron-Cohen. Heyes argues that the account of mirroring at layer 3 may not be readily applied to intransitive actions like facial expressions and gesture. Yet she points out that much of the evidence for the mirror system in humans comes from the copying of intransitive actions, rather than the instrumental actions modelled by SCM. Preston claims that there are no good reasons from either phylogeny or ontogeny to claim that control mechanisms like those found at layers 1 and 2 precede the mirroring mechanisms of layer 3. Again the worry seems to be that the appeal to control theory is inadequate when it comes to explaining the sort of mindreading involved in understanding others’ emotional experiences. Chakrabarti & Baron-Cohen wonder how SCM applies to the processing of facial expression. They suggest that the perception of emotions could recruit layers 4 and 5 to different extents, and that SCM makes no provision for such a possibility. Preston can be understood as raising a related concern when she cites evidence in support of the claim that the perception of emotional facial expressions activates semantic-level representations for specific emotions.

We suggest two lines of response. First, it should be recognised that Hurley never claimed to have identified a set of mechanisms that can account for every aspect of social cognition. SCM as it is described in the target article is offered as an account of the understanding of instrumental actions – actions that have a means-end structure. Hurley claims that mechanisms from control theory can explain this particular type of social cognition. She claims that there is a systematic relationship between the control and mirroring of instrumental actions. If it should turn out that there is no such systematic relationship between control and the mirroring of intransitive or expressive actions, this would not harm SCM, which claims only that such a relationship holds for the case of instrumental actions. It is certainly an interesting question as to whether SCM might be extended to account not just for our understanding of instrumental actions, but also for what Hurley calls “expressive actions.” Indeed, this is one of the questions Hurley raises in section 4.1.2, and we shall consider this possibility in more detail shortly. It is surely worthwhile to ask how many of our higher cognitive capacities can be explained by appeal to the same basic mechanisms we employ in sensorimotor behaviour. Natural selection often works by taking mechanisms that already exist and tinkering with them. It therefore makes good evolutionary sense to suppose that the very same mechanisms that are used to control sensorimotor behaviour might also serve a very different function in making possible mindreading and social intelligence more generally.

Goldman, Heyes, and Preston suggest, however, that a different set of mechanisms might be required to explain emotional mirroring from those described at layer 3. Consider the following example of emotional mirroring. I watch a couple arguing and I perceive the woman’s fear at her partner’s anger: I see the fear written on her face. When I see her fear, the same parts of my brain are active as when I myself feel fear. Williams et al. (2001), showed subjects Ekman faces expressing fear, and found that when fearful faces produced increases in skin conductance this response was also accompanied by increased activity in the amygdala. (Ekman faces are photographs of expressive faces used in emotion
In an unpublished review of Goldman’s (2006) appeals to explain the mirroring of instrumental actions, which builds on the sorts of control mechanisms Hurley whether an account of emotional mirroring could be given intend. Expressive actions like the facial expression of means to achieving any end. We don’t intend anything in this case, nor do we act in order to bring about what we intend. Expressive actions like the facial expression of emotions do not have a means-end structure, which is just to say that they are not instrumental actions.

By way of a second response, we want to briefly consider whether an account of emotional mirroring could be given which builds on the sorts of control mechanisms Hurley appeals to explain the mirroring of instrumental actions. In an unpublished review of Goldman’s (2006) Simulating Minds, Hurley writes that to understand how control and mirroring are related outside the context of instrumental actions and intention reading, an account must be given of “how instrumental and expressive actions are related.” She continues, “In my view, mirroring for expressive action builds on the more fundamental, control-related mirroring for instrumental action” (Hurley 2007, p. 11). She doesn’t expand on this comment, but we will try to fill out what she might have had in mind, first by contrasting her view of how simulations are involved in mirroring with that of Goldman. On the basis of this contrast one of us (Kiverstein) has developed a somewhat speculative suggestion about how Hurley might have thought about shared circuits as they arise in the context of emotional mirroring. Along the way, we will also have some things to say about how instrumental and expressive actions might be related.

Goldman and Sripada (2005) describe four accounts of emotional mirroring, and plump for what they call the unmediated resonance model (also see Goldman (2006, Ch. 6). According to this model, when we see a person’s facial expression of emotion, this directly causes in us a similar emotional state: “observation of the target’s face ‘directly’ without any mediation . . . triggers (sub-threshold) activation of the same neural substrate associated with the emotion in question” (Goldman & Sripada 2005, p. 207). We come to share the emotion that the other person displays because our observing this display causes in us an emotional experience of the same type. We can recognise the other’s emotion on this model because we come to occupy a state that resembles that of the target. Goldman and Sripada are proposing here a simulation-based account of mindreading for the emotions according to which we simulate the other person’s emotional state by instantiating a process which, when it functions properly, results in a state that resembles or matches the target’s mental state. Simulation is explained here in terms of the products (mental states and their contents) of a mental process of simulation. If this is correct, it is by first producing in ourselves a mental state that matches or resembles the mental state of the target we are seeking to understand that we become able to work out which mental state to attribute to the other.

Hurley suggests in her author meets critics review of Goldman’s book (2007), however, that there is a process/product ambiguity in Goldman’s account of simulation. One of the lines of evidence Goldman appeals to in support of his account of emotion experience is the finding that the same brain areas are active during the experience of, and the recognition of, emotions. When these areas are damaged, not only do subjects lack a capacity for a certain type of emotion, but they also have difficulties in recognising this emotion on the basis of facial expressions. This evidence suggests a similarity in the neural/functional processes that subserve emotion experience and emotion recognition within an individual. Goldman’s account of emotional mirroring takes this similarity in processes to be grounds for inferring a similarity in emotion experience between observer and target. It is this similarity or resemblance that forms the basis for the observer’s attribution of an emotion of a particular type to the target. Hurley argues that interpersonal similarity of mental state of the kind Goldman appeals to in his explanation of emotional mirroring is not sufficient for simulation. Interpersonal similarity of states does not count as simulation unless an individual reuses his own mental processes to drive his mindreading. Thus, Hurley proposes what she calls the re-use conception of simulation, according to which we come to recognise and understand what the other is feeling by using the processes that take place in us when we undergo an emotion episode of the same type. It is our re-using this same process that explains how we come to understand what the other is experiencing.

Suppose we accept that similarity of emotional state between observer and target is insufficient for mirroring, but that in addition what is required for true simulation is that the observer’s emotion recognition process must use her own emotion experiencing processes for the purpose of simulating the other. Now compare the case of emotion mirroring with mirroring of instrumental actions. The mirroring of instrumental actions is the result of learned associations between movement and the sensory consequences of movement. There will also be learned associations in the case of emotion expression. The changes in facial musculature, for instance, will have sensory consequences. If Adolphs et al. (2000) are right, there will also be somatosensory changes sometimes throughout the body that are associated with a given emotion. So just as in the case of instrumental actions, there will be associations between the execution of an expressive action and reafferent feedback.

There is, however, an important difference in the case of expressive actions that we should mention. In the case of instrumental actions, associations get set up between motor plans and visual experiences of one’s own movement, so that when an observer sees similar movements performed by others the same motor plan is evoked. (See the Hebbian account of mirroring sketched earlier and in the target article [sect. 3.3, para. 5].) However, the reafferent feedback that is available in the case of emotion expression won’t include visual feedback: We cannot see our own faces when we express an emotion (except when we use a mirror), nor do we see any of the other bodily changes that accompany an emotion experience. Thus, it would seem we run into the correspondence problem in attempting to explain emotion mirroring. The
correspondence problem is of course a quite general problem for explanations of imitation, and one for which various answers have been proposed. One approach invokes general purpose learning mechanisms (see Brass & Heyes 2005). Another introduces innate special purpose mechanisms (see Meltzoff 2002b). We have already seen how SCM suggests that the correspondence problem may be solved by a variety of different mechanisms, but here is not the place to suggest how SCM might tackle this problem as it arises for the case of emotions. We flag this as a problem to be solved in future work that develops an account of shared circuits for emotional expression.

We now have the ingredients in place to sketch a possible way in which an account of emotional mirroring might build on the connection between mirroring and control that SCM describes. We come to recognise the other’s emotion experience by re-using our own emotion-experiencing processes. In the instrumental action case, the processes for identifying the intentions that are the causes of an observed behaviour are the same processes we use to act ourselves. The same is true in the emotion case: The processes for recognising emotion are the same as (or better, they overlap considerably with) the processes that cause the expression of an emotion. Are the processes that cause the expression of an emotion the same as the processes that cause intentional action? There may be some crossover, but there will also be important differences. As Goldman, Heyes, and Preston note, there is nothing like a forward model and visual feedback in the case of emotional expression. This is why we cannot simply take layer 3 and apply it to the case of emotion mirroring. Let us, however, ignore these differences for the moment and focus on some broad similarities.

Our own emotion-experiencing processes will include motor processes that cause certain facial expressions and the sensory consequences of these motor processes. The actions that constitute the expression of an emotion will be associated with these various kinds of sensory consequence. Hurley argues that perceptual experience (see Hurley 1998) is the result of tracking relationships between sensory flows of information and motor behaviour (O’Regan & Noé 2001a) propose a similar view. A similar model would seem to be applicable to emotional expression. To experience emotion, on this model, is to track certain invariant relationships between an action that expresses the emotion (e.g., a facial expression) and the sensory consequences of this action. This tracking ability forms a part of SCM at layer 2. We can keep track of changes in flows of sensory information because we have learned to associate movements with certain sensory consequences. It is these associations that form the basis for the forward models introduced at layer 2 and that are re-used at layer 3. Our suggestion is that Hurley’s account of perceptual experience might be extended to emotion experience so that, when we undergo an emotion episode, what this involves is our tracking the relationships between an action that is the expression of this emotion and the sensory consequences of this action. When we come to recognise the other’s emotion, we do so by using the very same tracking abilities. We come to recognise the other’s emotion by re-using the same processes that form the basis for our own experiences of emotion. We take this to be one way in which emotional mirroring might re-use mechanisms introduced at layers 1 and 2 to account for emotion experience. Hence, we tentatively conclude that emotion experience doesn’t present an insurmountable problem for SCM. Rather, it presents an opportunity for the future development of the model.

Preston reports a behavioural study (Preston & Stansfield, in press) she interprets as challenging the sort of account of emotion experience we have just sketched. The findings of the study were that perception of an emotional expression not only results in mirroring but also, as Preston writes, “rapidly activates the semantic-level representation for the specific emotion.” Presumably, by “semantic-level” representation, she means that subjects can identify and recognise the emotion. However, on the simulation-based account of mirroring, this is predicted. The idea is that we use the same processes to recognise emotions that we use to experience emotions.

We turn now to Hurley’s account of how the self/other and actual/possible distinctions arise out of monitored inhibition, which Hurley describes as the second non-negotiable feature of SCM. Preston claims that it might not be monitoring of inhibited output that generates an understanding of the distinction between self and other, and suggests that there are many other mechanisms that might do this work. She does not, however, say exactly what she has in mind. Furey & Keenan pursue a similar worry and ask whether an account might be given of an understanding of the self/other distinction in terms of forward models and the work they do in distinguishing self-caused actions from externally generated actions. Furey & Keenan discuss the case of auditory verbal hallucination when subjects claim to hear voices in their heads. They suggest that the misattribution in this case might be explained by a malfunctioning forward model which doesn’t perform its normal function of enabling the subject to distinguish self-generated inner speech from externally generated speech. We find this suggestion very plausible. Furthermore, Furey & Keenan’s account of emotional mirroring might re-use mechanisms introduced at layers 1 and 2 to account for emotion experience.
these resources could help a mirroring system to differentiate information that relates to the self from information relating to others.

Northoff, meanwhile, asks some hard questions regarding the exact form of the neural coding implicated by Hurley, and suggests that the coding needs to capture the relations between different stimuli and between stimuli and motor actions. It seems to us that Hurley would agree, and that relational coding (insofar as we understand this notion) is indeed apt for many of the purposes of the SCM. Whether Hurley’s story further demands, as Northoff suggests, a radically new conception of the self so as to reflect these relational elements, is a matter we cannot resolve definitively. But elsewhere Hurley speaks intriguingly of the self as a “dynamic singularity” itself created out of a system of relations (see Hurley 1998, pp. 206–207).

Makino argues that the self/other distinction arises not when the motor system monitors its inhibited outputs but instead from a monitoring of failed actions. He raises some interesting and important questions about how the motor system works out which outputs to inhibit, and he suggests as an answer that the motor system will tend to inhibit output in cases when it is operating only with partial information. This may be one way of making this sort of decision, but it doesn’t seem obvious to us that all inhibited actions are ones that would fail were they to be performed. Furthermore, it wasn’t entirely clear to us how to understand the suggestion of monitoring failed actions. If this consists in monitoring actions that the system fails to perform, then it strikes us that this is just another way of talking about monitoring of inhibited input. We understand inhibition as the default principle that the motor system operates on, because inhibiting actions that the system has a tendency to copy is adaptive. This default is overridden in some cases, and the decision as to when this happens will be in the hands of executive systems in the brain. We will return to this related issue shortly in our discussion of the commentaries by Behrendt and Williams.

Hove provides some nice examples of what he calls “interpersonal synchrony” in which monitoring of inhibited output might fail to generate an understanding of the difference between self and other. In the sorts of cases he has in mind, our actions are synchronised with those of another person. Hove asks how we distinguish our own actions from those of the other in these sorts of cases. There is no motor output inhibition in the cases Hove describes, so it doesn’t look like we can appeal to this mechanism to solve the problem. Hove certainly raises an interesting problem here, but it seems to us that the sorts of cases his puzzle arises for are not the ones layer 4 is introduced to explain. His examples of interpersonal synchrony do not involve the copying of behaviour, so they do not meet Hurley’s definition of mirroring, where mirroring is the process that occurs when observing a behaviour primes the observer to perform the same behaviour himself. We turn next to questions relating to imitation.

R3. Imitation and mirroring

Williams and Behrendt both discuss the question of whether mirroring is an automatic process, but arrive at opposite answers. Behrendt asks how predictive simulation, which he takes to require consciousness, can be related to mirroring that happens automatically. It is not clear why Behrendt thinks predictive simulation must be conscious. Hurley takes layer 2 of her model, which is the layer at which predictive simulation occurs, to describe a subpersonal mechanism. Forward models do produce simulations that may often involve motor imagery; however, this imagery is not always conscious. Behrendt notes how movement plans can be formed automatically upon perception of a salient event. When this event is an observed action, we have just the sort of mirroring that layer 3 is introduced to explain. Behrendt points out that traits and stereotypes can automatically elicit patterns of behaviour, but again it is just this sort of case that SCM’s layer 3 was introduced to explain. Behrendt also points out that our motivation for copying in these cases may often be social approval. Nielsen provides some experimental results that support this idea. He suggests that 2-year-old infants will commonly be motivated to copy behaviour because they want to share an experience with the other, and he describes experiments in which 2-year-olds were less inclined to imitate when not engaged in social interaction. We think Hurley would have found these results extremely interesting, but that claims about what motivates us to imitate lie somewhat out of the purview of SCM. It is an important and interesting question as to just why we have a tendency to imitate, and it is a striking finding that this tendency changes as we develop. SCM, however, seeks to show that there is a systematic relationship between the control mechanisms we use in sensorimotor behaviour and the understanding of instrumental actions. It is a further question as to what reasons we have for imitating when we do so.

Williams attributes to Hurley the claim that imitation is automatic. He goes on to make a compelling case for the view that imitation is an “intentional,” “effortful,” and “selective” process. Williams describes how imitation requires continuous modifying of action plans with the goal of getting the agent’s actions to match the actions of the model. Hurley is certainly committed to the claim that the tendency to imitate is automatic, and that the performance of imitative behaviour must be inhibited. However, this seems to be distinct from the claim that the performance of imitative behaviour is automatic. This is precisely not the case. Normally, we copy behaviour covertly, not overtly. It is this covert copying which Hurley claims forms the basis for the simulation routines we use to understand the instrumental actions of others. Mature adult humans can sometimes fail to keep imitation covert when they have suffered damage to their prefrontal cortex or when their caudate nucleus is overactive (Kinsbourne 2005, p. 165). Kinsbourne notes:

Echopraxics do not walk through the world twitching in response to every movement around them. They do not imitate the rustling of the leaves and they do not imitate cars screeching to a halt. One elicits echopraxia by being a doctor, facing a patient, looking somber and purposeful, and giving the patient tasks. (Kinsbourne 2005, p. 166)

So, even in this case, imitation is not wholly uninhibited. Williams makes an interesting suggestion about echopraxia. As already explained, he rightly stresses the role of social interaction in imitative learning. He suggests that echopraxia may be understood as the result of an
“impaired capacity for social judgement and flexible rule learning.” Both capacities are required for imitative learning according to Williams. Thus, when either of these capacities is damaged, the result is that subjects can no longer imitate. We think that the executive areas of the brain responsible for inhibiting imitative behaviour are also very likely involved in regulating behaviour in accordance with social norms and in flexible rule learning. Thus, we wonder whether the difference between Hurley and Williams on this point might not be so great.

Whiten and Heyes both put pressure on the conception of imitation Hurley assumes in her target article. Whiten objects to Hurley’s account of the phylogeny of imitation where emulation comes first and imitation only rarely follows. He rejects what he describes as the “dichotomy of imitation or emulation,” arguing that emulation comes in a variety of different forms, some of which overlap with imitation. In his ingenious and important “ghost experiments,” chimpanzees witness only the environmental effects of the complex use of a tool. If the chimpanzees were able to learn through emulation, it ought to be sufficient for them to just observe the goal-directed action. However, Whiten found that chimpanzees could learn the complex use of a tool only by perceiving another chimpanzee (not a “ghost”) use the tool. The suggestion seems to be that, at least for complex techniques, chimpanzees cannot figure out their own means to achieving an end. To learn a complex technique, they must copy both ends and means.

Before we explore some ways in which Hurley might have thought about the relation between emulation and imitation, we should briefly note that she was certainly no sceptic about imitative behaviour in animals. She was keen to stress just how complex imitative behaviour is, requiring as it does that an animal execute movements from its behavioural repertoire in a new way to achieve some desired result. This requires what she describes as the “flexible interplay of copying ends and copying means; a given movement can be used for different ends and a given end pursued by various means” (sect. 2.1, para. 5). Hurley doesn’t deny that some animals are capable of this kind of complex behaviour. She notes, for instance, that in the artificial fruits experiments chimpanzees will imitate selectively only when the method for opening the fruit is the most efficient. Elsewhere (Hurley & Chater 2005b), she discusses and seems to endorse Byrne and Russon’s (1998) finding of program-level imitation in gorillas and orangutans. Program-level imitation occurs when animals learn to copy a specific sequence of behaviours for the performance of a task, such as the preparation of a particular type of plant for eating. Thus, although Hurley certainly insists on a distinction between imitation and emulation, and insists that imitation is phylogenetically rare, she was no sceptic about nonhuman imitation.

Hurley did, however, insist that the capacity for social learning varied across species. She identifies two factors that contribute to this variation (sect. 3.3, para. 9):

1. The grain and complexity of instrumental control capacities
2. Considerations concerning which of the many control capacities have associated mirroring functions and how richly and flexibly these mirroring circuits can be linked

Some animals will be capable of performing instrumental actions that are more complex than others, where complexity of behaviour is a function of “means/ends chains of differing grains and lengths” (sect. 3.3, para. 11). An animal that combines multiple behaviours in ways that are appropriate to achieving a given end will be capable of forming predictive models that are much richer in structure than will an animal that can perform only simple behaviours to achieve its ends. Mirroring takes the instrumental associations between means (a motor program) and ends (the consequences of performing an action) that provide the information for predictive simulations and uses these associations to mirror the cause of another’s movement. Animals that employ predictive models that are rich in structure will be capable of mirroring instrumental actions that are equally rich in structure. The potential for social learning in such an animal will be much greater than that in animals that are only capable of simple behaviours and predicting the consequences of those simple behaviours. Hence, Hurley concludes that, “Mirroring and simulation might provide information about the goals of certain observed movements, given fine-grained, complex means/ends associations but not given coarser control capacities” (sect. 3.3, para. 13). Animals whose behaviour has a rich means/ends structure will be capable of complex forms of mirroring, and it is this mirroring that forms the basis for social learning. Animals whose behaviour lacks this structure will be capable only of movement priming or perhaps of goal emulation.

Heyes questions what she describes as Hurley’s conjunctive conception of imitation. According to the conjunctive conception, imitation requires both (1) observational learning, as when an agent learns an instrumental relationship between a bodily movement and its effect, and (2) a capacity for copying, where this involves the ability to perform the observed body movement. Heyes suggests that observational learning should be distinguished from copying. It is not clear to us whether she thinks copying is sufficient for imitation. We would question such a claim. Copying seems to be a type of behaviour that could be manifested by creatures that are only capable of what Hurley calls stimulus enhancement: The action of another animal draws the observing animal’s attention to a stimulus, and the stimulus then triggers an innate or previously learned response. Copying also seems to occur in cases of movement priming, when observing a bodily movement primes an animal to perform a similar movement, but not as a means to an end. It seems to us that copying has to go together with observational learning if the animal is to be correctly described as having learned through imitation – that is to say, by performing some sequence of movements from its behavioural repertoire in a new way so as to achieve a desired result. We would not describe an animal as having learned through imitation if the animal doesn’t understand the instrumental relationship between performing some bodily movements and achieving its ends or goals.

Longo & Bertenthal describe experiments that seem to challenge the connection SCM describes between mirroring and imitation. They argue that in some contexts the motor system will just copy movements and in other contexts the motor system will copy goals. They describe a paradigm in which subjects are shown a computer-generated
hand performing movements, some of which are possible and others of which are impossible. They found that subjects attempt to copy both the possible and impossible movements unless their attention is explicitly drawn to the manner in which the movements are being performed. It is not clear to us what the instrumental action is in this experiment. What are the means and ends? What is the computer-generated hand moving its hand to do? Given that it is unclear what the goal of the movement is in this case, it is hard to assess whether subjects were just copying the movements in the first set-up and the goals in the second (once their attention was drawn to the manner in which the movement was performed). We think it would be interesting to run the same experiment again, but this time to have the computer-generated hand explicitly perform movements with the end of achieving some goal. In one case, the end could be performed by means of biomechanical movements that are impossible for the human hand, and, in the second case, by movements the human hand could perform.

R4. Beyond shared circuits

Apart from the large raft of questions concerning the details of the inter-layer transitions, relations, and interactions, a number of commentators raised questions of scope. How much can the shared circuits model, with its strong commitment to a single kind of model and mechanism (a control-theoretic account of simulation and mirroring, pursued through a cascade of stepwise refinements) really explain? As the advertising would have it, the model aims to reach and illuminate “imitation, deliberation, and mindreading.” But does it really have the resources to do so? More accurately, just how much of our understanding of the minds, goals, and intentions of other agents can be explained using the kinds of resources Hurley so ably displays?

In much this vein, Carpendale & Lewis worry that the model “fails to reach action understanding because it relies on mirroring as a driving force.” Their key charge is that mechanisms of mirroring are simply too unintelligent to yield much in the way of action understanding. I may see you point at something, and that may automatically activate a pointing tendency in me (even if it is inhibited!), but what does that tell me about why you are pointing? What is missing, they suggest, is “experience in shared routines.” Carpendale & Lewis are right, we think, to identify these kinds of limits in the direct application of the model. For mirroring circuits do, indeed, only deliver information about others’ goals and intentions for classes of actions whose purpose is already appreciated: the act of using a tool to get food, for example. Even to intelligently combine the meanings of already-understood actions (to understand, for example, someone’s pointing to a tool that is being used to get food) would be a cognitive task whose successful undertaking plausibly requires more than the kinds of circuitry discussed in the target article alone.

This kind of worry is also prominent in the commentary by Preston, who, while agreeing that many basic perception-action mechanisms are preserved in our higher-level understandings, notes that such mechanisms require the agent to already command an understanding of (or at least, some form of representation of) the type of action or state at issue. Closely related issues are raised by van Rooij, Haselager, & Bekkering [van Rooij et al.], who note that direct simulation (used to mirror the means-end structure of observed actions) is often inadequate to reveal the goal of (or the intentions behind) an action. This is because our actual understanding of the operative goals and intentions is often profoundly affected by the context in which the action occurs. As a result, there is no one-to-one relation between actions (conceived as sets of motor signals) and goals.

Chakrabarti & Baron-Cohen, in their incisive discussion of some possible shortcomings of the SCM, point out that in many cases one needs to understand that the intention of the other is that you should do something different to (but complementary to) their own action. For example, two people can carry a heavy log, but not by copying each other’s actions, which will be log-end-specific. Here, the automatic activity of the mirroring system seems more of a liability than an asset (but see our remarks later on in the commentary by Hove, for one possible solution, consistent with the spirit of the SCM).

Yet another way of raising the same kind of issue is usefully displayed by Paglieri & Castelfranchi, who suggest that SCM is hamstrung by its failure to consider some additional roles of goal-states, namely, their role in (not just control but also) “evaluation and motivation”: that is, in deciding upon appropriate goals for our own and others’ actions, and in evaluating the actions in terms of those newly arrived-at goals. These elements are clearly central both to the understanding and the generation of intentional action. But they do not seem to be naturally captured, at least in any of the more advanced flavours we have just been discussing, by the bedrock story about mirroring and shared information spaces for perception and action.

All this, we feel, is exactly as it should be. It was not Hurley’s aim to offer a single kind of mechanism as a cognitive panacea, capable (all on its own) of explaining all aspects of human intelligent performance. Rather, the story is better seen as an attempt to display one key ingredient in such stories – one that has the virtue of first appearing (in basic form) at quite low levels of cognitive sophistication, and then making a contribution at many later stages. But intelligent human performance is not to be understood as flowing solely from the operation of that key ingredient alone. Rather, the ingredient is one enabling element in a larger story, whose full shape has yet to be determined.

Several commentators made helpful suggestions concerning such additional mechanisms. The basic story, Whiten suggests, might well need to be combined with an account that displays a developing capacity for “secondary representation” (Perner 1991), allowing us to maintain multiple perspectives simultaneously on a single physical event. Iacoboni reports the exciting discovery, in human frontal cortex, of what Iacoboni and Dapretto (2006) dub “super mirror neurons” that seem to modulate the activity of standard mirror neurons in various ways. Such modulatory effects look to offer one possible mechanism by means of which an increasingly intelligent use of the mirror system itself may be enabled. Longo & Bertenthal, while also noting the general limitation that mirroring requires the presence of the target action in one’s own
Several commentators note a prima facie challenge to Hurley's heavy use of control theory as a framework for SCM. Thus, Goldman notes that although there is strong evidence for the role of efferent copy and reafferent input in the domain of perception and action, no such body of evidence exists for many of the other domains (such as that of pain, feelings, and emotions – for the latter case, see also the contribution by Preston and our own comments in section R2) where various shared circuits also seem to enable mirroring and simulation to occur. This calls into question, Goldman suggests, the guiding idea that a control-theoretic perspective is apt as a general framework for all "shared circuit"-style phenomena. In its place, he proposes a Hebbian learning paradigm in which associative learning binds together various forms of neural activation. Oberman & Ramachandran reject the Hebbian alternative as a sufficient account of the development of F5 mirror neurons themselves, on the grounds that one still needs to explain why some F5 neurons end up having mirror properties while others do not. Goldman might (indeed, probably would) accept the existence of what Oberman & Ramachandran call "specialized mechanisms and hardwired constraints" for this special population, while still rejecting (but again, see our comments in section R2) any generalization of the control-theoretic explanatory apparatus to other domains in which mirroring and mindreading also seem to occur. At least one of us (Clark) is inclined to the view that the choice of a single control-theoretic perspective to address all such phenomena would indeed be premature. It seems unlikely that all the work required can be achieved by any single kind of mechanism. Nonetheless, the attempt to display a wide variety of mirror-system phenomena from a control-theoretic perspective strikes us as eminently worthwhile. Subsequent departures from that perspective, and the exploration of additional kinds of mechanism and explanatory framework, can then be motivated and described on a case-by-case basis.

We would like to end by flagging, once again, what we take to be the central contribution of the SCM, which is the suggestion that social cognition is continuous with more basic cases in which we perceive the actions of others by means that involve (and not merely as collateral effects or learnt associations) one’s own capacities for similar actions. In this way, Hurley posits a “shared information space” as a starting point for our explorations of interpersonal space and as a lever for our coordinated action. The problem facing the intelligent agent is then, not so much how to learn about the minds of others, as how to separate her own mind from the minds of others. Insofar as this is correct, it turns much of the standard discussion inside out. “Mindreading” becomes the norm, though at the cost (Star Trek fans will recognize) of a Borg-like threat of mutual cognitive dissolution. By monitored inhibition of output, we nonetheless end up extruding a genuine (but perhaps fragile?) self/other distinction in the face of (and without ever disabling) those basic tendencies of automatic copying and simulation. Such a story, if it is true, matters in ways that go far beyond the immediate concerns of the cognitive scientific community. It matters for policy, for education, for psychiatry, and for our own self-understanding as a species. Hurley herself was keenly aware of this larger picture, and we would recommend that interested readers consult her powerful

Much, to be sure, remains unresolved. Hurley’s story, at least in its broadest outlines, is compatible (as many commentators rightly observed) with a wide variety of ways of “filling in the mechanisms” and of linking (or even identifying) the putative layers. But whatever the details, there seems something deeply right about the guiding spirit. That spirit is a vision of the human mind as fundamentally social, as an evolved organ not of solipsistic individual cognizing, but of social and communal co-cognizing. That kind of talk is not unfamiliar (especially to those working in developmental science), but it has not yet informed the shape of the cognitive scientific mainstream. Hurley’s great achievement is to place this kind of model center stage, and to do so in a way that – as we have seen – is both concrete enough to raise questions of detail, scope, and adequacy, yet generally enough to invite constructive elaboration for many years to come.

ACKNOWLEDGMENTS
This Response was prepared thanks to support from both authors from the AHRC, under the ESF Eurocores CNCC scheme, for the CONTACT (Consciousness in Interaction) project AH/ES511139/1. We thank Till Vierkant and members of the Philosophy, Psychology and Informatics reading group at Edinburgh for helpful discussion of the target article.

References

Hurley: The shared circuits model


References/Hurley: The shared circuits model


Hove, M. J. & Risen, J. L. (submitted) It’s all in the timing: Interpersonal synchrony increases affiliation. [MJH]


Hove, M. J. & Risen, J. L. (submitted) It’s all in the timing: Interpersonal synchrony increases affiliation. [MJH]


