7 The Shared Circuits Hypothesis: A Unified Functional Architecture for Control, Imitation, and Simulation

Susan Hurley

7.1 Introduction

Various researchers at the currently buzzing intersection of work on motor control, imitation, and simulation have suggested that these processes are closely connected or even co-constituted (see and cf. Rizzolatti, vol. 1, ch. 1; Gallese, 2000b and vol. 1, ch. 3; and Iacoboni, vol. 1, ch. 2; Meltzoff, vol. 2, ch. 1, on the AIM hypothesis; C. Frith et al., 2000; Jeannerod, 1997, 2001; Grush, 1995 and forthcoming; Gerrans, forthcoming; Gordon, 2002; Oztop & Arbib, 2002; Proust, forthcoming; Wolpert et al., 2003; Wolpert & Kawato, 1998; Blakemore & Decety, 2001; Gallese & Goldman, 1998). There is something intuitively right and important here, yet the suggested relationships are often partial or expressed in one of several overlapping technical jargons that may be inaccessible to those in other disciplines who are interested in essentially the same issues. At this point it is worth exposing a set of related substantive issues fully and clearly, in a way that cuts across disciplinary boundaries. Accordingly, I here put forward in plain terms[1] one version of a unified framework that makes the relationships among the mechanisms that enable control, imitation, and simulation explicit. I call it the *shared circuits hypothesis*. It can be regarded as a relation of the *common coding hypothesis* about perception and action (W. Prinz, vol. 1, ch. 5), although it describes the commonality in terms of the dynamics rather than the coding of perception and action (see also Arbib on Prinz, vol. 1, ch. 8.6, p. 215). It is also closely related to Gallese's *shared manifold hypothesis* (vol. 1, ch. 3), though it situates elements of Gallese's views explicitly within an overall framework.

The shared circuits hypothesis is a midlevel hypothesis about subpersonal functional architecture, cast at a level of description between that

1. Though with links to technical terms noted.

of neural implementation and of the personal level of conscious perception and intentional action.[2] While it may be too early to claim definitive empirical support for this particular specification, it may nevertheless have heuristic value in sharpening up questions and predictions at both higher and lower levels, while avoiding over-simple or a priori projections between subpersonal and personal level descriptions. Related work in this area has not always kept clear track of distinctions between neural, functional subpersonal, and personal levels of description. While the boundaries between levels are not wholly opaque, it will conduce to clarity and progress to recognize distinctions between levels, and to frame issues about interlevel relations, more explicitly. Looking downward from the functional shared circuits architecture, we can ask whether there is evidence that particular neural circuits implement parts of it. Looking upward, we can ask what its behavioral and cognitive implications are, by comparison with quite different architectures. If information about self and other is processed subpersonally along the lines suggested by the shared circuits hypothesis, what implications if any might that have for the role and uses of such information at the personal level? For example, if intersubjective information is prior, at the subpersonal level, to information that differentiates self and other, does this have any implications about the basis of our personal-level knowledge of other minds? Unfulfilled predictions or incompatible neural circuitry could lead either away from the general idea of shared circuits for control, imitation, and simulation, or to a better specification of those shared circuits.

I draw attention as I proceed to some striking aspects of the shared circuits hypothesis. In particular, this hypothesis connects a shared information space for action and perception with a shared information space for self and other, while at the same time illustrating how the distinctions between self and other and between the imagined and the real can be imposed on these shared information spaces. In this model, information about persons arrives in the subpersonal version of the first person plural: without distinction or inference between self and other. Moreover, the shared circuits hypothesis illustrates a horizontally modular architecture: it avoids the common conception of perception and action as separate and peripheral to central cognition (see Hurley, 1998, 2001). Rather, it views perception and action as dynamically co-constituted and sees cognitively significant resources, such as the self/other and imagined/real distinctions,

2. Read "animal level" for "personal level" where appropriate; for a defense of this move, see Hurley, 2003.

and information for action understanding and planning, as emerging from the information space that perception and action share.

The shared circuits hypothesis is a theoretical model that describes a functional architecture in five major stages or—better—layers. (The allusion to Brooksian subsumption architecture is intentional—another expression of what I call "horizontal modularity"; see Brooks, 1999.) Some of these could be further expanded into different sublayers. Multiple instances of the shared circuits structure could be linked together into a network of such shared circuits, for hierarchical yet flexible control permitting the decomposition and recombination of elements. Further questions arise about how the specific layers might map onto phylogenetic or ontogetic stages. The order of the layers is intended to be logically intuitive and to reflect increasing complexity, but not necessarily to represent the order of evolution, development, or learning. In particular, the order of layers 1 and 2, and of layers 4 and 5, is heuristic, as I shall explain below. What is essential to the shared circuits model is the conception of progressing from local simulation, via the idea of a reversed forward model, to higher level simulation of global significance to the system, and from the shared space for perception and action to the shared intersubjective space and to self/other and imagined/real distinctions. But whether this theoretical model describes paths of evolution, development, and/or learning is a further question.

7.2 First Layer. Basic Adaptive Feedback Control with Inverse Model

The first layer constitutes a simple adaptive control system or servomechanism for general purpose motor control, which can usefully be compared to a thermostat. The elements of this layer are: (1) a target or reference signal (such as desired room temperature, in the case of the thermostat); (2) an input signal (actual room temperature, in the case of the thermostat), which is the joint result of (3) exogenous events in the environment (such as nightfall) and the output of the control system (such as the level of heat output); (4) a comparator, which determines whether the target and input signals match and the degree of any mismatch or error (for example, the room is still 5 degrees below the desired temperature); (5) output, which is determined by comparison between target and input signals (for example, heat output is turned up); (6) a feedback loop, by which output has effects on the succeeding input signal (for example, actual room temperature rises). (See figure 7.1.)

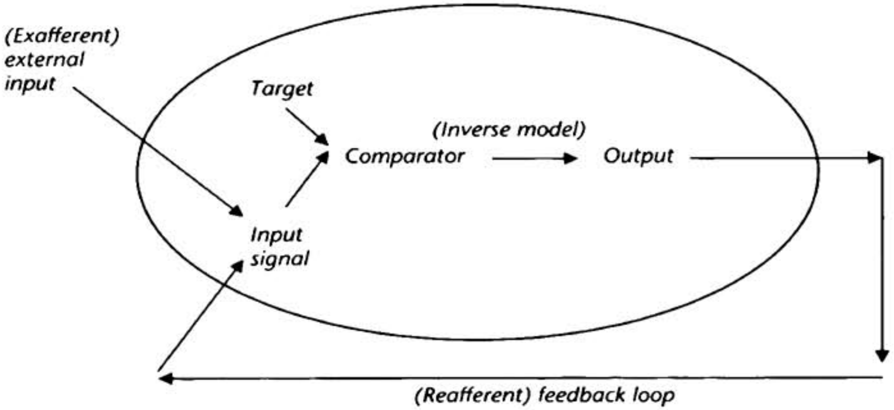

Figure 7.1
First layer: basic adaptive feedback control.

Some terminology and observations: the function that maps target signals onto output in the context of actual input signals is sometimes called an *inverse model*. In effect, it maps target to means, or specifies the means that will be used to approach the target, in given circumstances. The feedback loop at this layer is relatively slow, since it operates in real time (for example, the room takes a while to warm up after the heat has been turned up). In organisms, such feedback loops are often referred to as *reafferent feedback*: *reafference* is input to the system resulting from the organisms own activity, by contrast with *exafference*, input that results from exogenous events. Reafference, for example, includes visual and proprioceptive inputs resulting from movement of one's own hands, or movement through space, or manipulation of objects. Exafference captures inputs from events originating in the external environment, both inanimate and animate. It would include, for example, visual inputs resulting from movements by other creatures in a social group. This kind of system is *adaptive* because it adjusts itself to changing environmental conditions and compensates for exogenous disturbances: in the presence of different exogenous events, different output is needed to achieve the target. The control process is cyclical and dynamic; it does not have discrete steps or a nonarbitrary start or finish. Input is as much an effect as a cause of output. Information about inputs is not segregated from information about outputs; the dynamic relations among inputs and outputs are critical for control. This feature will be preserved as further layers are added; to the extent that perception and action

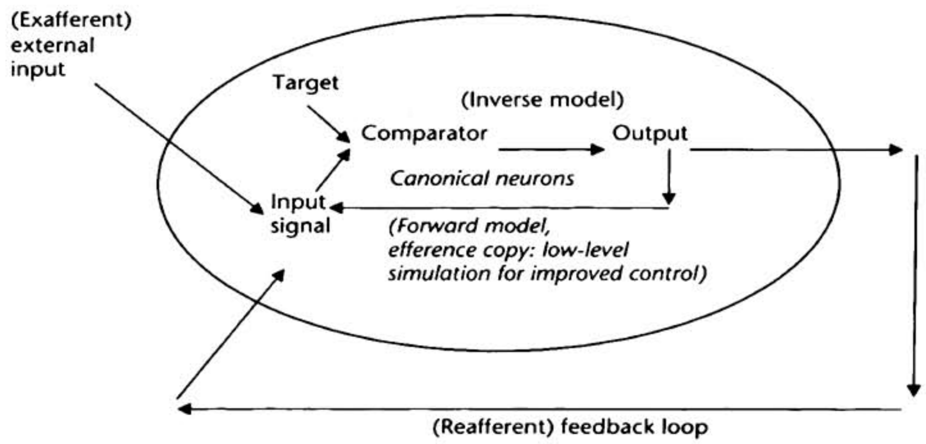

Figure 7.2
Second layer: simulation for improved control. Forward model added to basic adaptive feedback control.

arise out of a system with this basic feature, they share a fundamental information space (see Hurley, 2001, 1998).

7.3 Second Layer. Low-Level Simulation: Forward Model Added to Basic Adaptive Feedback Control

An inner loop is now added that maps the output signal back onto the input signal. In organisms this mapping is often understood in terms of *efference copy* (or *corollary discharge*); in engineering it is referred to as a *forward model*. Over time an association is established between copied output and subsequent input, so that efference copy comes to evoke the associated input signal. It can then operate as a forward model to predict the consequences of output on input. (See figure 7.2; new aspects are italicized.)

This process provides a general purpose improvement in the functioning of the motor control system, because the system does not need to wait on the real effects of output to produce reafferent feedback in real time. Rather, the system can bypass this relatively slow real-time process by learning and then anticipating the likely effects of output on input. In effect, efference copy produces a *simulation* of the expected effects of the system's output, which speeds up the control process and smoothes the appropriate behavioral trajectory. In the case of a significant mismatch between real and simulated input, a local switch can default back to real reafferent feedback while the forward model is further finetuned to improve its subsequent predictions (see Grush, forthcoming; Wolpert & Kawato, 1998; Haruno et al., 2001; Wolpert et al., 2003). This simulation is low-level in the sense that simulation can perform its speeding and smoothing functions without the system needing to monitor continuously or to access globally whether it is using actual or simulated feedback.

Recall that the order of the layers is heuristic and does not necessarily represent the order of evolution, development, or learning. For example, in the learning of particular tasks, layer 2's forward models may be acquired from feedback, enabling motor prediction, before layer 1's inverse model models are acquired, enabling motor control (Flanagan et al., 2003; here I am indebted to comments from Marco Iacoboni). One does not necessarily have to be pursuing a goal in order to learn to predict the sensory consequences of movement, even if it is natural to conceive of such prediction of feedback in an instrumental context.

Notice, however, that a system that includes reafference as well as efference copy has the resources to track the distinction between information about events in the world and information about goal-directed activity originating in the organism, that is, its behavior. When the train I am on pulls out of the station, I register movement relative to the train on the next platform, but this does not necessarily give me information about whether my train or the train on the next platform has begun to move. Comparison of efference copy with reafference gives an organism the resources to resolve the analogous subpersonal ambiguity, and hence provides information about the distinction between activity by the self and activity by the world ("self" here is neutral between persons and other animals). This is a familiar point (for discussion and references, see Hurley, 1998, pp. 140–141 and passim). Note that this information could provide part of the basis for the personal level distinction between action and perception, and that if so the distinction between action and perception emerges from shared processing resources, a shared subpersonal information space. Note also that information for the self/world and action/perception distinctions is prior to and more general than information for the self/other distinction (see layers 3 and 4 below). In this sense there are more and less fundamental layers of information about self.

At this point it would be predicted that cells or cell assemblies that mediate the association between efference copy and input signals might come to have both motor and sensory fields. Suppose an animal typically acts in a certain way on the perceived affordances of a certain kind of object: eating a certain kind of food in a certain way, for example. There will be associations between efference copy for the eating movements and a multimodal class of inputs characteristic of such objects and the eating of them. Any cells or cell groups that mediate this association might thus have both sensory and motor fields that between them capture the affordances of the objects in question. *Canonical neurons* are candidates for such sensorimotor affordance neurons (see Rizzolatti, vol. 1, ch. 1; Iacoboni, vol. 1, ch. 2; Gallese, vol. 1, ch. 3).

7.4 Third Layer. Reverse Forward Model for Priming, Emulation, and Imitation

Now consider how the system described so far would apply to visually transparent movements: movements that produce visual reafference, as when the creature watches his own hand movements. (The contrast here is with visually opaque movements, such as facial expressions: while they produce proprioceptive reafference, the creature cannot normally see his own facial expressions.) As the creature watches his own hand movements, an association is formed between efference copy and visual reafference from such movements. Here it would be predicted that cells that mediate this association might have matching sensory and motor fields. If the creature watches another perform hand movements of the same kind and he receives similar visual inputs, these will also activate his sensorimotor matching neurons with their motor fields. The sensory fields of such matching neurons cannot discriminate between his own actions of this kind and similar actions by others; the cells fire when he does something or observes someone else do the same thing. *Mirror neurons* are of course candidates for such matching sensorimotor neurons, and provide the neural underpinning for the kind of primitive blended intersubjective information space described by Gallese (vol. 1, ch. 3) in terms of a *shared manifold* and by Gordon (vol. 2, ch. 3) in terms of *constitutive mirroring*. Note the intimate relationship between the sharing of circuits for action and perception and for self and other: the blended intersubjective information space is a specification of and presupposes the generic blended sensorimotor information space.

Assume now that the sensorimotor matching association is bidirectional. Then, as well as efference copy simulating input signals, as in forward models described so far, input signals can also evoke efference or motor output. The forward model can, in effect, run in reverse (see and cf. Gallese

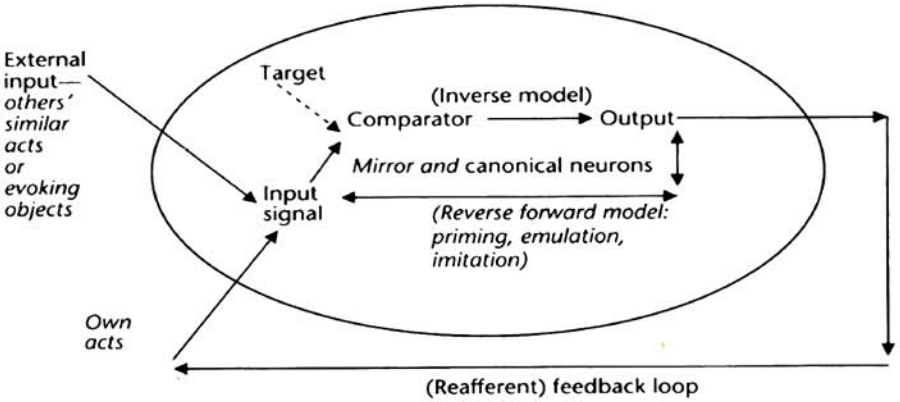

Figure 7.3
Third layer: reverse forward model for priming, emulation, imitation.

& Goldman, 1998; after I wrote this my attention was drawn to a similar idea in Blakemore & Decety, 2001). The predicted result would be motor copying at some level or levels. (See figure 7.3.) If a particular shared circuit controls details of movement (Rizzolatti's *low-level resonance*; vol. 1, ch. 1, pp. 65, 76), a predicted effect would be motor priming and copying of similar movement. If it controls the result of movement (Rizzolatti's *high-level resonance*, as in the monkeys in whom mirror neurons were discovered; vol. 1, ch. 1, p. 63) rather than the detailed movements that are the means to these results, a predicted effect would be emulation. If shared circuits for both motor means and results are themselves flexibly associated and work together, they could enable full-fledged imitation in which means as well as ends are copied (as revealed by the two-action methodology for identifying imitation; see vol. 1, chapters 9 through 12; see also Tomasello, 1999). Such full-fledged imitation would be predicted to be rarer than either response priming or emulation separately, since it would require circuits for both means and ends, appropriately linked together. And indeed it is rarer (see vol. 1, part II, on imitation in animals).

The distinction between an inverse model and a reverse forward model is functional; the neural paths that perform these functions might largely overlap. An inverse model functions instrumentally to bring about a goal by matching a target within a comparator system. A reverse forward model does not in itself have this instrumental function (see and cf. Peterson & Trapold, 1982). The priming of my own action by observing someone else's similar action is rather a by-product of the presence of the forward model,

which functions at layer 2 to provide predictions that improve the functioning of the control system. However, this priming may in due course be exapted for other functions, such as those associated with imitation and simulation.

The neural mechanism by which such reverse functionality might be acquired is a matter of speculation. Perhaps co-firing associated with the operation of the forward model strengthens backprojecting connections thus unmasking backprojections (cf. Heyes, vol. 1, ch. 6, p. 170, on "firing together and wiring together").

Circuits with this reverse forward model aspect could function in a variety of useful ways. They could operate to generate behavioral building blocks or modules that could be strung together in program level imitation, of sequences (Byrne, vol. 1, ch. 9) or of hierarchical structures (Whiten et al., vol. 1, ch. 11). They could allow an infant to form three-way associations among observed behavior by its parents (who have survived to reproduce, so may have adaptive behaviors not all of which are heritable), observed circumstances in which its parents perform such behavior, and its own similar behavior. Such associations could drive contextual imitation: act like that, when the environment is like this (Byrne, vol. 1, ch. 9, p. 228).

Note that the sensorimotor affordance associations described in the second layer (mediated by canonical neurons?) could also be bilateral. If so, observation of an object that affords some type of action would be predicted to prime the type of action afforded (see and cf. Lhermitte's utilization syndrome patients; Lhermitte, 1983, 1986; Lhermitte et al., 1986).

So far, the reverse forward model account does not explain imitation of visually opaque acts. How can a correspondence be established between one's own acts and similar acts by others, when there is no reafference in the same modality as observations of others' acts? For example, a creature receives visual input when observing another's facial expressions, but normally only receives proprioceptive, not visual reafference from its own facial expressions. How then can an association be established between my seeing another's facial expression and my making a similar expression myself? One answer is that some such correspondences are innate (Meltzoff, vol. 2, ch. 1). Another is that they are acquired in a variety of ways, through experience with mirrors, or with being imitated (Heyes, vol. 1, ch. 6).

The shared circuits model is compatible with these suggestions; it has no commitments about whether opaque correspondences are innate, acquired, or both. It also naturally accommodates another suggestion: that *stimulus enhancement* can establish associations between one's own and others' similar acts for visually opaque actions. Suppose a social creature repeatedly visual observes others' actions of a certain type, and its attention is thereby drawn to the characteristic objects of such actions. Such stimulus enhancement repeatedly evokes in the observer an innate or otherwise acquired response to those objects. As a result, an association is formed between visual observations of others' actions and one's own similar action. This is not initially imitation or any kind of copying; the object independently evokes others' and one's own acts. But while the link is initially indirect, nevertheless an association between own and others' acts may be established. Cells that mediate this association may acquire mirror properties such that subsequently merely observing another's act comes to prime similar action by the observer. In this way mere stimulus enhancement may develop into copying, and an indirect stimulus enhancement link may develop into a direct sensorimotor matching link. This suggestion about how opaque correspondences could be established is similar to one Heyes makes (in vol. 1, ch. 6) about the mediating role of words, but it is more general, and applies to stimulus enhancement at large.[3]

7.5 Fourth Layer. Simulation for Action Understanding with Output Inhibition

Next consider the possibility that a creature might observe another's act, which primes a similar act in the usual way, yet its own action is inhibited so that the observed behavior is not actually copied. In effect, the output of the reversed forward model is taken off line prior to motor output. Since observing the other's act is still associated with motor priming even when copying is inhibited, such observation could be interpreted as providing the observer with a simulation of what it is like to perform that kind of act, or a kind of understanding of the action: doing *that* is like *this*. Simulation for action understanding is off-line copying. (Cf. Rizzolatti, vol. 1, ch. 1, on action understanding preceding imitation; the views are consistent if priming of a movement and emulation of a goal are distinguished from full-fledged imitation, even though all are forms of copying. See also Meltzoff, vol. 2, ch. 1.) Applied to emulation circuits that control the result of movement, the simulation would provide information about the goal to which the other's movement is directed. The ability to pick up the information that another's movement is directed toward a certain goal can be regarded as enabling an early stage in understanding other agents and hence other minds. (See figure 7.4.)

3. Heyes's ASL model (Heyes, vol. 1, ch. 6) claims that visual and motor representations are linked according to the same Hebbian principles whether the movement is perceptually opaque or transparent. The only difference is that in the transparent but not in the opaque case, self observation will lead to the formation of links between movements that are the same from a third party perspective. What I am here regarding as stimulus enhancement could be regarded as acquired equivalence learning. The ASL model cites words as examples of the kind of stimuli that could act as the "third term" in acquired equivalence learning, but acknowledges that, as in most experiments on acquired equivalence in animals, the third term is often a non-linguistic stimulus. Thanks here to Cecilia Heyes.

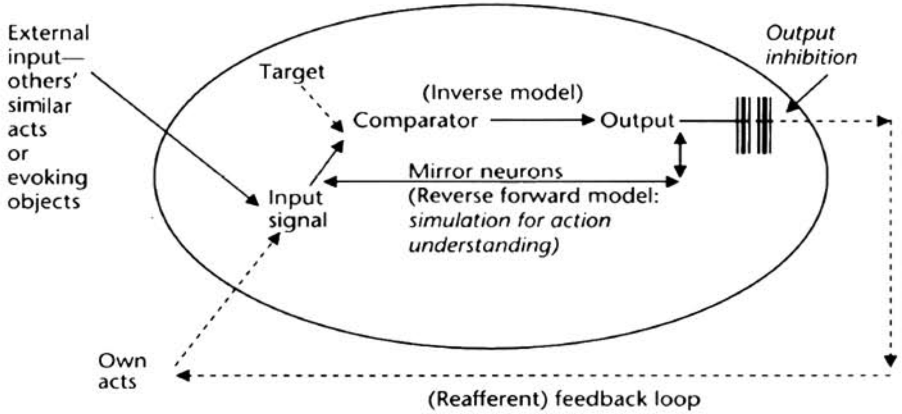

Figure 7.4
Fourth layer: simulation for action understanding with inhibition of output.

Although it uses the same circuit in reverse, simulation for action understanding can function at a higher level than the simulation for speeding and smoothing control described in the second layer (by "control" here, I refer to the overall function of the control system, not merely to that of the inverse model component). Recall that the basic functions of a forward model in a control system do not require the system to monitor continuously whether it is relying on actual reafference or on the forward model, even though it should be able to switch between them as needed. In other words, as long as the forward model works well and there is no significant mismatch in retrospect, the system does not need to know that it is using the forward model to improve control. The distinction between actual and simulated feedback does not have global significance for the system. By contrast, for simulation flexibly to subserve, as needed, understanding as opposed to copying an action, the system has to track the distinction between states in which the its output is inhibited and states in which it is not; this distinction *is* of global significance. Information about whether a movement is another's or one's own now overlays the primitive blended intersubjective manifold. Information about the distinction between self and other thus emerges. (Recall that the level of description of this information is subpersonal in this hypothesis; while this information is enabling, it is a further question how it is used at the personal level).

In particular, the shared intersubjective space is here prior to the self/other distinction, so that subpersonal information about persons in effect arrives in the first person plural, in a form that does not distinguish or infer between self and others. Subpersonal processing of information about other agents is more a matter of simulated recentering of first-personal or self-information processing than of inference from first person information to third person information. At the level of subpersonal information, the problem of "knowledge" of other minds is reconfigured: it is neither one of starting from information about the self and constructing a bridge across a gulf to information about other persons, nor one of starting from information about other persons and from the resources it provides somehow generating information about the self. The shared circuits hypothesis gives concrete if subpersonal form to the interdependence and parity of information about self and other minds.

Again, it is a further question how these subpersonal relations are reflected at the personal level. Do they give any support to a parallel priority of the first person plural at the personal level? How should "priority" indeed be understood in this question: as a question about development, or about the structure of mature understanding of other persons, and what is the relation between these? Can personal level understanding and knowledge of other minds be noninferentially based on or enabled by reliable subpersonal information? Is there any reason, conceptual or empirical, to believe that the problem of knowledge of other minds is similarly reconfigured at the personal level, so that it is neither one of starting from the first person perspective and constructing a bridge across a gulf to the third person, nor one of starting from the third person perspective and from the resources it provides somehow creating the first person perspective? Careful further thought is needed here. We should not simply help ourselves to an isomorphic projection from the subpersonal to the personal levels, but nor should we assume that the structure of subpersonal information processing has no implications for the personal level.

One way of responding to these issues is suggested by the affinities between the shared circuits hypothesis and Gordon's version of simulation theory (see especially 1995b, pp. 56, 58, 68; see also 2002; vol. 2, ch. 3). In Gordon's felicitous phrase, constitutive mirroring *multiplies the first person*, through a process of making sense of observed behavior and the self's matching response together, under a common scheme of reasons, a process that assigns incoherent mental states to different persons (Gordon, vol. 2, ch. 3, p. 103). While the shared circuits model offers a subpersonal description in which first person plural information is prior to first person singular and third person singular information, Gordon's account of the multiplication of the first person under a scheme of reasons is more ambitious in linking subpersonal constitutive mirroring to personal level understanding of other minds.

Gordon appeals to *ascent routines* to explain how simulation can underwrite mind reading without depending on inference from the first to the third person, as other versions of simulation theory do (see Gordon, 1995a; vol. 2, ch. 3; compare Gallese & Goldman, 1998). When I use an ascent routine, I answer a meta-question about my own or another's mental states by looking at the world; ascent routines are as well suited in principle to answering questions about another's mental states as about one's own. For example, to answer a question about whether I believe *p*, I consider whether *p* is true; to answer a question about whether another believes *p*, I perform an egocentric shift and imaginatively recenter myself to the other's perspective, and then again consider whether *p* is true. Similarly, for questions about what I or another perceive or intend: I look out at the world and the reasons it provides, though in the case of others having first transformed myself imaginatively. Note that on this view, to answer questions about what I or others believe, perceive, or intend, someone must first have the ability to perceive and act in the world. There is here another parallel, between Gordon's conception of ascent routines and the first aspect of the shared circuits model I noted earlier: the way a shared intersubjective space is distilled out of and simulatively employs the shared perception/action information space.

7.6 Fifth Layer. Counterfactual Input Simulation for Deliberation and Planning

Finally, the system can be taken off-line on the input side as well as the output side. Counterfactual inputs of possible acts and affordances can be simulated and the resulting motor activations entertained and compared without commitment to action and its costs; circuits for means and ends can be linked and recombined flexibly. Simulation at both ends could provide information that would enable deliberation and planning, and

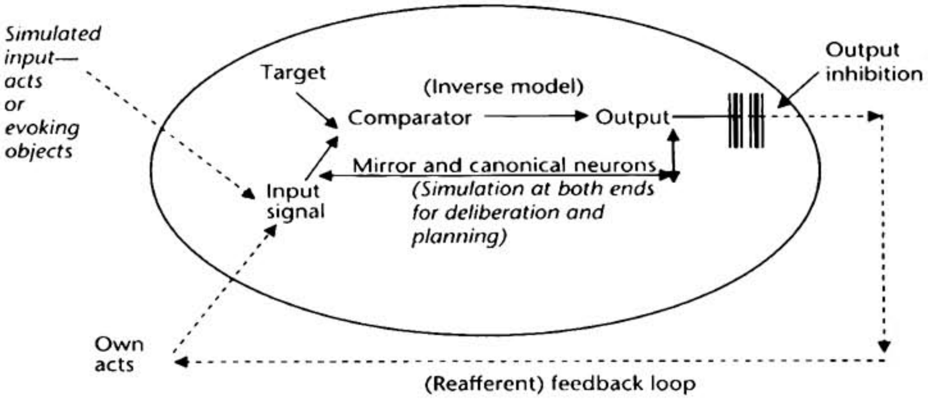

Figure 7.5
Fifth layer: counterfactual input simulation for deliberation and planning.

counterfactual and instrumental reasoning. For these functions, it would be essential for the system to keep track of whether it is simulating or not; the distinction between the imagined and the real thus emerges, close on the heels of the self/other distinction. (See figure 7.5.)

However, keep in mind that the order of the layers presented here is heuristic and does not necessarily represent the order of evolution, development, or learning; those are questions for further investigation. Layer 4's simulation and inhibition of output may accompany or follow rather than precede layer 5's simulation of input. That is, the shared circuits hypothesis does not specify the phylogentic or developmental priority between subpersonal information about self versus other and subpersonal information about the imagined versus the real. Rather, it provides generic, adaptable tools for framing more specific hypotheses.

7.7 Concluding Remarks

The five-layer shared circuits hypothesis I have sketched provides a unified subpersonal architecture for control, imitation, and simulation at a middle level of description: a functional level above that of neural implementation but below that of the normatively constrained and/or conscious personal level. It raises a variety of questions about how this functional architecture might map onto the neural and personal levels; the model may thus play a useful heuristic role even if it proves to be wrong in details (although care is needed to avoid over-simple interlevel projections and isomorphism

assumptions). For example, looking down to the level of neural implementation, we can ask: where might the postulated comparators be located? (PF? STS? Compare Rizzolatti in vol. 1, ch. 1 and Iacoboni in vol. 1, ch. 2.) Are shared circuits for the results of action found in prefrontal areas while those for detailed movements are in parietal areas? Does the model bear any relationship to the distinction between ventral and dorsal processing streams? Does it cast any light on the presence and function of mirror neurons in Broca's area and their relation to linguistic capacities? I have sketched the dynamics of the shared circuits model in cybernetic terms, but if neural implementations can be found, their interactive behavior through time could be represented as the evolution of a phase space in the manner of dynamical systems theory, and its attractor structure investigated.

Looking up to the personal level, we can ask: What behavioral, cognitive and functional predictions does the model provide? Intentional agents achieve their goals by means that can be given successively finer specifications, related by an asymmetrical "do *x* by doing *y*" relation: for example, I turn on the light by flipping the switch by moving my fingers. If we envisage a series of spectra with control of the ultimate result or goal of action at one extreme, and control of detailed fine movements that are the means to the result at the other extreme, then the shared circuits model could apply at successive linked points along such spectra. Thus the means outputted to the target of one circuit could be the target of the next circuit. A network of such linked circuits would support hierarchical control while permitting the flexible decomposition and recombination of goals and means. What relationship might such recombinant flexibility have to the recombinant flexibility characteristic of language? What does the model suggest about the functional relationships among three distinctive human capacities: for imitation, mind reading, and language (cf. Rizzolatti & Arbib, 1998; Iacoboni, vol. 1, ch. 2; Arbib on Iacoboni, vol. 1, ch. 8.2, p. 200; Meltzoff, vol. 2, ch. 3)? What implications does it have for the issue of whether simulation approaches to mind reading require an inference from the first to the third person (cf. Gordon, ch. 3; Goldman, ch. 2; Meltzoff, ch. 1; all in vol. 2)? What constraints does the model suggest on the relationships among various personal level distinctions: between action and perception, between self and other, between reality and appearance? Can the model play any role in distinguishing conscious and unconscious mental states and processes (see and cf. Hesslow, 2002; Frith et al., 2000; Gray, forthcoming; Jeannerod, 1997)? How might it be extended to include the emotional mirroring postulated by various researchers (see vol. 1, chapters 1 through

4, by Rizzolatti, Iacoboni, Gallese, and Decety & Chaminade, respectively)? Might the layers of the model usefully be mapped onto evolutionary or developmental stages, in theorizing the imitative and mind reading abilities of other animals or children?

I conclude by summarizing the shared circuits hypothesis. Theories about the control, imitative, and simulative functions of the mirror system, and evidence from imitation studies for ideomotor and common coding theories, suggest that perception and action share a fundamental information space that is preserved as higher cognitive capacities and distinctions are built on it. The distinction between results and the means to those results, essential to goal-directed, perceptually guided intentional action as well as to imitative learning, emerges as a flexible articulation of this shared processing. But perception remains fundamentally enactive, in a way that challenges orthodox views of perception and action as separately constituted and of perception as motivationally inert (see and cf. Kinsbourne, vol. 2, ch. 7; Noë, in press; Hurley, 1998).

The intersubjectivity characteristic of human beings, their distinctive capacity to understand and empathize with one another, is enabled as a specialization of enactive perception: I perceive your action enactively, in a way that immediately engages my own potential similar action, thus enabling me to understand, or to imitate, your action. Shared processing of the actions of other and self is a special aspect of the shared processing of perception and action. In an enabling role, this subpersonal informational structure may have implications for the epistemology of other minds. Within this informational structure, it is not so much that intersubjectivity bridges a self/other gap as that the self/other distinction is imposed on the fundamental information space that self and other share. Simulation theories of mindreading can be right about shared processing for self and other with respect to this fundamental intersubjectivity, even if more advanced aspects of mindreading require theorizing, in ways enabled by language.

Three aspects of the shared circuits hypothesis are noteworthy. First, it connects a shared information space for action and perception (understood in terms of control processes) with a shared information space for self and other (enabling imitation, intersubjective identification, and action understanding). In effect, the shared intersubjective space is distilled out of the shared perception/action space. Second, it illustrates how the distinctions between perception and action, between self and other, and between the imagined and the real, which provide information that enables the mental lives of persons, can be imposed on these shared information

spaces.[4] In particular, the shared intersubjective space is here prior to the self/other distinction, and information about persons arrives in the first person plural, in a form that does not register the self-other distinction. Processing information about other agents is more a matter of simulated recentering of the first person than of inference from the first person to the third person. At the subpersonal level, the problem of "knowledge" of other minds is reconfigured: it is neither one of starting from information about the self and constructing a bridge across a gulf to information about other persons, nor one of starting from information about other persons, and from the resources it provides somehow generating information about the self. The shared circuits hypothesis gives concrete form to the interdependence and parity of self understanding and understanding other minds. Finally, the shared circuits hypothesis thus illustrates what I call a *horizontally modular* architecture (Hurley, 1998, 2001): it avoids the common conception of perception and action as separate and peripheral to central cognition. Rather, perception and action are dynamically co-constituted, and cognitively significant resources, such as the distinctions between self and other and between the imagined and the real and information for action understanding and planning, emerge from the information space that perception and action share.[5]

Acknowledgments

For comments on an earlier draft, thanks to Chris Frith, Vittorio Gallese, Robert Gordon, Jeffrey Gray, Cecilia Heyes, Marco Iacoboni, Hanna Pickard, and Nicholas Rawlins.

4. In a way that converges with the broadly Kantian view that these two distinctions are intimately connected; cf. Strawson (1959, 1966).

5. For discussion relevant to this chapter, see especially Gallese, vol. 1, ch. 3, and Gordon, vol. 2, ch. 3.