《國立政治大學哲學學報》 第三十九期 (2018 年 1 月) 頁 37-70

On Susan Hurley's Shared Circuits Model of Action Learning

HUNG Tzu-Wei

Institute of European and American Studies, Academia Sinica

Address: No. 128, Sec. 2, Academia Rd., Nankang Dist., Taipei City 11529, Taiwan

E-mail: htw@gate.sinica.edu.tw

Abstract

Hurley's shared circuits model (SCM) is a cognitive architecture that specifies behavior-related sociocognition in terms of the dynamic interactions among perception, action, and the world. Despite its explanatory power for action learning, this model confronts a challenge: although Hurley introduces representation to the SCM to describe higher cognitive skills, she rejects classical representations with a

投稿日期：2016.03.25；接受刊登日期：2017.06.07
責任校對：劉又仁、劉鎧銘

domain-specific processor. The SCM has thus been questioned in terms of its ability to accommodate representation and computation. In this paper, I present a solution that integrates a motor selection mechanism into the SCM. I show how this integration, which requires neither an additional specialized processor nor representation in any classical sense, explains action learning and provides a basis for even higher sociocognitive skills in terms of general computational processes.

Keywords: Action Learning, Instrumental Action, Susan Hurley, shared Circuits Model · HMOSAIC Model

On Susan Hurley's Shared Circuits Model of Action Learning

I. Introduction

In the debate over cognitive architecture, classicists hold that the human mind is a digital computer. It contains *an input system* for encoding external stimuli into information recognizable by *a central processor* that handles the information in a sequential manner and then sends it to *an output system* for producing output to the environment. This information, called symbolic/classical representations, has a combinatory structure with semantics and syntax that resemble our language. For classicists, the manipulation of representation amounts to computation, as does cognition. Alternatively, connectionists hold that the mind is a neural-like network consisting of distributed units, each with an activation state (e.g., excitatory or inhibitory) determined by the weight of the connections among units. Some hold that each activation state, indicated by a vector, is a non-symbolic representation. Non-symbolic representations can be used to simulate symbolic representations and capture computational operations in the way that classicists describe. According to this view, cognition amounts to computing functions that are defined over vectors (Smolensky 1995).

A third paradigm, called the dynamic system approach, holds that the mind primarily contains input and output systems. The complex interaction among input, output, and the world suffices to the emergence of cognition (Brooks 1999). According to this view, representation is a problematic notion and earns no explanatory keep; thus, it should be replaced with *functional explanation*, in which the cognitive system is analyzed through its components, subcomponents, and the way in which they interact. Some scholars also suggest studying cognition without appealing to computation (Van Gelder 1995). Within this approach, Susan Hurley's (2008) shared circuits model (SCM) aims to offer a functional explanation of cognition through its functionally individuated layers. This model is useful for clarifying behavior-related sociocognitive skills. For example, the SCM explains why a baby can easily mimic an adult's facial expression without seeing its own face, just as the baby mimics an adult's observable actions. It elucidates how a hunter, without actually attempting all possible strategies, can conduct mental rehearsals when choosing to set a trap or attack a quarry. It also explains why a social agent might understand others' instrumental actions (i.e., intentional behavior with a means-end structure). These sociocognitive skills have a close relationship with real-time environmental change and rely on dynamic action-perception coordination. They can be handled fairly well by the SCM.

Empirical studies support the basic idea of the model and its underlying functional layers. For instance, one-third of neural cells with mirroring properties (i.e., properties related to a person's ability to reproduce observed actions) in the mesial-frontal lobe can implement the imitation functions that the SCM describes (Iacoboni, 2008). This mirroring occurs during the interaction between perception and action (Calvo-Merino, Glaser, Grezes, Passingham, and Haggard, 2005; Longo and Bertenthal, 2008). Emotion-related phenomena, such as empathy, can also emerge from perception-action mechanisms (Preston and de Waal, 2002; Preston, 2008).

Nonetheless, the SCM has its problems. Although this model is good at explaining behavior-related capacities, it is unclear how it explains higher cognition involving language. This is because Hurley rejected classical representations and the central processor, while their language-like structure is allegedly necessary to explicate linguistic structure. Hurley instead introduced a nonclassical notion of representation to a higher level of the SCM. However, what it exactly is and how it is manipulated were not made clear before Hurley passed away in 2007. Even if they were made clear, this SCM might not be compatible with the dynamics system approach. As Chemero and Cordeiro (2000) criticize, Hurley might have misused the concept of dynamic interaction because neither the representation nor the computation is involved in the dynamic system approach. Hence, the SCM confronts a core problem (C):

> (C) How can the SCM justifiably use "representation" and "computation" in the dynamic approach?

Although this difficulty seems to be a specific issue that focuses on a particular model in cognitive science, the **philosophical significance** of this problem is far beyond the scope of the SCM. Cognitive models or hypotheses labeled "situated cognition," "embodied cognition," "extended mind," and others also lack the above feature and thus confront the same difficulty. A solution for the SCM could provide a solution for similar models and hypotheses.

To solve the identified problem, one should specify the notion of representation, as well as how it is exploited in the SCM (or a revised SCM). The sense in which the SCM's representation and computation are compatible with the dynamic system approach should also be clarified. Most importantly, as Hurley's SCM is a model at the level of functional explanation, it is important to demonstrate how any revised model elucidates a skill (which is well-explained in the original SCM) in terms of new components, subcomponents, and the way in which they interact. To do so, Section 2 begins with a layer-by-layer description of Hurley's original SCM. Section 3 then rephrases the SCM in terms of three components and explains the potential benefits of the proposed integration. Next, Section 4 illustrates how this modified SCM can explicate behavior-related skills such as action learning. Finally, Section 5 outlines the further sociocognitive skills that might be elucidated by the revised SCM. In short, the central proposal that will be defended here is that the SCM, without assuming an additional specialized processor or representation in any classical sense, can process representation in learning action.

II. The original SCM

Hurley's SCM consists of five functionally distinctive layers (L1-L5). The most fundamental layer is basic adaptive feedback control (L1), which includes four constituents (Figure 1, in uppercase):

> REFERENCE SIGNAL: the target or goal of a system;
>
> SENSORY INPUT: environment stimuli and/or external feedback;

COMPARATOR: determines whether sensory input and the reference signal match;

MOTOR OUTPUT: regulated by comparison between sensory input and reference signal.

When receiving SENSORY INPUT, the system's COMPARATOR compares it with the REFERENCE SIGNAL to determine whether there is a gap or error between them. If there is, MOTOR OUTPUT will be modified to minimize the gap. This output is sent back to the system for further comparison and correction via an external feedback route, which enables the system to behave adaptively to the environment. As an example, consider when a rookie baseball pitcher attempts to throw a forkball *after* observing a coach's demonstration. The pitcher's visual memory of the coach's bodily movements constitutes the REFERENCE SIGNAL, the target action to be learned. When beginning the throw, the pitcher's own movements, the MOTOR OUTPUT, are returned to the system through an external feedback route, which, along with external stimuli, constitutes the system's SENSORY INPUT. The COMPARATOR then determines whether there is a gap between sensory input and the reference signal (e.g., whether the pitcher twists too fast). Therefore, the system can improve the next motor output and thereby facilitate the imitative learning of pitching at the most basic level.

However, the actual feedback may be too slow to guide the action in some circumstances. Simulative prediction (L2) provides predictive simulation by mapping motor output (efference copy of motor commands) onto sensory input (reafference). Besides, when seeing an action of someone else, an observer's functional mirroring (L3) gathers what causes would be needed for the observer to produce a similar motor action. This mirroring of causes will result in motor activation that, if not inhibited, generates an action similar to observed one (e.g., we sometimes smile unconsciously when seeing other people smile or hum a tune after hearing someone singing that tune). L2 and L3 differ in both the direction of signal transmission (output to input vs. input to output) and function (predicting output vs. copying input). However, at this stage, the observer's own actions may be confused with the observed actions because both L2 and L3 process signals between input and output. Output inhibition (L4) helps to resolve this confusion. L4 monitors and restrains L3's simulation of motor activation, thereby letting the system learn that inhibited actions are the copy of observed actions (i.e. the self/other distinction). L4 may also work with L2. L2's simulated results, if inhibited, provide information about possible actions. Otherwise, they provide information about actual actions (i.e. the actual/possible distinction). The cooperation of L2 and L4 enables counterfactual deliberation and choice among alternative possible actions. Finally, monitored input simulation (L5) can simulate counterfactual input to allow for mental trial and error and the detection of the intentions behind other people's actions. The main different between L2 and L5 is that L2 receives actual stimuli from the environment while L5 receives simulated stimuli from the system, e.g., sensory memory that is not reference signal.

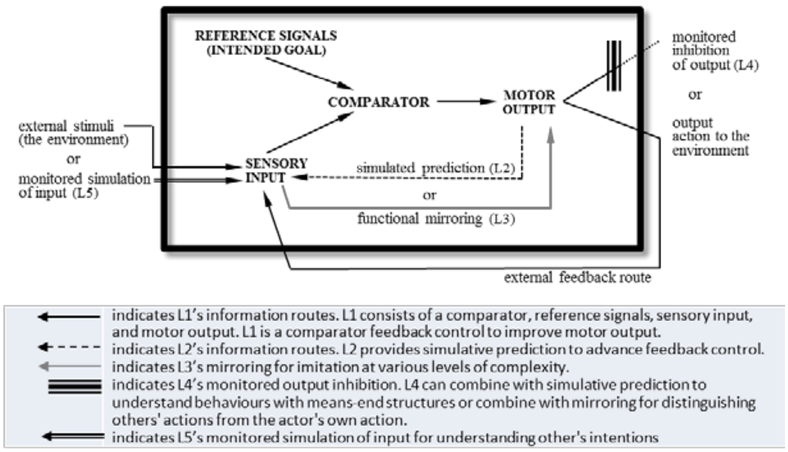

Figure 1**.** Reformulation of Hurley's SCM (2008)

III. Proposed Integration

Hurley (2008) only introduced representation in the higher layers of the SCM, but I hold that the entire model can be reinterpreted computationally (in which the term "computation" is used broadly to include both classical algorithms and neural networks). Although representation plays explanatory roles in some cases, computation does not require representation.[1] Thus, whether the SCM is computational does not depend on whether all its layers can manipulate representation.

Then, in what sense could the revised SCM be computational? Piccinini (2007) distinguishes *computational modeling* from *computational explanation*. The former amounts to using a computational system's output to explain another system's behavior (or using a computational system's internal states to describe another system's internal states). The target system itself need not be computational. Conversely, the latter is a special case of functional explanation, in which a target system's behavior is described by the system's own computational processes and their properties. As this paper explains a cognitive skills by postulating the underlying mechanisms, it offers a functional explanation, which becomes a computational explanation if the processes of its mechanisms are computational processes (Piccinini 2007). This happens when a well-established computational model is incorporated into the SCM, in which the interactions among its layers can be given with general rules and described algorithmically. It is in this sense that the framework to be proposed is *computational*. It is also in this sense that this paper explains a mental phenomenon by using a computational model (rather than describing the phenomenon with a computational model).

[1] For classicists, computation must involve symbolic representation (Crane 2004, Fodor 1981, Pylyshyn 1984, Sprevak 2010). However, as Piccinini (2007, 2008) points out, the main problem of this view is that it contradicts the standard view in computation science—namely, computational scientists individuate symbol strings by their formal properties instead of their content. Piccinini thus suggests defining computational states in terms of their functions and not representation.

The proposed framework has three components: (A) a minimal representation; (B) a pair composed of a representation producer and consumer; and (C) a mechanism for motor selection, the HMOSAIC. I shall explain them in turn.

A. Three Components

The first component is minimal representation (MR). In the first approximation, MRs are the framework's internal states that map onto specific external states. In the debate over nonclassicist representation, connectionist notions of representation are often questioned about the exact content they bear (Fodor 2000) and whether they play any legitimate role in explaining the system's behavior (Ramsey 1997). The notion of MR is hence explained in terms of these two points.

On the one hand, Hurley's (2008) SCM is inspired by Brooks' (1999) behavior-based robotics. While Brooks (1999) argues that (classical) representation is a problematic notion and earns no explanatory keep, his system also involves transformation of information, in which one number is passed from one process to another, and the number can be given an interpretation as well. If we map such a system and its state to another domain, we may define a minimal representation that the "numbers and topological connections between processes somehow encode" (Brooks 1999：90). Likewise, the brain's visual signals can be interpreted, so that some signals can be mapped onto an object's shape, some onto its color, some onto its motion, etc. After interpretation, these signals have the power of representing the aspects of the entire object.

On the other hand, MR has diverse types. For example, some MRs may stand for observed action (exafference input), some for intended action (reference signal), some for the next possible output (motor command), some for the copy of that command (efference), and some for the external feedback of previous output (reafference). Accordingly, when a fielder intends to intercept an approaching ball (reference signal), her movements (reafference) and environmental change (exafference) will be compared; this will result in the adjustment of motor commands for outputting action. Therefore, if the fielder suddenly changes her trajectory, this change is likely caused by the change of motor commands, which in turn results from the change of reafference and exafference. Hence, MRs explain not only information changes from one state to another, but also the framework's behavior.[2]

The second component is a pair composed of a producer for generating MRs and a consumer for manipulating MRs. The former includes an input receptor and a reference signal generator while the latter contains a comparator and a motor controller. For instance, the pitcher's retinas (the input receptor) transform the visible light reflected from the

[2] One might wonder how, given that the SCM is good at explaining cognition involved real-time environmental change, MRs can help clarify offline processing. Wheeler (2004) proposes an inner surrogate to explain offline processing (and linguistic mental rehearsal) in embodied–embedded systems. This surrogate is internal representation that reproduces the aspects (e.g., syntax) of previous received stimuli (e.g., a spoken sentence). It requires no central processor and thus differs from classicist representation. If we rephrase the information transmitted in the SCM into MRs, then the SCM will encode sensory stimuli into MRs in online processing. However, when external stimuli are unavailable, the SCM's L5 can re-input the stored MRs to simulate sensory input. As these MRs copy some aspects (syntax) of original stimuli, MRs help the SCM better explaining offline processing.

pitcher's own movements into signals that convey information about the pitcher's throwing motion (MRs). This information, together with the MRs of the reference signal generator (visual memory of the coach's movements), is sent to the comparator for analysis. The analyzed results are used by the motor controller to adjust output motor commands that cause muscle contractions for completing the entire action.[3] These motor commands are MRs because they stand for the subsequent states of bodily movements. The output movements and environmental change are then re-inputted into the receptor, which closes the information loop.

One may wonder why motor commands count as MRs, especially what justifies the relationship between commands and movements as representational, instead of merely causal. To answer this question, let us first compare a representational relationship to a causal relationship.[4] During *acquisition*, the two relationships share a similarity. In both cases, a learner needs to observe an event followed by another in order to establish a connection. For example, a learner needs to observe the burning of fire followed by the boiling of water to causally associate the two. An observer also needs to hear the word "apple," followed by seeing an apple to semantically associate them. However, their *processing* involves at least two differences. First, representations can be sequentially

---

[3] As the MRs sent from the producer to the consumer have the power to depict (the previous actual movements) and guide (the next possible movements) simultaneously, it is similar to Millikan's (2004) functional description of cognitive processes using pushmi–pullyu representations.

[4] Here, we may define a causal relationship between *a* and *b* if *a*'s presence is always followed by *b*'s presence, in which *a*'s presence is neither necessary nor sufficient to *b*'s presence. Also, two entities have a representational relationship if one has the power to stands for the other.

arranged to stand for something new. A bee's movements may represent different states (e.g., direction or distance of nectar) and can be arranged sequentially (bee dance) to represent an integrated state (the location of nectar). Conversely, it is not always clear what sequentially arranged causes associate to. Second, a representation can provide information about its referent in the absence of that referent, even where contextual information is limited. But a causal relationship is more flexible and diverse. The presence of a single event reveals little about what it results in. For example, even we learned that burning could cause many effects (warming someone, drying clothes, boiling water, cooking food, and lighting the dark, etc.), the presence of burning alone tells little about what effect it will pairs to. But even without contextual information, a representation "apple" is usually about an apple.[5]

If we apply the above distinction to the SCM, it becomes obvious why motor commands are MRs. Motor commands can be sequentially arranged to stand for new states, and they still offer information about possible movements in the absence of actual movements (e.g, forward model's offline processing). Therefore, the relationship between the commands and movements is representational. These commands have no combinatory structure with semantic and syntax, but simply map onto specific states. Therefore, they are MRs. In other words, the second component specifies where MRs come from and how they are processed in the SCM.

---

[5] Although some representations (e.g. "bank") may have more than one referent, an event of cause always has much more effects.

The third component is a motor selection mechanism proposed by Wolpert, Doya, and Kawato (2003)—that is, the modular selection and identification for control (MOSAIC) model. As previously noted, the pitcher's cognitive system must activate a set of appropriate controllers that cause a series of bodily movements to throw a forkball. However, determining which controllers to select for a given bodily state and informational context presents a problem. The MOSAIC consists of multiple controller–predictor pairs and can determine the correct controllers. It randomly initiates several controllers. Each controller generates a motor command to cause a movement and outputs the efference copy of that command to a paired predictor to simulate the outcome. This simulated outcome is sent to comparators for signal calibration, and only controllers whose simulated outcomes are within a given error range of comparison will be selected. The comparison results are stored in a memory component according to probabilistic rules, thereby allowing an experienced pitcher to start with a more precise initiation of controllers (i.e., with a bias). In addition, the MOSAICs can be arranged hierarchically to achieve more accurate selection (known as HMOSAIC). The HMOSAIC consists of numerous MOSAICs at multiple levels, including MOSAICs at lower levels for elemental movement control, middle levels for generating correct sequences of movements, and higher levels for tracking intentions or goals.

For example, when a pitcher intends to contain a baserunner, the reference signals (stored MRs about checking a baserunner) will be sent to the higher-level MOSAICs to decide which of the controller's set of movements should be selected (e.g., only watching the runner or passing the ball to a infielder). The result will be output to the middle-level MOSAIC to decide which activation order of these controllers is the most likely. Each controller of the set will activate lower-level MOSAICs to determine further movements required for passing the ball (e.g., twisting the waist, moving the arm, releasing the ball). The lower-level MOSAICs will calibrate the motor output against the sensory input (windspeed, the position of the batter, etc.) to fine-tune the next motor commands and thus produce a smooth movement for the entire action.[6] In other words, HMOSIAC facilitates the production and learning of hierarchically organized movements. The HMOSAIC, just as Hurley's L2, can predict the future state of a system (also called forward models). Yet the HMOSAIC has a greater number of subcomponents for subtle predictions. Thus, it can be used to replace the function of L2 so that the SCM can operate in a more refined manner.[7]

B. Proposed framework

How are the three components integrated into the SCM and how do they relate to L1–L5? The revised SCM is illustrated as follows. The most fundamental layer is an adaptive feedback control (L1), containing three components (and subcomponents):

---

[6] When using Bayesian terms to describe cross-level communication in this system, controllers at higher levels only receive posterior probabilities from MOSAICs at subordinate levels. Predictors at higher levels generate prior probabilities for MOSAICs at subordinate levels.

[7] Pickering and Clark (2014) distinguish an auxiliary forward model from the integral forward model. The former is extra circuitry adding to the core mechanisms of perception and action whereas the latter lies at the heart of all forms of perception and action. The use of the forward model in the revised SCM is closer to the latter sense.

MR: information transmitted among sub/components.

MR PRODUCER: includes an input receptor and a reference signal generator.

MR CONSUMER: includes a HMOSAIC (for replacing the comparator and motor control in the original SCM).

The basic idea of L1 is that the PRODUCER encodes external stimuli into MRs to guide the CONSUMER in revising the next actual action. A slightly more detailed description is that, when the coach intends to throw a forkball, her higher-level MOSAICs receive reference signals and activate a series of controllers that are required to complete the entire action (Figure 2 top). The middle-level MOSAICs predict the suitable activation order of these controllers. The lower-level MOSAICs then calibrate the above predictions against exafference and reafference. Thus, the output is improved by actual feedback (L1) and the predictions of the next states (L2).

Conversely, when watching the coach's throwing, an observer's input receptor inputs MRs into his lower-level MOSAICs through L3's information route (Figure 2 bottom). These MOSAICs will activate a series of controllers, representing the possible segmented movements of the observed action. (Each controller will output a motor command, which—if not inhibited by L4—will make the observer automatically reproduce the observed action.) The middle-level MOSAICs then determine whether these activation orders contradict any known sequence. If not, the segmented movements and their order are sent to the higher level for understanding the entire action. The visual memory of throwing a forkball can serve as simulated input (L5), thereby facilitating the observer's mental rehearsal of throwing. (How the observer learns the observed action will be unpacked in Section IV.)

The proposed integration has new features that extend each of the individual features. For example, Hurley's (2008) SCM is weak with regard to accommodating representation and computation, but the HMOSAIC is a well-established computational model. It can be complementary to the SCM with a computational base. In addition, the HMOSAIC was originally designed to solve problems with motor control whereas Hurley's L5 was proposed to simulate the possible action goal. As the HMOSAIC has more predictor–controller pairs, it can analyze more subtle input and achieve more exact predictions of the intended goal when combined with the SCM's L5. Moreover, Wolpert et al. (2003) emphasized actions involving sophisticated motor skills (e.g., balancing a walking stick on a finger) but paid little attention to simulated input (L5), inhibited output (L4), and functional mirroring (L3). Yet these signals are vital to other degrees of complexity in the imitative learning of action, ranging from low-level emulation to the learning of action. Thus, the HMOSAIC can be advanced by the SCM's L3–L5.

Employing MR in the SCM is not incompatible with the dynamic system approach (DSA). To clarify, what the DSA rejects is classical representation rather than all forms of representation. Brook's (1999) behavior-based robotics system also involves some sort of representation, in that the system and its states can be mapped onto another domain. Besides, some dynamic systems include common coding between perception and action. Since common coding conveys content between action and perception (Brass 1999, Prinz 1987), it has to entail some form of representation (but need not to be classicist one). This explains why the SCM, even at its lowest layer, requires some representations (i.e. MRs).

Likewise, employing computation in the SCM is not incompatible with the DSA. DSA does not reject all forms of computation, only the classicist view of computation. This view includes that computation is identical with the manipulation of symbolic representation and that a central processor of symbolic processing is indispensable. As previously noted, the revised SCM is computational not because of its capability of manipulating representation, but because its underlying processes can be given with rules and can be described algorithmically. Furthermore, there is no central processor involved in the revised SCM. The SCM still conforms to the DSA's definition regarding cognition as emerging from the complex interaction among input, output, and the world. Therefore, the revised SCM and the DSA are compatible.

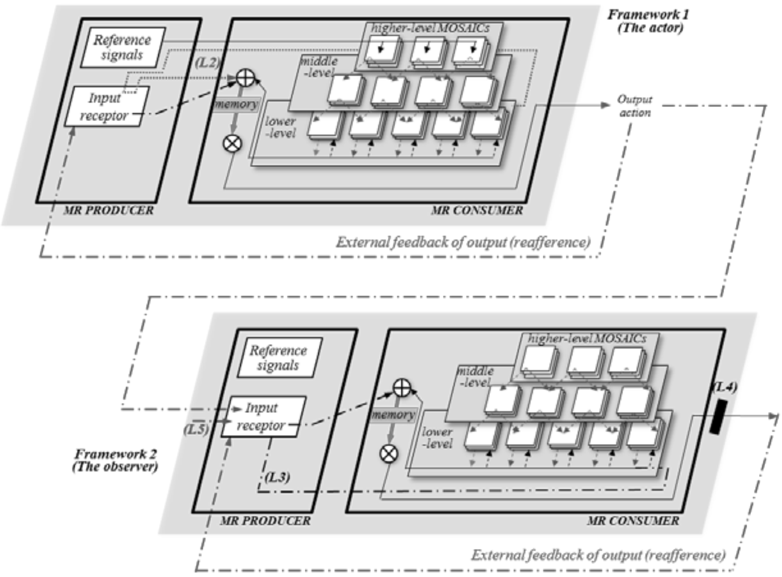

Figure 2. The interaction between an actor (top) and an observer (bottom)

IV. How the Framework Works

To test whether this proposed framework functionally explains the learning of instrumental actions, let us consider the non-instructed case of action learning in human beings. The task of action learning is divided into the following three parts: (A) segmenting and learning the elemental movements; (B) abstracting the sequential order of these movements; and (C) identifying the action goal (Wolpert et al., 2003). I shall demonstrate how the framework explains each part in turn.

A. Segmenting and learning the elemental movements

First, suppose that when observing a coach throw a forkball, a rookie pitcher's cognitive system (represented by Figure 3) activates the set of lower-level controllers A through D (Figure 3, bottom) through L3's direct copying based on the observed trajectory of the action and the observer's own bodily state. Rather than generating an actual action, these controllers output efference copies of the motor commands (dotted lines, Figure 3, bottom) to paired predictors to predict the coach's likely next movements. These initial predictions amount to prior probabilities, whereas predictions adjusted after comparisons with reafferent input constitute posterior probabilities. When the prediction of a local trajectory (for example, D) matches the reafferent input (e.g., the coach's waist twisting), then this trajectory becomes a properly segmented element of the entire action, and the elemental movement is stored in memory and learned. In this sense, the pitcher's MOSAIC serves as a tool to segment the coach's action (Wolpert et al., 2003), which partly explains the mind's tendency to parse continuous stimuli into manageable components (Hoffman & Richards, 1984; Hummel & Biederman, 1992; Tversky, Zacks, & Hard, 2008).

B. Abstracting the sequential order of movements

Second, the activation order of the pitcher's lower-level predictor, say, C→D→A→B, can be sent to the middle levels (Figure 3, top) to check whether this order contradicts any known sequences or restrictions. If not, then this order will be encoded and stored in memory, enabling the framework to learn the sequential pattern of the actor's continuous trajectory.[8] Nonetheless, one may argue that mimicking elemental movements and their sequences does not facilitate the creation of new actions in a system. A system may simply duplicate the entire sequence that is suitable for a given situation without rearranging the elements. Fortunately, the integrated framework depends on feedback processes to create a series of comparisons of its own inputs, which are essentially recursive. When recursive processing functions with the learning mechanism for elements and sequences, the framework can select elements in novel sequences that complement environmental changes, which allows the framework to create a potentially infinite number of actions based on a finite number of elemental movements.

[8] Haruno, Wolpert, and Kawato's (2003) simulation experiment supported that the HMOSAIC could be trained to learn action sequences. The HMOSAIC can select controllers in the correct sequence in response to various stimuli. This result holds when 5% noise is added to the sensory feedback.

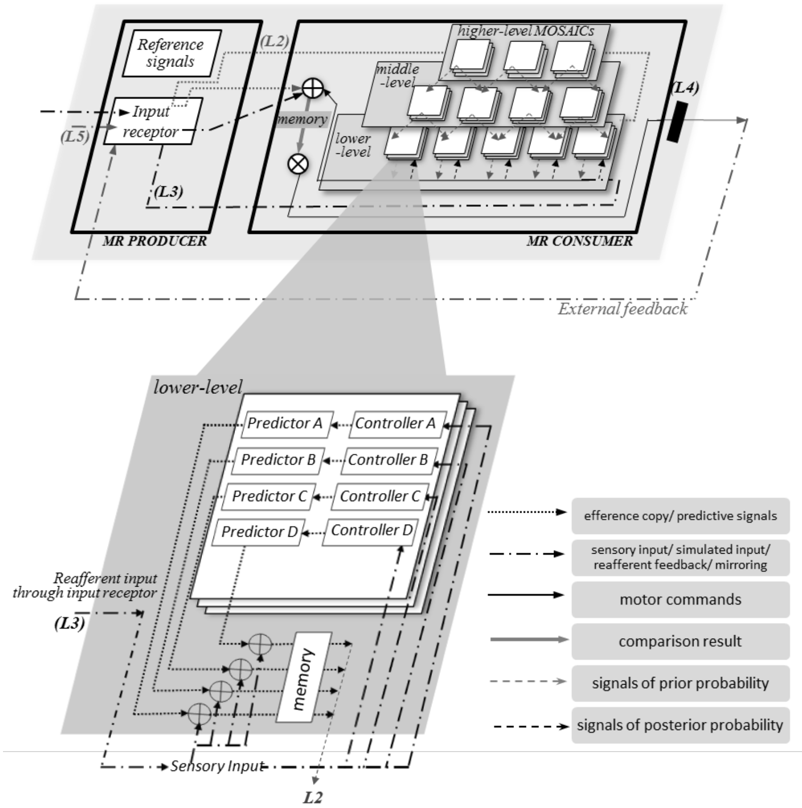

Figure 3. The proposed framework (top) and the enlargement of one of its lower-level MOSAICs (bottom)

C. Identifying the action goal

The final and likely most challenging task is to understand the action goal. Despite their success in simulating movement and sequence learning, Wolpert et al. (2003: 600) only noted that their model “dramatically reduces computational problems in action understanding”. Additionally, Hurley’s direct copying of observed actions (L3) is insufficient to generate a goal or imitation at various levels (Carpendale & Lewis, 2008; Preston, 2008; van Rooij, Haselager, & Bekkering, 2008; Uithol, van Rooij, Bekkering, & Haselager, 2011).

In fact, to identify the goal of an action, the observer has to infer from what is already known (an observable action) to what is yet unknown (the goal of the actor). In online reasoning, a dynamic system can derive information by directly interacting with the environment without the need of representation. For example, when seeing someone shaking a tree, and then apples falling from the tree, the observer may conclude that the shaking causes the apples to fall. However, external stimuli may not be always available. In offline or counterfactual reasoning, there is no external event for the system to engage with. Hence the system requires some surrogate (e.g., MRs) to represent possible events in offline and counterfactual reasoning. Hence, MR is important in reasoning from an observable action to an unobservable action goal.[9]

But how does the framework infer to the best prediction of goal? Briefly, the framework may initiate a number of higher-level MOSAICs

[9] I have described mathematically how an unobservable goal can be derived from an observable action (Hung 2015). The basic idea is that, according to Oztop et al. (2005), an actor’s intention may parameterize the motor control to generate actions. Therefore, the observer can analyze observed actions to predict the actor’s motor parameters and then the goal. In the framework, the observer receives exafference (i.e. observed parameters) and generates efference copy for predictions (i.e. predictive parameters). These predictions are compared with the immediate next exafference for possible revisions. If they match, then the estimated goal of the actor can be derived from the comparison. Since those internal signals (e.g., efference, exafference, etc.) are MRs by definition, they facilitate reasoning from an observable action to an unobservable goal.

to produce predictions about an actor's intended goal (e.g., to deceive a batter with a forkball or to strike out a batter with a fastball). These predictions are MRs because they stand for various possible goals. These MRs, via middle-level and then lower-level controllers, are tested downward with internal feedback (L2, Figure 3, top) and external feedback (Figure 3, top). When the comparator $\oplus$ indicates a gap, either these higher-level predictors are dismissed or their predictions are revised to minimize the difference. Although it is not sufficient to reveal the intended goal, this top-down processing may help to narrow the number of possible goal predictions. Similarly, the framework allows bottom-up processing for simulative mirroring. In unfamiliar situations, mirroring alone is insufficient to identify an action goal. However, when higher-level predictions that are activated by mirroring are cross-compared with those from inferential processing, the framework produces a more precise way of detecting the goal. For instance, suppose that some familiar segments of a novel action activate lower-level, middle-level, and higher-level MOSAICs through L3. The predictions of these higher-level MOSAICs are cross-compared with the results from top-down processing, such that the predictions with the highest likelihood (i.e., predictions with minimal gaps) will be singled out. Therefore, the intended goal can be selected from among the framework's initial predictions. Identifying the actor's intentions behind actions requires the repetitive processing of predictions, comparisons, and revisions.

Empirical research also supports the basic idea of the proposed framework or similar integration on Hurley's SCM. For instance, based on Hurley's SCM, Boza, Guerra, and Gajate (2011) designed an artificial control system to emulate general sociocognitive capacities. They also upgrade their modified SCM to improve the efficiency of imitation and mind-reading and to enable a self-optimization strategy for the control goals (Guerra et al. 2012). Likewise, Kilner, Friston, and Frith (2007) propose a Bayesian predictive coding for inferring intentions from observed action, which resembles the framework's use of HMOSAIC in identifying the action goal. Finally, Cacioppo, Fontang, Patel, and Decety (2014) argue that an experienced tennis player can accurately predict the intended action of a server using a prediction mechanism, which is also similar to the strategy of the framework's identification of the action goal.

V. Conclusions and Further Questions

To summarize, this paper first explains what the SCM is and then clarifies why and how to integrate the SCM with the HMOSAIC. Next, it discusses how this integration can be used to describe action learning, which involves MR and MR consumer-producer pairs. Consequently, the SCM can accommodate representation and computation without assuming classical notions of representation and computation.

A further merit of this integration is that it provides a basis for higher cognition such as speech processing. It is reported that the human capacity for understanding and producing actions is relevant to the ability to understand and produce language (Byrne, 2006; Garrod & Pickering, 2008; Hung, 2014; Kiverstein & Clark, 2008; Wolpert et al., 2003). To interpret a sentence, a listener's cognitive system also needs to segment continuous (auditory) flow into constituents (e.g., words) and to abstract the sequence of the constituents. These two tasks require similar processing procedures as segmenting movements and learning their sequence. In fact, a model based on this integration may associate a segmented word with its referent (Hung, 2014) and identify the syntactic rules of word combination (Hung, in press). With additional development, this model may be extended from action learning to language learning. This expansion might shed light on the relationship between the two significant human capacities and constitute a valuable focus for future studies.

參考文獻

西文：

Boza, A. S., Guerra, R. H., & Gajate, A. 2011. Artificial cognitive control System based on the shared Circuits Model of sociocognitive Capacities. A first approach. *Engineering Applications of Artificial Intelligence*, 24, 2: 209-219.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. 2007. Investigating Action Understanding: Inferential Processes versus Action Simulation. *Current Biology*, 17: 2117-2121.

Brooks, R. 1999. *Cambrian Intelligence: The early History of the New AI*. Cambridge, MA: MIT Press.

Buss, D. 2007. *Evolutionary psychology: The new Science of the Mind*. Boston: Allyn & Bacon.

Byrne, R. W. 2006. Parsing Behaviour. A Mundane Origin for an Extraordinary Ability?. In *The Roots of Human Sociality*. Ed. by N. Enfield & S. Levinson. 478-505. New York: Berg.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. 2005. Action Observation and acquired Motor Skills: An FMRI Study with Expert Dancers. *Cerebral Cortex*, 15: 1243-1249.

Carpendale, J. I. M., & Lewis, C. 2008. Mirroring cannot account for

Understanding Action. *Behavioral and Brain Sciences*, 31: 23-24.

Carruthers, P. 2003. The Mind is a System of Modules shaped by natural Selection. In *Contemporary Debates in the Philosophy of Science.* Ed. by C. Hitchcock. 293-311. Oxford, UK: Blackwell.

——. 2004. The Case for massively modular Models of Mind. In *Contemporary debates in the Philosophy of Science.* Ed. by R. Stainton. 3-21. Oxford, UK: Blackwell.

——. 2006. *The Architecture of the Mind*. Oxford, UK: Oxford University Press.

——. 2008. An Architecture for dual Reasoning. In *In two Minds: Dual Processes and beyond*. Ed. by J. Evans & K. Frankish. 109-128. Oxford, UK: Oxford University Press.

Chemero, A., & Cordeiro, W. 2000. *Dynamical, ecological Sub-persons.* http://host.uniroma3.it/progetti/kant/field/hurleysymp_chemero_cordeiro.htm. Retrieved January 5, 2010.

Cosmides, L., & Tooby, J. 1994. Origins of Domain-specificity: The Evolution of functional Organization. In *Mapping the mind: Domain-specificity in cognition and culture*. Ed. by L. Hirschfeld & S. Gelman. 85-116. New York: Cambridge University Press.

Crane, T. 2004. *The mechanical Mind*. London: Rutledge.

Garrod, S., & Pickering, M. J. 2008. Shared Circuits in Language and Communication. *Behavioural and Brain Sciences*, 31: 26-27.

Guerra, R. E. H., Boza, A. S., Gajate, A., & del Toro, R. M. 2012. Modified Shared Circuits Model for Manufacturing Processes Control. In *Brain Informatics*. 232-242. Heidelberg: Springer.

Haruno, M., Wolpert, D. M., & Kawato, M. 2003. Hierarchical MOSAIC for Movement Generation. *International Congress Series*, 1250: 575-590.

Hoffman, D. D., & Richards, W. A. 1984. Parts of Recognition. *Cognition*, 18: 65-96.

Hummel, J. E., & Biederman, I. 1992. Binding in a neural Network for shape Recogniton. *Psychological Review*, 99: 480-517.

Hung, T. W. (in press). Syntactic Processing in the behavior-based Architecture of the Mind. In *Topics in Intelligent Engineering and Informatics*. Ed. by J. Romportl, P. Ircing, E. Zackova, M. Polak, & R. Schuster. Heidelberg: Springer.

——. 2014. What Action Comprehension tells us about Meaning Interpretation. In *Communicative Action: Selected Papers of the 2013 IEAS Conference on Language and Action.* Ed. by T. W. Hung. 125-137. Singapore: Springer.

——. 2015. How Sensorimotor Interactions Enable Sentence Imitation. *Minds and Machines*, 25, 4: 321-338.

Hurley, S. 2006. Active Perception and Perceiving Action. In *Perceptual Experience.* Ed. by T. Gendler & J. Hawthorne. 205-259. Oxford, UK: Oxford University Press.

——. 2007. Neural Dominance, neural Deference, and Sensorimotor Dynamics. In *The Blackwell Companion to Consciousnesss.* Ed. by M. Velmans & S. Schneider. 640-656. Oxford, UK: Blackwell.

——. 2008. The shared Circuits Model (SCM): How Control, Mirroring, and Simulation can enable Imitation, Deliberation, and Mind

Reading. *Behavioral and Brain Sciences*, 31: 1-22.

Iacoboni, M. 2008. Mesial frontal Cortex and Super Mirror Neurons. *Behavioral and Brain Sciences,* 31: 30.

Jackendoff, R. 1987. *Consciousness and the computational Mind.* Cambridge, MA: MIT Press.

——. 1992. *Languages of the Mind: Essays on mental Representation*. Cambridge, MA: MIT Press.

——. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Kilner, J. M., Friston, K. J., & Frith, C. D. 2007. The mirror-neuron System: a Bayesian Perspective. *Neuroreport*, 18, 6: 619-623.

Kiverstein, J., & Clark, A. 2008. Bootstrapping the Mind. *Behavioural and Brain Sciences*, 31: 41-52.

Longo, M. R., & Bertenthal, B. I. 2008. Flexibility and Development of Mirroring Mechanism. *Behavioural and Brain Sciences*, 31: 31.

Millikan, R. G. 2004. *The Varieties of Meaning: The 2002 Jean Nicod Lectures*. Cambridge, MA: MIT Press.

Over, H., & Gattis, M. 2010. Verbal Imitation is based on Intention Understanding. *Cognitive Development*, 25: 46-55.

Oztop, E., Wolpert, D., & Kawato, M. 2005. Mental State Inference using visual Control Parameters. *Cognitive Brain Research*, 22, 2: 129-151.

Piccinini, G. 2007. Computational Modelling vs. Computational Explanation: Is Everything a Turing Machine, and does it matter to the Philosophy of Mind? 1. *Australasian Journal of Philosophy*, 85, 1: 93-115.

Piccinini, G. 2008. Computation without Representation. *Philosophical Studies*, 137, 2: 205-241.

Pinker, S. 1994. *The Language Instinct: How the Mind creates Language.* New York: Harper.

——. 1999. *How the Mind works.* New York: Norton Collins.

Pinker, S., & Jackendoff, R. 2005. The Faculty of Language: What's special about it? *Cognition*, 95: 201-236.

Plotkin, H. 2000. *Evolution in Mind: An Introduction to Evolutionary Psychology*. Cambridge, MA: Harvard University Press.

Preston, S. D. 2008. Putting the subjective back into Intersubjectivity: The Importance of Person specific, distributed, neural Representations in Perception-action Mechanisms. *Behavioural and Brain Sciences*, 31: 36-37.

Preston, S. D., & de Waal, F. B. M. 2002. Empathy: Its ultimate and proximate Bases. *Behavioural and Brain Sciences*, 25: 1-72.

Pylyshyn, Z. W. 1984. *Computation and Cognition: Toward a Foundation for cognitive Science*. Cambridge, MA: MIT Press.

Tversky, B., Zacks, J. M., & Hard, B. M. 2008. The Structure of Experience. In *Understanding events.* Ed. by T. F. Shipley & J. M. Zacks. 436-465. Oxford, UK: Oxford University Press.

Uithol, S., van Rooij, I., Bekkering, H., & Haselager, P. 2011. What do Mirror Neurons mirror?. *Philosophical Psychology*, 24, 5: 1-17.

van Gelder, T. J. 1999. Dynamic Approaches to Cognition. In *The MIT encyclopedia of cognitive sciences.* Ed. by R. Wilson & F. Keil. 244-246. Cambridge, MA: MIT Press.

van Rooij, I., Haselager, W., & Bekkering, H. 2008. Goals are not implied by Actions, but inferred from Actions and Contexts. *Behavioral and Brain Sciences*, 31: 38-39.

Wolpert, D., Doya, K., & Kawato, M. 2003. A Unifying computational Framework for Motor control and social Interaction. *Philosophical Transactions of the Royal Society of London B*, 358, 1431: 593-602.

論荷莉行動學習的迴路共享模型

洪子偉

中央研究院歐美研究所

地址：11529 台北市南港區研究院路二段 128 號

E-mail: htw@gate.sinica.edu.tw

摘要

英國哲學家蘇姍・荷莉所提出的迴路共享模型，是一個透過感知、運動與外在環境三者的動態系統來解釋人類行動的社會認知架構。雖然這個模型在解釋工具行動之學習上有很多優點，但由於荷莉並不接受古典表徵與相應的處理器，卻又引進表徵概念來處理高等認知能力，因此該模型不但被批評與動態系統不相容，其處理計算和表徵的能力也備受質疑。本文提出一個解決方案，將荷莉的模型與一運動控制機制相整合。在此整合架構下，不但無需預設古典表徵與處理器就能進行資訊處理，此架構也替未來描述更高層的認知能力（例如涉及命題、概念的認知能力）提供有利基礎。

關鍵詞：行動學習、工具性行動、蘇姍荷莉、迴路共享模型、階層式運動篩選與辨識機制