The skill of translating thought into action: framing the problem

Wayne Christensen
wayne.christensen@gmail.com
Philosophy, University of Barcelona, Barcelona, Spain

This is the accepted draft of the paper published in *Review of Philosophy and Psychology*, 16 November 2020.

Abstract

The nature of the cognition-motor interface has been brought to prominence by Butterfill & Sinigaglia (2014), who argue that the representations employed by the cognitive and motor systems should not be able to interact with each other. Here I argue that recent empirical evidence concerning the interface contradicts several of the assumptions incorporated in Butterfill & Sinigaglia's account, and I seek to develop a theoretical picture that will allow us to explain the structure of the interface presented by this evidence. The central idea is that neural plasticity incorporates metarepresentational rules for constructing representational systems and linking them. The structure of the cognition-motor interface is constructed flexibly during development and skill learning based on information processing demands.

1 Introduction

The way that cognitive intentions interface with the mechanisms of motor control has been brought into focus by Butterfill & Sinigaglia (2014), who describe a puzzle that they say confronts efforts to understand this interface. This is, in essence, that the representations employed by the cognitive and motor systems should not be able to interact with each other, since they have different formats, and translation is dismissed as a basis for interaction. The problem is then to explain how they do interact. Several solutions to this problem have now been proposed, including those of Sinigaglia & Butterfill (2015), Mylopoulos & Pacherie (2017), Shepherd (2017, 2019), and Ferretti & Caiani (2018). All arguably contribute valuable insights, but also suffer from problems that partly stem from adhering too closely to Butterfill & Sinigaglia's formulation of the interface problem. This characterisation reflects a widespread 'broad brushstrokes' conception of the interface, but adds some fairly specific theoretical assumptions concerning representational format and translation. Some of these assumptions are common (e.g., representations must be in the same format to directly

interact), while others are not (the ban on translation). But as I argue here, recent evidence indicates that the 'broad brushstrokes' view of the interface is not correct, and this evidence also contradicts several of Butterfill & Sinigaglia's assumptions. This requires us to reformulate the problem and raises questions about existing solutions.

Mylopoulos & Pacherie (2017) adhere most closely to the terms of the problem according to Butterfill & Sinigaglia, suggesting that intentions and motor representations are linked by associations rather than translation. Such associations are likely to be part of the overall solution, but this account is too limited. The evidence strongly suggests that a much richer set of representational interactions is involved, including some involving translation. Sinigaglia & Butterfill (2015) propose that links between intentions and motor representations can be constructed by imagining performing an action. Mylopoulos & Pacherie criticise this on the grounds that it tacitly violates the ban on translation, and because motor representations are standardly understood to be inaccessible to awareness. But in addition to evidence supporting translation, recent evidence indicates that some motor representations are consciously accessible. Butterfill & Sinigaglia's solution should thus not be rejected on those grounds, though it too fails to be sufficiently comprehensive. Shepherd (2017) suggests that intentions may have both a propositional and a motor format. The former allows them to interact with the cognitive processes involved in intention formation, while the latter allows them to interact with the motor processes of action execution. However, Shepherd also accepts translation, and it is therefore not clear why intentions need to be in a propositional format in order to interact with intention formation processes, or in a motor format to interact with action execution processes. Shepherd (2019) goes further and claims that action guidance involves interactions between representations in multiple formats. Ferretti & Caiani (2018) suggest that intentions and motor representations can interact because they have the same format, on the basis of evidence that some conceptual representations are involved in action control. But they tacitly abandon the assumption of no translation, and it is therefore not clear why intentions and motor representations need to be in the same format.

All these solutions have merits, but the nature of the problem itself needs to be systematically rethought. Here I undertake two tasks. The first is to summarize some of the main lines of empirical evidence concerning the overall structure of the cognition-motor interface and codify the resultant picture as a model. This allows us to identify assumptions employed by Butterfill & Sinigaglia that should be questioned, and it provides a target for theory. Our problem is then to explain the structure of the cognition-motor interface that emerges from this evidence. The second task is to take up this challenge. I do so by exploring lines of theoretical reasoning that could be used to support Butterfill & Sinigaglia's account of the interface problem, with the aim of developing, through contrast, a revised set

of ideas concerning cognitive architecture that will allow us to explain the structure of the model of the interface presented by the evidence. I draw on neuroscientific evidence and evolutionary considerations to identify a set of mechanisms that allow cognitive and motor representations to interact. These include metarepresentational rules for constructing representational systems and linking them that are incorporated in neural plasticity. These mechanisms allow the structure of the cognition-motor interface to be constructed flexibly based on information processing demands during development and skill learning.

The account offered here focuses on marshalling the general theoretical resources needed for explaining the interface. A follow-up paper will build on this framework, providing a more detailed explanation for the structure of the model of the interface characterised here and comparing this account to the other proposed solutions described above.

Since this account involves analysing and tracing interconnections between a large number of ideas I have used numerous labels for codification and ease of reference, but this in turn confronts the reader with some challenging working memory problems.[1] To ease this difficulty the labels are defined in Table 1 and figure 2.

Table 1 - Labels & definitions

Butterfill & Sinigaglia's interface problem (BS-IP)	Intentions and motor representations can't interact representationally, raising the problem of understanding how they do.
No multiformat processes (NO-MFP)	Representations in different formats can't participate in a common process.
No common planning process (NO-CPP)	Intentions and motor representations of goals can't participate in a common planning process.
No translation (NO-TRANS)	Intentions can't be translated into motor representations.
The Classical View (CV)	The cognitive system is responsible for conscious thought, the motor system is responsible for action execution, and there is strong functional segregation between the systems.
Strong functional segregation	The functions performed by a system are performed largely autonomously.
Milner & Goodale's *Two visual pathways theory* (2VP)	The ventral visual system is responsible for conscious scene perception and 'high-level action planning', while the dorsal system is responsible for 'fine-grained action planning'.
The *two motor systems* (2MS) model of Keele et al. (2003)	The motor system includes two major subsystems, a modular and an integrative system.
Weak functional segregation	A system performs many of its functions in interaction with other systems.
A revised version of two-pathways theory (2VP-R)	The ventral and dorsal systems exhibit weak functional segregation.
Linkage	Any way of relating representations.

[1] Thanks to both reviewers for their protests on this issue.

Systematic linkage	Linkage between representational systems such that novel representations in one can be linked to novel appropriate representations in the other in virtue of content.
Narrow translation	Production of approximately the same content.
Broad translation	Production of an appropriately related content.
Butterfill's (2007) *computational theory of modularity* (B-CTM)	Explains the encapsulation of modules in terms of format difference.
Representation-process mutual specificity (RP-MS)	Representation types and process types are mutually specific.
No common representations (NO-CR)	Modules and the general reasoning system cannot share representations.
The definition of 'format' argument (DFA)	An argument that RP-MS is true by definition.
Superformat	A format that encompasses multiple sub-formats.
The format principles argument (FPA)	Formats of different types are governed by different, non-overlapping principles (and thus have essences) and so cannot be combined in a single coherent representational system.
The iconic-discursive distinction according to Quilty-Dunn (2019)	(1) Parts of icons correspond to parts of what they represent, while this need not be true of discursive representations. (2) Icons represent holistically in the sense that parts of icons represent multiple properties at once, whereas the parts of discursive representations typically represent particular individuals and features.
Invariance and selectivity in visual processing	Principles of hierarchical information processing employed in the visual system, whereby generalization and specificity are achieved by integrating over lower order representations.
The format expense argument (FEA)	Multiformat processes are rare because mapping between formats is expensive.
Representational distance	Determined by the similarities and differences between two representational systems which make it easy or hard to map between them.
Metarepresentational rules	Rules for constructing and modifying representational systems.

2 Butterfill & Sinigaglia's interface problem

In outline, the problem as Butterfill & Sinigaglia define it (BS-IP) is that intentions and motor representations both represent action outcomes, but they each have a different format: intentions have a propositional format while motor representations have a motor format. However, representations in different formats can't participate in a common process (NO-MFP), which implies that intentions and motor representations of goals can't participate in a common planning process (NO-CPP)[2]. How, then, do they come to be appropriately

[2] Thus, they say "…some motor representations are like intentions in representing action outcomes while also remaining sufficiently unlike intentions in that no single planning process can integrate both intention and motor representation" (2014, p. 120).

matched? Butterfill and Sinigaglia (B&S) rule out translation from one to the other as a solution (NO-TRANS).

3 Outlining the empirical structure of the interface

3.1 The classical view

The NO-CPP claim can be viewed as an interpretation of what I will call the *classical view* (CV) of the boundary between the cognitive and motor systems. This incorporates the idea that the cognitive system is responsible for conscious thought while the motor system is responsible for action execution, and that there is strong functional segregation between the systems. *Strong functional segregation* means that the functions performed by a system are performed largely autonomously. The conception of the boundary as being between conscious thought and action execution is inherently vague, but it can be made somewhat clearer by specifying that the cognitive system determines the goals for intentional action and in some cases selects the action type. Goals are typically to bring about a change in the world (such as illuminate the room by turning on the light) or gain information (such as learn about the meal options by reading a menu). Planning and control of the movements that achieve these goals is performed by the motor system.

An important qualification is that the motor system is wholly responsible for action execution only when actions are fully automated. Just which actions these are is unclear, although actions paradigmatically thought to be automated, such as reaching to pick up a cup or kicking a ball, are relatively simple and unfold over relatively short timescales. Actions that have a complex, variable sequential structure and unfold over longer time scales, like cooking a meal or playing a game of soccer, are often thought to incorporate cognitive decision making in their execution. But actions as complex as making a cup of tea have been treated as being automated (e.g. Cooper and Shallice 2000). When cognitive decision making does contribute to the execution of complex actions it is for the selection of appropriate actions to support the overarching goal. Except in special cases, such as learning a new skill, the cognitive system does not control movement.

Figure 1 presents this interpretation of CV, with 1a and 1b depicting variations. The vertical axis represents functional hierarchy while the horizontal axis represents time. In figure 1a the individual forms an intention to achieve an outcome and this automatically generates an action that achieves the outcome. More specifically, a cognitive representation of the situation prompts a decision to achieve a goal. This generates in the motor system a corresponding motor goal, which is not a representation of the cognitive goal as such but when achieved will effectively bring the cognitive goal about. This motor representation of

the goal serves to select an action type which is then programmed (the parameters are specified) and executed.

For example, the situation representation might be that you have a fresh cup of coffee on the table in front of you, prompting the formation of an intention to ‘take a sip’. This intention induces in the motor system a goal that will bring about the intention. This goal is used to plan and control the production of a motor sequence that brings about the motor goal, and thereby the intention. Described in English, the motor sequence is reaching to and grasping the cup, lifting it to your lips, and sipping. The difference in 1b is that the intention specifies both the outcome and the type of action to perform. In other words, the intention is to bring about the outcome by performing a particular action. In the coffee example the intention might be something like ‘reach to the coffee and take a sip’. Since the action type is specified by the intention this will be incorporated into the motor goal.

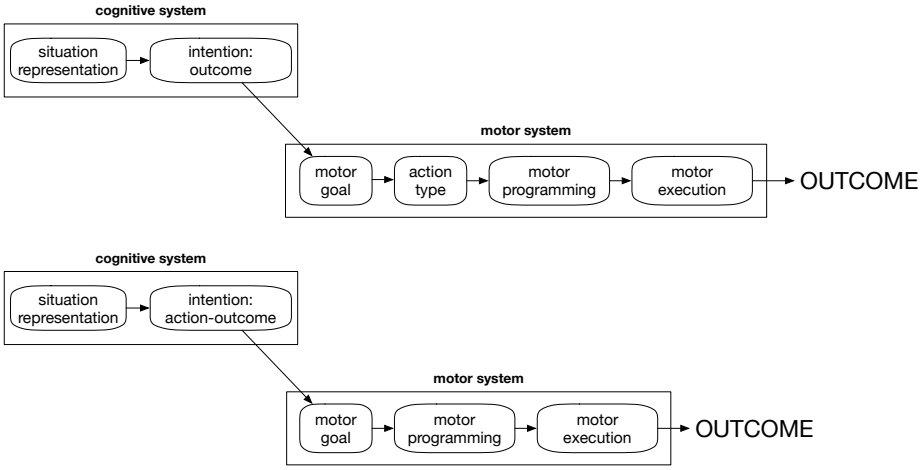

Figure 1: The classical view (CV)

The strong segregation between systems proposed by CV has an intuitive basis in the experience in everyday action of being unaware of the processes that govern the detailed structure of action execution. It has scientific foundations in a variety of single and double-dissociations associated with the multiple memory systems framework. An early and very influential finding was that, after surgery removing his medial temporal lobes, patient HM showed a preserved ability to acquire sensorimotor skills despite an inability to acquire new conscious memories (Milner 1962). A wide range of subsequent research supported this general picture, which resulted in the formulation of the multiple memory systems account and its primary distinction between declarative and nondeclarative memory (Squire 2009). The former is divided into *semantic* and *episodic memory*, where the latter is broadly defined as memory for personal events and the former is broadly defined as memory for facts. Declarative memory can be flexibly accessed and processed in working memory (Baddeley

2000, Dehaene & Naccache, 2001), whereas non-declarative memory is thought to be nonconscious and expressed in performance (Squire 2009).

The classical view has been widespread; for influential examples in psychology see Masters (1992) and Beilock & Carr (2001). For critical discussions see Christensen et al. (2016) and Montero (2016).

3.2 Difficulties for the classical view

BS-IP can be viewed as an interpretation of CV, but the fit is not entirely easy. Firstly, contrary to NO-TRANS, it has been standardly assumed in motor control research that cognitively specified action goals are provided to the motor system by means of translation (e.g. Willingham 1998, Haruno et al. 2003). Secondly, on the classical view motor control operates largely autonomously *only* when actions are well-learned and conditions are normal. When a skill is being acquired, or when conditions are unusual, cognitive processes control aspects of motor execution. This appears to violate NO-CPP, inasmuch as some motor planning and control processes are occurring in the cognitive system, which presumably include motor representations.

We need to clarify what counts as a motor representation, however. On a *systems-based* definition within the classical framework, cognitive and motor representations are distinguished by the fact that they belong to the cognitive and motor systems, respectively. On a *feature-based* definition they are distinguished by having distinct features. Mylopoulos & Pacherie (2017), for instance, characterise motor representations in terms of a list of characteristic attributes.[3] On a *role-based* definition they are defined specifically by role, regardless of system membership or other features. The role-based conception is the most important for type differentiation in a system predominantly characterised by functional organisation. Within the classical framework, the role of cognitive action representations is, broadly, determining what actions to perform, while the role of motor representations is to plan and control action execution. Level of control must be specified, however, since action control can have a multiple levels in which goals govern implementation processes. Motor representations, as the name suggests, are usually thought of as operating at the level (or levels) of control concerned with the movements involved in action execution. On this definition, then, representations that function to specify and control movement are motor representations. However, this implies that representations of the cognitive system that are

[3] However, Mylopoulos & Pacherie include functional role amongst these features. It is possible that they might regard role as fundamental.

used to control movement during skill learning and in unusual conditions are motor representations. This would be a violation of NO-CPP.

An alternative interpretation is that 'motor representations' are representations of the motor system, and that the cognitive control of movement that occurs during skill learning and in unusual conditions does not violate NO-CPP. Instead, the cognitive system simply provides more detailed, movement-specific goals to the motor system. This account is in tension with the role-based approach to type differentiation of representations, but it can be further claimed that, although cognitive representations can be used to control movement, this is not their primary specialisation and they are not well suited to it. For this reason, the motor system takes over control of movement during skill learning and performs this function autonomously once a skill has been well-learned. This position is made untenable, however, by evidence that control of movement is the normal function of some cognitive representations.

Hints of this issue are already evident in figure 1b, where the intention specifies the action type, with the example in the text being reaching for and sipping a cup of coffee. This action type has characteristic movement structure, namely the type of movements involved in reaching and sipping a cup of coffee. This level of movement specification is abstract, but motor control theories postulate generalized representations that capture standard patterns that occur across situations (Schmidt 1975, Wolpert et al. 1995). Based on role, then, these are motor representations: they specify movement features and contribute to producing them. CV claims that there is a clear functional distinction between cognitive and motor representation, but the distinction is less clear-cut on close examination, even for simple examples usually taken to be compatible with CV. Several theories of motor control violate the distinction systematically by proposing that some aspects of motor control are performed by the system associated with working memory and conscious, conceptual cognition.

The *two visual pathways theory* (2VP) of Milner & Goodale (1995, 2008) has been linked to the memory systems framework and is often viewed as supporting CV, encouraged by the characterization of the systems as "vision for perception" and "vision for action". On close examination however, the account is significantly different to the interpretation of CV just given and warrants being considered a distinct conception of the cognition-motor boundary. The theory retains the strong functional segregation of CV, and one of the core forms of evidence on which it is based is an apparent double dissociation between *optic ataxia* and *visual agnosia*. In the case of optic ataxia, which involves damage to the dorsal stream, the standard characterization has been that action planning and control is impaired while object recognition remains intact. In contrast, damage to the ventral system gives rise to visual

agnosia, which has been characterized as involving impaired object recognition while leaving intact the ability to make accurate reaching and grasping movements.

Nevertheless, two visual pathways theory differs from CV in proposing that the ventral system, whose representations can be conscious, contributes to motor planning. More specifically, Milner & Goodale (2008) characterize the contrast between the two systems as follows. The ventral visual system is responsible for conscious scene perception and 'high-level action planning', understood as the identification of objects, parts of objects that are targets for manipulation, and selection of actions to perform on them. In contrast, the dorsal system is responsible for 'fine-grained action planning' (e.g. reach direction and anticipative setting of grip aperture) and control of action execution. The crucial point to note is that this 'high level action planning' includes functions that play a role in programming the movements the execute the action.

The difference between CV and 2VP can be illustrated with respect to the kind of dissociations we should find between the systems. Taking the example of the action of reaching for and sipping a cup of coffee, CV predicts that damage to the cognitive system might result in impairments at deciding whether to reach for and sip a cup of coffee, but it should leave intact the ability to execute reaching for the cup, grasping it, bringing it to the lips, and sipping. In contrast, 2VP predicts that some of the structure of these movements will be impaired if the cognitive components of motor planning are impaired. And indeed, the empirical basis for Milner & Goodale's claim that the ventral system performs high level action planning is evidence of this kind. Patient DF, who has visual agnosia resulting from ventral stream damage, and is one of the primary sources of evidence for 2VP, exhibits motor planning deficits. Carey et al. (1996) found that DF had difficulty selecting the use-appropriate part of an everyday object for grasping, such as the handle, even though the grasping itself was executed with normal ability.

Other research has similarly distinguished between two motor systems, one of which employs representations and processes accessible to consciousness. According to what I will call the *two motor systems* (2MS) model of Keele et al. (2003) and Clark & Ivry (2010), one of the motor systems is a modular system which exhibits learning that is slow, integrates information only within stimulus modalities, is insensitive to the conceptual categories employed in task performance, and is inaccessible to awareness. The other is a non-modular system which can learn rapidly, can associate information across modalities, is sensitive to the conceptual categories employed in task control, and involves processes that can be, but are not always, accessible to awareness. Keele et al. point out that their account undermines the standard distinction between procedural and declarative memory.

If the nonmodular motor system is integrated with conscious cognition then high-level action planning and control processes should tax working memory, and a significant strand of research provides support for this (Weigelt et al. 2009; Logan & Fischman 2011, 2015; Spiegel, Koester & Schack 2013). Weigelt and colleagues had participants perform a task that combined motor demands (opening drawers and turning over cups) with a memory task (memorizing letters on the underside of the cups). They found that memory performance was impaired, indicating that the motor component of the task placed demands on working memory.

There is, thus, a significant body of convergent evidence supporting the view that the motor system is differentiated into multiple systems, and that part of the motor system is integrated with the cognitive system associated with conscious, conceptual control. This evidence is contrary to CV. The story does not end here, however. In recent years M&G's two pathways theory has been criticized for overstating the degree of functional segregation between the ventral and dorsal systems. It has been argued that the differences between optic ataxia and visual agnosia are more complex than traditionally thought (Pisella et al. 2006, Schenk et al. 2011, Rossetti et al. 2017). The impairments associated with optic ataxia, for instance, are present only in peripheral vision, not central vision. Patient DF, who had been interpreted as showing impaired object recognition with preserved (low-level) motor function, also shows impaired motor performance for pointing in peripheral vision. Multiple subsystems within the dorsal stream have been identified (Rizzolatti and Matelli 2003, Binkofski & Buxbaum 2013), and there are extensive connections between the dorsal and ventral streams (Bullier et al. 1996). Such connections afford the possibility for rich interactions between the ventral and dorsal streams. There is evidence that such interaction occurs during action execution and that its nature depends on the control requirements of the action (Grol et al. 2007, Verhagen et al. 2013).

Taking this evidence into account yields a revised version of two-pathways theory (2VP-R) that abandons the strong segregation claim in favor of a *weak segregation* view. Weak functional segregation between component systems occurs when there is rich functional interaction between systems and comparatively weak internal processing autonomy. Functions, such as motor planning and control, are often performed by means of cooperative input from multiple systems.

In the context of two pathways theory, van Polanen & Davare (2015) propose a model of how such interactions might operate. This account holds that there is greater interaction with increasing action complexity. van Polanen & Davare broadly follow Milner & Goodale's (2008) claim that there is a progressive increase in dorsal stream contribution in the course of skill learning, with the ventral system playing the dominant role in controlling novel actions and

the dorsal stream becoming dominant as automaticity develops. In this respect Milner & Goodale's theory is similar to Fitts & Posner's (1967) account of the changing roles of cognitive control and automaticity during skill acquisition. van Polanen & Davare depart from this picture, however, by proposing that the ventral stream continues to play a role in well learned skills that becomes greater when idiosyncratic features of the situation are important for control and as the complexity of the action increases. Here their account breaks with Fitts & Posner's model and is similar to that of Christensen et al. (2016). In the case of simple, familiar actions, like simple grasping, the dorsal stream may operate largely autonomously. When object properties and context are important, such as when taking into account the shape, weight and material (e.g. ceramic or plastic) of a particular cup of tea, the ventral stream contributes information for control. In the case of complex tool use van Polanen & Davare say that the ventral system will make a substantial contribution because conceptual knowledge of object functions is important. Interactions between the systems will be optimized in skilled actions by mechanisms that selectively employ the most predictive information in a given context.

Stepping back from the details of this model, we can codify the overall picture by means of an interactive, multi-level architecture (IMLA; figure 2). Not only are there more levels, but interactions between levels are considerably richer than is recognized by CV. The uppermost level corresponds to the standard conception of the cognitive system as a general reasoning system (GRS). The next level down (SAS) is a system responsible for situated action control which controls action in relation to situation and task structure, such as using a screw driver to screw in a screw while attaching a leg of an Ikea chair. This system interacts with an integrative motor system (IMS) which performs the kind of high level motor planning described by 2VP, together with the control of execution described by 2VP-R. SAS and IMS jointly control action execution. At the lowest level is a modular motor system (MMS) which encompasses the dorsal system.

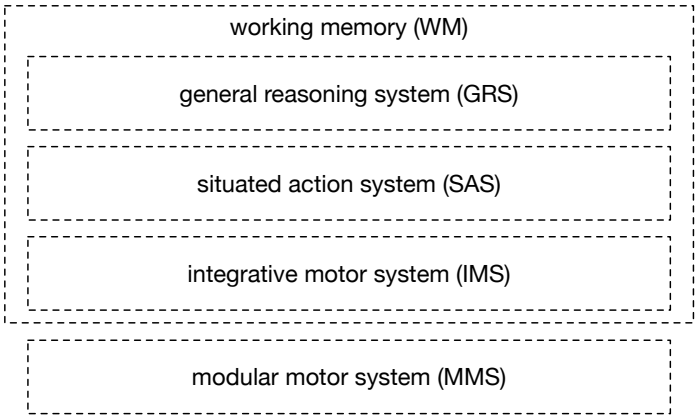

Figure 2: The interactive multilevel architecture (IMLA).

3.3 Preliminary review of implications

IMLA is a summary of recent evidence concerning the interface between cognition and motor control. It is intended to incorporate minimal theoretical interpretation, rather serving to define the target for theoretical explanation. It only addresses the broad outlines of the interface, and there is a great deal more relevant evidence that must be surveyed for a detailed picture. Nevertheless, even this much is sufficient to warrant a substantial theoretical reorientation in comparison with CV and BS-IP.

IMLA incorporates significant functional integration between the systems, contrary to CV and its picture of strong functional segregation between systems. Moreover, IMLA shows a complex pattern of functional integration. Representations and processes of the GRS, SAS and IMS can form part of conscious awareness, while representations and processes of the MMS cannot.[4] Notwithstanding the absence of co-present representations in consciousness, SAS/IMS and MMS exhibit rich functional integration. Relating this picture to BS-IP, the integration in working memory between GRS, SAS and IMS contradicts NO-CPP. If we assume that there are format differences between the general reasoning system, action system, and low-level motor system, at least some of the integration between these systems is contrary to NO-MFP.

More broadly, CV postulates a clear functional distinction between cognitive and motor representations that is questionable even for simple, familiar actions like reaching for a cup of coffee. The distinction is also dubious in light of classical theories of skill acquisition such as that of Fitts & Posner (1967). For humans learning new skills is a common occurrence and of the highest adaptive importance – the capacity to acquire a very wide range of skills is one of our defining characteristics as a species (Gibson 2002). Furthermore, even in the classical picture the attainment of high levels of automaticity takes a long time. These points imply that cognitive control is required to perform a great deal of motor control and must do so with reasonably high degrees of proficiency. On these grounds we should expect that for some cognitive representations motor control is a normal function and they are reasonably well suited for this function. IMLA confirms this and identifies two systems in working memory that jointly perform motor control functions.

[4] The claim that IMS is integrated with conscious awareness is based on the 2MS model of Keele et al. (2003), the dual task experiments cited above which show that motor planning taxes working memory, of which Weigelt et al. (2009) is representative, and the evidence described by Milner & Goodale that the ventral system participates in high level action planning.

4 A revised set of assumptions for understanding the interface

I now turn to the problem of assembling a set of theoretical resources that will allow us to explain the structure of IMLA. I will do so by contrast, examining reasoning that could be given to support the picture of cognitive architecture that underlies BS-IP, and identifying alternative assumptions that better fit IMLA. Here I focus on modularity and representational interactions across formats. Why should we find rich interactions between systems, as depicted by IMLA and contrary to the strong functional segregation view? What mechanisms support these interactions? Some interactions are likely to be non-representational, but many will be representational, and some are likely to involve translation.

In this respect some clarification is required. I'll refer to any way of relating representations as *linkage.* Some forms of linkage are perhaps too simple to be considered forms of translation, such as direct associative links between representations. It is unclear what should be considered translation, but I will take it that *systematic linkage* between distinct representational schemes counts. By this I mean that novel representations in one scheme can be linked appropriately to novel representations in the other in virtue of their content. The production of approximately the same content will be termed *narrow translation* and production of an appropriately related content as *broad translation.* When motor researchers such as Willingham (1998) claim that cognitive intentions are 'translated' into motor representations, this is best understood as broad translation because the contents are not the same, but rather related ones appropriate for the subordinate level of control.

In addition to the evidence codified by IMLA we can add a further piece of evidence which indicates that it is likely that the cognition-motor interface does involve translation. This is that fluent action control is productive: it is possible to fluently produce many novel actions within a skill domain. This in turn implies that action intentions and motor representations are each productive, and there are systematic links between them. An associationist account can't fully explain such linkage because associations can only be formed between existing representational tokens.

4.1 Butterfill's argument that format is a hard constraint on module interface structure

As described above, BS-IP explains strong segregation by means of NO-MFP. This appears to draw on Butterfill's (2007) computational theory of modularity (B-CTM). This account aims to explain the properties that Fodor (1983) attributes to modules. The central idea is representation types and process types are mutually specific (RP-MS). That is, representation types are in part defined by the kinds of processes they can participate in, and vice versa. Modules will employ different kinds of processes to the general reasoning

system, and hence have different kinds of representations. Butterfill draws the conclusion that modules and the general reasoning system cannot share representations (NO-CR).

An initial problem is that NO-CR and NO-MFP appear to be too severe to explain even classical, strong views of module encapsulation. The boundary between two systems is strongly segregated if they function largely autonomously, but this is compatible with some amount of representational interaction. As described above, motor control theorists have commonly assumed that cognitive intentions are translated into motor representations. On this version of CV the interface has a selective unidirectional opacity: motor representations are not accessed by the cognitive system, but some cognitive representations are accessed by the motor system. In contrast, NO-MFP implies a strict bidirectional opacity.[5] Translation is ruled out because a translation process would span multiple formats.[6] B&S embrace the proscription on translation, but the bidirectional opacity of NO-MFP is also at odds with classical views of the perception-cognition interface. Here there is unidirectional opacity, though in the opposite direction to that of the cognition-motor interface. That is, the perceptual system is unable to access cognitive representations, but the cognitive system takes as input perceptual representations. Thus, the Müller-Lyer illusion is interpreted as showing the insensitivity of perception to knowledge (e.g. Fodor 1983, p. 66), but the illusion depends on the cognitive system accessing perceptual representations.

In keeping with IMLA, more recent views of cognitive modularity regard modules as more highly connected than the Fodorian picture recognized (Spunt & Adolphs 2017). Theories of working memory (WM) are especially relevant because they clearly violate NO-MFP. The most influential is Baddeley's multicomponent model, which in its current version incorporates four major components: a central executive, a visuospatial sketchpad, a phonological loop, and an episodic buffer (Baddeley 2000, 2011). Each of these components is a module, but together they function as an integrated workspace. The visuospatial sketchpad and phonological loop are specialized systems for processing visual and verbal representations respectively, but these representations are integrated into a unified

[5] B&S say that "it does seem that motor representations are available in some sense" (2014, p. 134), which appears to be at odds with strict bidirectional opacity. The question is what sense, since their contents can't be available without violating NO-MFP and NO-CPP. Sinigaglia & Butterfill (2015) propose that motor representations shape experience, and they suggest that this is analogous to the way that visual experiences shape conscious experience without being inferentially integrated with cognitive representations. This proposal is problematic, however. As described in the text, the contents of visual representations are accessed in visual experience and are integrated with cognitive representations in processes of visual working memory.

[6] B&S explicitly justify NO-TRANS on the grounds that nothing is known about translation between intentions and motor representations. This can just as easily be taken as grounds for investigating translation, given that translation is widely viewed as the obvious candidate mechanism linking intentions and motor representations. In any case, NO-TRANS appears to be required by NO-MFP.

representation of the situation in the episodic buffer, while the central executive is able to control operations in any of the other three modules. Other theories of WM differ in details but also see it as encompassing representations with different formats. *Sensorimotor recruitment* theories are particularly notable in this regard because they propose that working memory activates the same representations that are involved in perception and action (D'Esposito & Postle 2015).

More fundamentally, B-CTM provides no account of how to individuate types of representations and processes. Many forms of individuation will render RP-MS false: abduction and deduction are different kinds of processes, but they can operate on the same kinds of representations, for instance. Further, it seems as though NO-MFP will prevent not only representational interactions between modules and the general reasoning system, but also module-module interactions and perceptual input to modules, thus encapsulating modules rather too well. There may be an account of representation and process type individuation that prevents this, but without an explicit analysis the theory is hard to interpret.

4.2 Module interface structure is based substantially on function rather than constraint

The strategy of explaining the structure of system architecture in terms of representational format arguably puts the cart before the horse. That is, it is more likely that format differences will be *caused by* the major determinants of architectural structure, rather than *be* the major determinant of architectural structure. In general, factors influencing the architecture of a cognitive system should exhibit the following hierarchy: function > information processing requirements > specific representational mechanisms. That is, an adapted or well-designed system should have representational mechanisms that are shaped to its functional and information processing requirements. If the representational mechanisms themselves are the primary determinant of macro-architecture structure, this will usually be because they are acting as a constraint that limits variation. An evolutionary adaptive explanation for a biological trait T1 requires plasticity of organismic form: the mechanisms by which T1 is constructed are capable of producing alternative forms (T2, T3, …), and the particular form T1 is prevalent in the population because it serves a function and has been selected for. Thus, an adaptive explanation for zebra stripes requires that the biological mechanisms of coloration in zebras are capable of producing various coloration patterns, and zebras have black and white stripes because this particular coloration pattern serves one or more adaptive functions.

Butterfill's explanation for module interface structure is based on constraint: evolution could not do otherwise than build a cognitive module interface that obeys NO-MFP. But he also claims that this interface structure is adaptive. Butterfill (2007) suggests that it is required for

robustness, while Apperly & Butterfill (2009) argue that cognitive modules are subject to the requirements of speed and efficiency, and this in turn requires that they are not exposed to the strong informational connectivity characteristic of the general reasoning system. Each claim is individually surprising, and their combination also warrants explanation. Other things being equal, we should expect that there is wide variation in the functionally optimal connectivity of modules. Some modules may benefit from minimal connectivity while others may require extensive connectivity to perform their functions. In comparison, modularity in biology more generally exhibits just this kind of variation (Melo et al. 2016).

Constraints certainly play a profound role in evolution, but it would be very surprising if the interface structure of cognitive modules is entirely based on constraint. Neural systems have been evolving for more than seven hundred million years (Roth 2013), while module interfaces are a major determinant of cognitive architecture with pervasive adaptive significance. If the functional requirements of cognitive module connectivity do vary then there has been ample time for the emergence of mechanisms that allow this. Constraint and function can coincide (Gould & Vrba 1982), but that they should do so over such a time scale and in such varying circumstances is a claim that requires extraordinary support. The modular organization of multicellular animals has been evolving over the same period and, indeed, in close interdependence with neural organization. The wide variation found in the modular structures of metazoans provides a guide to what could be expected for cognitive modularity, and, as noted above, cognitive neuroscience does indeed yield a similar picture.

According to current theories in developmental cognitive neuroscience, the organization of cognitive systems, including patterns of connectivity and segregation, is based substantially on activity-dependent plasticity (Chou et al. 2013, Johnson & de Haan 2015). This allows these patterns of connectivity to reflect function, and the information processing requirements of function. The development of ocular dominance columns in layer 4 of the primary visual cortex provides an illustration. Neurons in these columns initially receive inputs from both eyes, but during development axons from one eye withdraw, leaving a particular column responsive to just one eye (Held 1993). The effect of this is to segregate information from the two eyes in a process called *parcellation*. Parcellation at the level of layer 4 allows for more sophisticated forms of binocular integration at later levels of visual processing (Johnson & de Haan 2015). The formation of larger scale cognitive systems requires both selective segregation and the formation of long-range connections between distant neural regions. Johnson & de Haan (2015) propose that one of the mechanisms involved in the latter process is Hebbian learning – essentially, a larger scale version of a key learning process by which neurons alter their functional connections.

Experiments with ferrets illustrate the remarkable power of neural plasticity to construct novel functional organization. When projections from the retina are rerouted into auditory cortex, the affected sections of auditory cortex can take on organization characteristic of visual cortex, and the ferrets behaviourally treat signals along this unusual pathway as visual (Sharma et al. 2000, von Melchner et al. 2000).

4.3 The definition of 'format' does not preclude multi-format processes

The rationale for NO-MFP is RP-MS, but what reasoning might be given in support of RP-MS? Conversely, what theoretical grounds can be given for rejecting RP-MS?

One possible argument for RP-MS holds that it is true by definition, given the nature of formats (DFA). This is not an argument to be taken seriously, but it is useful as a starting point. More specifically, the argument is that, if we understand a format to be a system of rules that govern the interdependent structures of representations and processes that operate on them, then insofar as a cognitive system functions as a single, coherent, integrated representational system, it has a single format. Two representationally distinct systems will have different formats, by definition. And having different formats, they will be unable to interpret each other's representations. This argument is superficial, however. We can simply define a 'superformat' as a format that encompasses multiple sub-formats. Linking multiple formats is, in essence, just a matter of more format.

It is worth noting that, as a matter of fact, superformats are common. Maps, graphs, diagrams and cartoons, for instance, can include graphical and linguistic representational elements. Natural languages are also superformat systems. English, for example, includes both verbal and written representations, while verbal communication includes both gestural and facial expression components. Theories of comprehension propose that it involves multiple levels of processing which progress through different forms of representation, with the culminating representation not being linguistically structured, but rather having a non-propositional situation model format (Kintsch 1998, Ferstl 2018).

We can also deny that common format is the only way to interrelate representations. As detailed below, there are metarepresentational mechanisms for identifying representational relations across formats. These mechanisms can be used to construct linking format.

4.4 Format types do not have essences which preclude multi-format processes

A more substantial argument for RP-MS holds that formats of different types are governed by different, non-overlapping principles and so cannot be combined in a single coherent

representational system (FPA). What counts as a relevant difference in format type is explained by example, appealing to iconic and discursive representational formats. This model is apt because the iconic-discursive distinction is commonly viewed as explaining the perception-cognition contrast, and it would seem reasonable to suppose that a similar format distinction might apply to the cognition-motor interface. Specifying FPA in terms of a major format distinction of this kind leaves room for the possibility that some formats can overlap in principles, and RP-MS will hence not apply to these format differences, or apply in a weaker form that does not prevent translation or representation sharing. The contrary position is that the iconic-discursive distinction doesn't provide grounds for thinking that perception and cognition are governed by non-overlapping principles which prevent representational interaction, and there is accordingly no strong reason to suppose that such a format contrast is to be found at the cognition-motor interface.

The idea that iconic and discursive formats are governed by distinct, non-overlapping essences is common: Haugeland (1991), Fodor (2008) and Quilty-Dunn (2019) provide illustrative analyses that seek to articulate this view. Nevertheless, the position faces substantial difficulties, even on the face of it. Representations often considered to be iconic are diverse, including photographs, paintings, maps, diagrams and graphs. Problematically, iconic and discursive formats can share representational attributes. Maps, diagrams, and languages are all symbolic, for instance. It is sometimes assumed that there is a close link between format and content. B&S and Mylopoulos & Pacherie (2017), for example, argue that cognitive representation has a propositional format on the basis that it can represent propositional contents. The link is not so tight, however. Diagrammatic logic systems, such as Venn diagrams, are capable of expressive equivalence with monadic first order predicate logic (Shin 1994, Camp 2007). Johnson-Laird (1992) presents a theory of human propositional reasoning that is based on mental models rather than sentential representations. It thus cannot be inferred that cognition has a propositional format, where this is understood in the usual sense of sentential format, from the fact that cognition represents propositional contents and performs propositional inferences.

The literature on the iconic-discursive distinction as a basis for the perception-cognition distinction is too large to survey here, so I will take Quilty-Dunn's (2019) account as illustrative on the grounds that it is recent and well-informed. Quilty-Dunn (2019) argues that some perceptual representations are discursive, contrary to the idea that the contrast between perceptual and cognitive representations corresponds to the iconic-discursive distinction. Moreover, he claims that this can explain the ease with which perceptual representations can enter cognition because they can be integrated into cognition without translation. He thus provides an example of a position which holds that perceptual and

cognitive formats can overlap, though this is not based on an overlap between iconic and discursive representation. He also holds that many perceptual representations are iconic.

Problems with his characterization of the iconic-discursive distinction, however, suggest that these putative format types are not well understood and don't provide a sound basis for understanding visual representation. Quilty-Dunn posits two key differences: (1) parts of icons correspond to parts of what they represent, while this need not be true of discursive representations, and (2) icons represent holistically in the sense that parts of icons represent multiple properties at once, whereas the parts of discursive representations typically represent particular individuals and features. He illustrates this contrast by comparing the sentence "Bob Dylan is wearing a checkered shirt" with a photograph of Bob Dylan wearing a checkered shirt. The holistic criterion for iconicity is apparent in the photograph in the way that the spatial extent, shape, texture and color of Dylan's shirt are all represented together by sections of the pixel array.

Quilty-Dunn points to evidence for object file representations in visual perception as evidence that some perceptual representations are discursive. These representations track objects over time as integrated entities with bound features, and they can maintain tracking across dramatic changes in features, such as when a cartoon frog transforms into a prince (Kahneman et al. 1992). In other words, visual perception represents objects discretely. Quilty-Dunn describes a variety of forms of evidence for iconicity in perceptual representation. One kind is topographic representation in sensory cortices, where for instance edge orientation and spatial location are bound. Another is phenomena associated with mental imagery, such as mental rotation and image scanning, where processing speeds correspond to spatial structure of the objects represented.

That rotation and scanning of mental images respect spatial structure doesn't show that mental images represent object features in a way that is inextricably bound, however. In controlled mental imagery it is possible to arbitrarily modify individual features of a mental image, such as color, shape and size. Features are thus representationally separable, even if some kinds of operations treat feature complexes as units. This is further supported by neuroscientific evidence concerning visual processing. It is true that the retinotopic mapping of cells in V1 structurally combines retinal location and oriented line segment information. This is a very narrow feature combination, however, and it is quite different to the way that photographs represent.

To understand visual representation we need to consider the structure of visual processing architecture. Hubel & Wiesel's (1962) model of hierarchical representation in the visual system remains a foundation for contemporary computational models of visual processing

(Mély & Serre 2017). This model describes a three-level hierarchy, the base of which is the center-surround receptive field of retinal ganglion cells, which discriminate point light sources in a retinotopic framework. At the next level, simple cells in V1 have a receptive field structure that makes them sensitive to oriented line segments at a retinal location, achieved by integrating over multiple ganglion cells (via the lateral geniculate nucleus). Complex cells are also sensitive to oriented line segments, but are less sensitive to specific retinal location. This discrimination is based on integrating across simple cells with receptive fields that have the same orientation sensitivity but differing retinal location.

This hierarchical organization incorporates principles that are thought to be fundamental to visual perception. *Invariance* and *selectivity* are both achieved through appropriate forms of integration over lower level representations. Invariance is a loss of certain kinds of specificity, evident in the reduced sensitivity of complex cells to specific retinal location. Ultimately, the representation of features of the world requires representation in world-based coordinates, not retina-based coordinates. This is achieved through progressive abstraction through integration, and at higher levels of the visual system neurons show sensitivity to specific features anywhere in the visual field. Selectivity is achieved by integrating over features represented at lower levels so as to differentiate a more complex, specific feature. This is evident in the integrative properties of simple cells. According to current theories, these principles are the basis for the representation at higher levels of the visual system of complex features, like shape, color, motion, objects of various kinds, scenes, and so on. These principles aren't exhaustive, however, because scene and object segmentation also incorporate other processes, such as 'filling in' (e.g. Lee & Nguyen, 2001).

Visual representation thus has a structure contrary to the spirit of Quilty-Dunn's second criterion for iconicity and is very unlike that of photographs. Photographs, understood narrowly as physical objects or digital representations on a screen, don't 'represent together' features such as the shape and texture of Bob Dylan's shirt; they don't represent these features at all. Photographs, strictly, represent variations in color by means of a color point array.[7] Features such as objects, shape, and texture are latent in this array and differentiated by downstream processes in the visual system of the individual looking at the picture. In contrast to photographs, visual representations exhibit specific selectivity for a wide range of features, achieved through the hierarchical organization described.

Three points can be drawn from this. The first is that Quilty-Dunn's second criterion presents the contrast between 'holistic' and particulate, feature-specific representation as a dichotomy when it can be a matter of degree. The 'holism' of even early visual

[7] I'm including greyscale as 'color'.

representations is highly selective, with ‘holism’ becoming increasingly restricted at higher levels of processing. But if we amend the second criterion accordingly the distinction between iconic and discursive representations is no longer clear cut.

The second is that Quilty-Dunn’s analysis doesn’t provide strong grounds for viewing iconic and discursive representation as well-understood natural kinds. If these were well understood natural kinds then classification should be relatively straightforward for a significant number of cases (though not necessarily all), and classifying a particular case as belonging to one of the categories should allow us to explain many of its features. Thus, whales can be clearly identified as mammals, and this classification allows us to explain many of the features of whales. In contrast, visual representation is not readily assimilated to either iconic or discursive representation, as Quilty-Dunn conceptualizes them, and it remains to be shown that either category provides a substantial basis for explaining many of the features of some or all forms of visual representation.

The third point is that the feature-oriented nature of visual representation provides overlap with cognitive representational formats, which, in the form of mental models or linguistically structured representations, often employ constituents with particulate format and feature-specific content. Indeed, it may allow visual representations to be constituents of mental models in visual working memory, consistent with sensorimotor recruitment theories of working memory as described above, which claim that working memory can employ the same representations that are involved in perception and action. This is consistent with Quilty-Dunn’s idea that there is format overlap between perception and cognition, but it goes further.

In light of these considerations, FPA lacks support. The idea that iconic and discursive formats are non-overlapping is particularly difficult to sustain, and efforts to use these conceptions to explain the perception-cognition distinction face serious difficulties. Since the claim that cognitive and motor representations have non-overlapping essences draws its inspiration from the iconic-discursive contrast, and its application to the perception-cognition distinction, there isn’t a strong reason to expect that such a format divide can be found at the cognition-motor interface. In contrast, there is extensive empirical support for the view that there is overlap between perceptual and cognitive formats, and that this allows representational interaction.

Overlap between representational systems can be sufficiently great that they share representational resources. Evidence from humans indicates that different modalities can share representations of space. Two areas of the brain, the parahippocampal place area (PPA) and the retrosplenial complex (RSC), have been found to respond selectively to visual

scenes in contrast with objects. Wolbers et al. (2011) found that the PPA and RSC were also activated when scenes were perceived haptically (the scenes were presented as Lego dioramas). They ruled out the possibility that this activation was based on the formation of visual images of the scenes during touch-based scene perception in part by demonstrating PPA and RSC activation in congenitally blind participants while performing the task. According to some prominent theories of visual perception (Hochstein & Ahissar 2002) and visual working memory (Logie & Della Sala 2012), conscious visual awareness is conceptual. That is, conscious visual representations of space, objects, and object relations are represented in a way that employs concepts such as 'teapot', 'kitchen bench', 'on' and so on. This claim is based on evidence that conscious perception involves activation of the same representations in long-term memory involved in conceptual representation. Thus, on Hochstein & Ahissar's 'reverse hierarchy' theory conscious perception is based on top-down 'analytic' processes that come after bottom-up processing along the full perceptual processing hierarchy, which includes areas involved in representation of long-term semantic memory.[8] The conceptuality of these representations can help link them to representations in other formats that also employ these concepts, such as propositional representation of the information that the teapot was given to you as a present by a friend.

With this as background, we can identify potential forms of overlap that can help explain the pattern of interactions identified by IMLA. Mylopoulos & Pacherie (2017) characterize the contrast between cognitive and motor representations in a way that is consistent with CV and FPA, emphasizing a fundamental contrast in the format and principles of each. Intentions are propositional and governed by principles of rationality while motor representations represent information relevant to action and are governed by principles concerning efficient control of movement. But Pacherie's own three-level model of action control (Pacherie 2008, Mylopoulos & Pacherie 2018) suggests a more complex picture. This distinguishes between distal, proximal, and motor levels of control. Proximal control, which roughly corresponds to SAS in IMLA, should represent objects, features and spatiotemporal relations immediately relevant to action control. Moreover, these representations should reflect efficiency principles which govern efficient action control.

Putting the overlap in terms of IMLA, we should expect that MMS, IMS and SAS all represent objects, features and spatiotemporal relations immediately relevant to action control. They will do so in ways that are not identical, suited to the information process requirements of their own level of control. SAS and IMS are likely to share representational resources, in particular semantic memory. Translation will at minimum be required across the

[8] Whether perceptual awareness is or can be conceptual has been a matter of controversy amongst philosophers (e.g. Burge 2010, Siegel 2010, Block 2014). The debate is too large to address here, but the evidence described in the text supports the view that it can be.

allocentric-egocentric division between the ventral and dorsal systems. But representational overlap between these systems provides a substantial basis for translation.

4.7 Multiformat integration is not so expensive that it is rare

A different kind of argument might be made to arrive at a picture of modularity congenial to BS-IP, drawing on the same broad theoretical context it is informed by. This argument proposes that multi-format processes are not impossible, but rather rare because they are expensive (FEA). Relating representations across formats is, in a sense, 'just' a mapping problem, but it requires infrastructure which is costly, and prohibitively so in the case of mapping between the general reasoning system and modules. The contrary position is that relating representations across formats is not so expensive that it is rare, and it occurs between modules and the general reasoning system based on information processing requirements.

One factor that contributes to the expense of translation is the problem of finding the appropriate mappings across representational systems. A representational format is itself a very difficult achievement: a system of rules that are so constructed as to respect contents of a certain kind. The more complex and precise the contents are, the more elaborate and precise must be the rule system. Correspondingly, particular contents will depend in elaborate, variable ways on the rule system. Respecting content is thus already a hard problem within a rule system specialized for a particular kind of content. Translation must respect content across two rule systems specialized for different kinds of content. Viewed as a search problem, then, the problem of translation appears formidable. From a purely structural perspective, the translator must find correspondences between contents that in each system depend in complex ways on a complex rule system. Moreover, a translator will require criteria for determining correspondences and finding such criteria will be difficult, since it amounts to the construction of a metarepresentational system that encompasses the contents of the two zero-order systems.

These problems are made more challenging when there is a high degree of *representational distance* between two representational systems. FEA recognizes that representational systems can overlap, but the degree to which they do so varies. The representational distance between two representational systems is determined by similarities and dissimilarities that make it easy or hard to map between them. Factors that affect similarity can include both type of content and type of format. For example, the representational distance between English and Spanish is much closer than between English and Mandarin because English and Spanish are structurally and conceptually/culturally much more similar

to each other than English is to Mandarin. It will accordingly be a great deal more difficult to translate a Jane Austen novel into Mandarin than into Spanish.

A picture can be constructed based on these ideas that allows perceptual input and limited module-module interactions while preventing sharing of representations between modules and the general reasoning system. Modules are separated by relatively small representational distances, making the cost of translation adaptively feasible, though perhaps only in limited forms. The representational distance between modules and the general reasoning system, however, is too large for translation because of the dramatically greater expressive capacity of the latter. Indeed, translation between the general reasoning system and a module might seem to require an especially implausible homunculus. This translator should 'understand' the representations of the general reasoning system, commensurate with conscious understanding, and also the representations of the module, and be able to map between them. Such a system would thus have greater representational power than conscious cognition.

The argument has a number of problems, which can be itemized as follows:

P1: It rules out the unidirectional opacity attributed to visual perception by the classic modularity picture. We can try to accommodate this by further proposing that, because visual content is restricted while propositional content is not, mapping from the former to the latter is feasible whereas the reverse is not. This makes it unclear why we should find the bidirectional opacity attributed to the cognitive-motor interface, however.

P2: FEA supposes that the zero-order problem of constructing a system of content-respecting rules is very difficult. However, no detailed analysis is given of the complexity of representational systems and the difficulty will depend on the resources available and the complexity and precision required.

P3: FEA focuses on cost without considering benefit. To determine whether a trait will be adaptively favored we need to assess the cost in relation to the benefit. A trait that is extremely resource intensive can still be selected for if it provides enough adaptive benefit – the primate visual system provides a salient illustration of this. If cross-format linkage is adaptively valuable then substantial resources might be devoted to it.

P4: FEA supposes that the cost of linkage is based on the overall representational complexity of the two systems linked, without considering the possibility that linkage is restricted to a subset of the representations of each system. In fact, the human cognition-motor interface cannot generate a direct link between any conceivable action intention and

an appropriate corresponding motor representation. It also cannot generate a link from any motor representation to a corresponding cognitive representation. But neither of these abilities is needed for a functional interface. The human cognition-motor interface *can* generate direct intention-motor mappings for fluent action abilities. It can also construct novel actions of kinds that are not familiar, with fluency declining as the nature of the action is increasingly unfamiliar. These basic facts indicate that only a restricted set of cognitive and motor representations are employed by the interface. This simplifies the linkage problem.

P5: FEA assumes that the cost of linkage is based on the complexity of the respective systems in their mature states, neglecting the possibility that linkage might be constructed progressively, beginning with simple action abilities in which the correspondences are comparatively simple, and gaining the ability to form more complex correspondences only slowly as action abilities gain greater complexity and sophistication.

P6: FEA regards the finding of metarepresentational criteria for content-relatedness as a difficult problem, and hence costly. However, there are candidate criteria which are simple and cheap. One such is temporal correlation of activation, employed by Hebbian learning. Other criteria may be more complex and less cheap, but nevertheless 'affordable'. Templates provide one kind of example, discussed below. Templates, in turn, can be used as the basis for constructing coordinated frameworks. Coordinated frameworks, in turn, can be employed for tracking systematic content relations across representational systems.

To expand on P2, if mental representation depends on rule systems that have similar complexity and precision requirements to formal languages like logics or computer languages then the construction of these systems will indeed be very difficult. But this difficulty is a strong reason for doubting that these are good models for mental representation. Logics and computer programs are not tolerant of variation in their fundamental components: random variations in the definitions of symbols and rules in such systems will usually not result in a functional system. In contrast, biological systems evolve through incremental hill climbing and exhibit high tolerance to variation in components. Variation of traits is the norm in biology (Dobzhansky 1962, Hallgrímsson & Hall 2005).

How is it that complex biological systems can tolerate substantial variation in component structure despite exhibiting complex, holistic dependencies? In part, because they must. Designed systems, like computer languages and airliners, can be constructed that have extremely precise dependencies, or in other words, very low tolerance for variation. Consider the tolerances of the components of the engines of a modern airliner, for example. This is a realm of function space that is simply unavailable to biological evolution. Biological evolution

is restricted to the construction of complex systems that are tolerant of variation. This implies that they are subject to norms that are permissive: they allow substantial variations from the ideal. Indeed, the nature of incremental evolution indicates that biological systems can be adaptive while being very far from applicable ideal norms. Consider, for instance, the functioning of the earliest animal eyes in comparison with function ideals for visual perception. If the norms applicable to mental representation are permissive, then coarse approximations can be constructed that are still useful enough to be adaptively valuable. Such approximations can be refined incrementally, both through evolutionary adaptation and developmentally by means of neural plasticity mechanisms.

Another feature that is crucial in allowing biological systems to tolerate variation is that they are self-regulating (Elman et al. 1996). Neural plasticity again supports this for mental representation. The ferret example discussed above indicates that mental representation does indeed exhibit a high degree of plasticity and self-regulation. This example also bears on P3. The infrastructure supporting neural plasticity is undoubtedly not cheap, requiring extensive genetic and cellular machinery and being metabolically resource intensive. But it has powerful adaptive benefits and is evidently worth the cost. The adaptive benefits are worth emphasizing. For example, plasticity helps to ameliorate an apparent problem facing the evolution of perception. This looks like it should be difficult because each adaptive change seems to require at least three simultaneous genetic modifications: a change that alters the sensory periphery in a way that provides a novel type of information, a change which modifies the channel to transmit this information for central processing, and a change to central processing which interprets the new information. Based on the ferret example we can see how activity-dependent plasticity can facilitate evolutionary change by allowing perceptual systems to ‘figure out’ how to interpret new information.

This point bears on P6. The rules of neural plasticity include, in effect, meta-rules for constructing representational systems. Taking into account the ability to flexibly construct links between regions, described above, they appear to include metarepresentational criteria needed for constructing linkage between representational systems. Some of these are simple, such as Hebbian correlation. Others will be more complex, and finding them may not have been easy, from an evolutionary perspective. But, as noted above, neural systems have been evolving for more than seven hundred million years, and in light of the profound role of plasticity in animal evolution it is plausible that a rich set of these rules has been acquired.

Templates provide a basis for more complex forms of linkage than does simple Hebbian learning, and their operation is illustrated in the maintenance of alignment between visual and auditory representations of space (Knudsen 2002). The auditory representation of space in owls is based on the discrimination of cues such as interaural timing differences and

interaural level differences. Values of these cues are associated with locations in space. The computation of these values is affected by the fact that the size and shape of the head varies between individuals and for an individual as it matures. It is also affected by factors such as hearing loss. The maintenance of spatial mapping has been studied by altering auditory and visual perception in owls and examining the effects. Audition has been modified by blocking an ear with an earplug, while vision has been modified by means of spectacles which displace the visual field. In response to these modifications young owls initially mislocalize perceptual sources, but regain accuracy after several weeks of experience. The neural basis for these behavioral changes includes neural remapping in the midbrain spatial localization pathway. For instance, after horizontal displacement of the visual field, neurons in several areas shift their auditory receptive fields by an amount that corresponds to the visual field displacement. Neural modification includes a change in axonal projections, which are topographic – new projections are formed which reflected the altered topographic relationships. Changes in auditory mapping are based on an instructive signal from the visual system in the form of a spatial template – a topographic template of the visual field. In other words, the visual and auditory fields are aligned with each other in a continuously updated manner by means of signals which inform each region about the structure of the other.

The coordination of coordinate frameworks has particular significance for understanding translation because it provides a basis for understanding how systematic relations between different representational systems can be tracked. Spatial frameworks are especially straightforward (and hence easy to study) and may serve as the basis for constructing more abstract frameworks for coordinating representations across systems. For example, coordination between the egocentric and allocentric spatial representations of the dorsal and ventral systems may allow the discovery of more complex information relations across the systems, and the construction of functionally appropriate linkage.

What of the problem of representational distance? A solution that has been employed in theories of action control architecture is multi-level hierarchy (Lashley 1951, MacKay 1982, 1987, Fitch & Martins 2014). To illustrate, take verbal language production as our example and consider the very large representational distance between the uppermost and lowest levels of the hierarchy. At the top are communicative intentions while at the bottom are the representations that control muscle activations of the mouth and tongue in sound production, and the facial expressions and gestures involved in the visual components of verbal communication. Mapping directly from the former to the latter would seem to present formidable difficulties due to the profound differences between these representational domains. These difficulties are ameliorated, however, by the interposed roles of intermediate levels of the hierarchy. A communicative intention to express an idea, for example, can often be achieved by a variety of verbal expressions. A sentence planning process constructs a

particular verbal expression that will fit the communicative intention. A word selection process selects words that are suitable for the sentence structure. A morphological selection process selects morphological representations appropriate to words. Phonemic planning determines the components of these morphological representations. Motor planning, in turn, determines the movements that will produce the sounds. And these representations determine what patterns of muscle activation are required.

The representational distance between representations at each level is relatively small. Consider the relatively straightforward relations between lexical, morphological and phonemic representations, and between phonemic and motoric representations. On close examination these relations are substantially more complex than they may seem intuitively, as revealed by the complexity of the relevant theories in psycholinguistics (e.g. Wheeldon & Konopka 2018), but the mapping problems are nevertheless clearly tractable enough to allow rapid, largely automated cross-level (and cross-format) linkage. Perceptual processing exhibits a similar hierarchical structure, although substantially bottom-up rather than top-down.

Conclusions

The preceding discussion has provided us with a set of resources for explaining the structure of the cognition-motor interface. To review and further elaborate, interfaces between cognitive systems are likely to be based substantially on function rather than constraint, meaning that patterns of integration and segregation will reflect the functions and information processing requirements of the various systems to a significant degree. Processes of neural plasticity can construct representational connections between systems where this is useful and create segregation where they are not. Empirical evidence indicates that cognitive systems can exhibit a wide range of patterns of modularity, including the integration of multiple modules and multiple representational formats in an integrated representational workspace. Neural plasticity is in part based on metarepresentational criteria which support the construction of representational systems and linkage between them. Consideration of the evolutionary context, coupled with evidence that neural plasticity is extremely powerful, indicates that the brain is endowed with a rich and sophisticated set of metarepresentational mechanisms.

Representational formats are often open, in the sense that they are not tightly restricted to a specific class of information processing principles. The information processing principles applying to cognition and motor control are also open, in the sense that they can permit a range of solutions. Together, these points have several implications. Firstly, different formats can share contents and information processing principles. Secondly, a given format can be

progressively refined from poor (but still useful) approximations to key principles to better (but perhaps rarely perfect) approximations. Thirdly, a given format can encompass multiple kinds of information processing principles. These features of mental representation enable superformats and translation.

Representational distance is an important constraint on linkage between systems, and one solution is multi-level hierarchy. These hierarchies are evident both in bottom-up processes of perception and top-down processes of action control. The more elaborate hierarchy of IMLA, in comparison with CV, likely reflects this strategy.

Progressive learning processes are another important ingredient for understanding how linkage can be constructed. The basic structure of the cognition-motor interface will be established in infancy as fundamental sensory-motor-cognitive relations are discovered/constructed. Action abilities in the earliest stages of development are extremely rudimentary. The sensory, motor and cognitive signals that must be integrated will be very noisy but also fairly simple. The complexity of the information that must be integrated will increase slowly, in a manner interdependent with advances in action abilities. The interfaces involved in action control will continue to develop in adulthood as new skills are acquired. Skill-specific interfaces will build on more general representational abilities and support the information integration needed for domain-specific representations and control.

A task that remains is to employ these resources to explain the details of the structure of IMLA. I tackle this in a follow-up paper, which will also compare the solution proposed there to other solutions to the interface problem that have been recently advanced.

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. (2011). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29.

Beilock, S. L., & Carr, T. H. (2001). On the Fragility of Skilled Performance: What Governs Choking Under Pressure? *Journal of Experimental Psychology: General*, *130*(4), 701–725.

Binkofski, F., & Buxbaum, L. J. (2013). Two action systems in the human brain. *Brain and Language*, *127*(2), 222–229.

Block, N. (2014). Seeing-As in the Light of Vision Science. *Philosophy and Phenomenological Research*, *89*(3), 560–572.

Bullier, J., Schall, J. D., & Morel, A. (1996). Functional streams in occipito-frontal connections in the monkey. *Behavioural Brain Research*, *76*(1), 89–97.

Burge, T. (2010). *Origins of Objectivity*. Oxford University Press, USA.

Butterfill, S. (2007). What Are Modules and What Is Their Role in Development? *Mind & Language*, *22*(4), 450–473.

Butterfill, S. A., & Sinigaglia, C. (2014). Intention and Motor Representation in Purposive Action. *Philosophy and Phenomenological Research*, *88*(1), 119-145.

Camp, E. (2007). Thinking with Maps. *Philosophical Perspectives*, *21*, 145–182.

Carey, D. P., Harvey, M., & Milner, A. D. (1996). Visuomotor sensitivity for shape and orientation in a patient with visual form agnosia. *Neuropsychologia*, *34*(5), 329–337.

Chou, S.-J., Babot, Z., Leingärtner, A., Studer, M., Nakagawa, Y., & O'Leary, D. D. M. (2013). Geniculocortical Input Drives Genetic Distinctions Between Primary and Higher-Order Visual Areas. *Science*, *340*(6137), 1239–1242.

Christensen, W., Sutton, J., & McIlwain, D. J. F. (2016). Cognition in Skilled Action: Meshed Control and the Varieties of Skill Experience. *Mind & Language*, *31*(1), 37–66.

Christensen, W., Sutton, J., & Bicknell, K. (2019). Memory systems and the control of skilled action. *Philosophical Psychology*, *32*(5), 692–718.

Clark, D., & Ivry, R. B. (2010). Multiple systems for motor skill learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 461–467.

Cooper, R., & Shallice, T. (2000). Contention Scheduling and the Control of Routine Activities. *Cognitive Neuropsychology*, *17*(4), 297–338.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37.

Dobzhansky, T. (1962). *Mankind Evolving*. New Haven: Yale University Press.

Elman, J. L., Bates, E. A., & Johnson, M. H. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press.

Ferretti, G., & Caiani, S. Z. (2019). Solving the Interface Problem Without Translation: The Same Format Thesis. *Pacific Philosophical Quarterly*, *100*(1), 301–333.

Ferstl, E. (2018). Text comprehension. In S.-A. Rueschemeyer & M. G. Gaskell (Eds.), *The Oxford Handbook of Psycholinguistics*. Oxford University Press.

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, *1316*(1), 87–104.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Wadsworth.

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press.

Fodor, J. A. (2008). *LOT 2: The Language of Thought Revisited*. OUP Oxford.

Gibson, K. R. (2002). Evolution of Human Intelligence: The Roles of Brain Size and Mental Construction. *Brain, Behavior and Evolution*, *59*(1–2), 10–20.

Gould, S. J., & Vrba, E. S. (1982). Exaptation—A Missing Term in the Science of Form. *Paleobiology*, *8*(1), 4–15.

Grol, M. J., Majdandžić, J., Stephan, K. E., Verhagen, L., Dijkerman, H. C., Bekkering, H., … Toni, I. (2007). Parieto-Frontal Connectivity during Visually Guided Grasping. *Journal of Neuroscience*, *27*(44), 11877–11887.

Hallgrímsson, B., & Hall, B. K. (2005). *Variation: A Central Concept in Biology*. Elsevier.

Haruno, M., Wolpert, D. M., & Kawato, M. (2003). Hierarchical MOSAIC for movement generation. In T. Ono, G. Matsumoto, R. R. Llinas, A. Bethoz, R. Norgren, H. Nishijo, & R. Tamura (Eds.), *Excepta Medica International Congress Series* (Vol. 1250). Amsterdam: Elsevier Science.

Haugeland, J. (1991). Representational genera. In W. Ramsey, D. Rumelhart, & S. Stich (Eds.), *Philosophy and Connectionist Theory*. New York: Psychology Press.

Held, R. (1993). Development of binocular vision revisited. In M. H. Johnson (Ed.), *Brain development and cognition: A reader.* (pp. 159–166). Oxford: Blackwell.

Hochstein, S., & Ahissar, M. (2002). View from the Top: Hierarchies and Reverse Hierarchies in the Visual System. *Neuron*, *36*(5), 791–804.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154.

Johnson-Laird, P. N., Byrne, R. M., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, *99*(3), 418–439.

Johnson, Mark H., & Haan, M. de. (2015). *Developmental Cognitive Neuroscience: An Introduction*. John Wiley & Sons.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. Cognitive Psychology, *24*(2), 175–219.

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review*, *110*(2), 316–339.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge University Press.

Knudsen, E. I. (2002). Instructed learning in the auditory localization pathway of the barn owl. *Nature*, *417*(6886), 322–328.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–131). New York: Wiley.

Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences*, *98*(4), 1907–1911.

Logan, S. W., & Fischman, M. G. (2015). The death of recency: Relationship between end-state comfort and serial position effects in serial recall: Logan and Fischman (2011) revisited. *Human Movement Science*, *44*, 11–21.

Logie, R. H., & Della Sala, S. (2012). Disorders of visuospatial working memory. In P. Shah & A. Miyake (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 81–120). Cambridge University Press.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review*, *89*(5), 483–506.

MacKay, D. G. (1987). *The Organization of Perception and Action: A Theory for Language and Other Cognitive Skills*. Springer-Verlag.

Masters, R. S. W. (1992). Knowledge, knerves and know-how: The role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. *British Journal of Psychology*, *83*(3), 343–358.

Melchner, L. von, Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, *404*(6780), 871–876.

Melo, D., Porto, A., Cheverud, J. M., & Marroig, G. (2016). Modularity: Genes, development and evolution. *Annual Review of Ecology, Evolution, and Systematics*, 47, 463–486.

Mély, D. A., & Serre, T. (2017). Towards a Theory of Computation in the Visual Cortex. In *Computational and Cognitive Neuroscience of Vision*.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785.

Milner, B. (1962). Les troubles de la memoire accompagnant des lesions hippocampiques bilaterales. *Physiologie de l'hippocampe*, *107*, 257–272.

Milner, D., & Goodale, M. (1995). *The Visual Brain in Action*. Oxford, UK: Oxford University Press.

Montero, B. G. (2016). *Thought in Action: Expertise and the Conscious Mind*. Oxford University Press.

Mylopoulos, M., & Pacherie, E. (2017). Intentions and Motor Representations: The Interface Challenge. *Review of Philosophy and Psychology*, *8*(2), 317–336.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*(1), 179–217.

Pisella, L., Binkofski, F., Lasek, K., Toni, I., & Rossetti, Y. (2006). No double-dissociation between optic ataxia and visual agnosia: Multiple sub-streams for multiple visuo-manual integrations. *Neuropsychologia*, *44*(13), 2734–2748.

Quilty-Dunn, J. (2019). Perceptual Pluralism. *Noûs*.

Rizzolatti, G., & Matelli, M. (2003). Two different streams form the dorsal visual system: Anatomy and functions. *Experimental Brain Research*, *153*(2), 146–157.

Rossetti, Y., Pisella, L., & McIntosh, R. D. (2017). Rise and fall of the two visual systems theory. *Annals of Physical and Rehabilitation Medicine*, *60*(3), 130–140.

Roth, G. (2013). *The Long Evolution of Brains and Minds*. Springer Science & Business Media.

Schenk, T., Franz, V., & Bruno, N. (2011). Vision-for-perception and vision-for-action: Which model is compatible with the available psychophysical and neuropsychological data? *Vision Research*, *51*(8), 812–818.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*(4), 225–260.

Sharma, J., Angelucci, A., & Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature*, *404*(6780), 841–847.

Shepherd, J. (2017). Skilled Action and the Double Life of Intention. *Philosophy and Phenomenological Research*, *98*(2), 286–305.

Shepherd, J. (2018). Intelligent action guidance and the use of mixed representational formats. *Synthese*.

Shin, S.-J. (1994). *The Logical Status of Diagrams*. Cambridge University Press.

Siegel, S. (2010). *The Contents of Visual Experience*. Oxford University Press.

Sinigaglia, C., & Butterfill, S. A. (2015). On a puzzle about relations between thought, experience and the motoric. *Synthese*, 1–14.

Spiegel, M. A., Koester, & Schack, T. (2013). The functional role of working memory in the (re-)planning and execution of grasping movements. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1326–1339.

Spunt, R. P., & Adolphs, R. (2017). A new look at domain specificity: Insights from social neuroscience. *Nature Reviews Neuroscience*, *18*(9), 559–567.

Squire, L. R. (2009). Memory and Brain Systems: 1969–2009. *Journal of Neuroscience*, *29*(41), 12711–12716.

van Polanen, V., & Davare, M. (2015). Interactions between dorsal and ventral streams for controlling skilled grasp. *Neuropsychologia*, *79, Part B*, 186–191.

Verhagen, L., Dijkerman, H. C., Medendorp, W. P., & Toni, I. (2013). Hierarchical Organization of Parietofrontal Circuits during Goal-Directed Action. *Journal of Neuroscience*, *33*(15), 6492–6503.

Weigelt, M., Rosenbaum, D. A., Huelshorst, S., & Schack, T. (2009). Moving and memorizing: Motor planning modulates the recency effect in serial and free recall. *Acta Psychologica*, *132*(1), 68–79.

Wheeldon, L. R., & Konopka, A. (2018). Spoken Word Production: Representation, retrieval, and integration. In S.-A. Rueschemeyer & M. G. Gaskell (Eds.), *The Oxford Handbook of Psycholinguistics* (2nd ed.). Oxford University Press.

Wolbers, T., Klatzky, R. L., Loomis, J. M., Wutte, M. G., & Giudice, N. A. (2011). Modality-Independent Coding of Spatial Layout in the Human Brain. *Current Biology*, *21*(11), 984–989.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882.

Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review*, *105*(3), 558–584.