UNIVERSITY OF COPENHAGEN

The Perception-Action Model

Counting Computational Mechanisms

Grünbaum, Thor

The Perception-Action Model: Counting Computational Mechanisms

Thor Grünbaum (University of Copenhagen)

Abstract

Milner and Goodale's Two Visual Systems Hypothesis (TVSH) is regarded as common ground in recent discussions of visual consciousness. A central part of TVSH is a functional model of vision and action (a functional perception-action model, PAM for short). In this paper, I provide a brief overview of these current discussions and argue that there is ambiguity between the strong and the weak version of PAM. I argue that, given a standard way of individuating computational mechanisms, the available evidence cannot be used to distinguish between these versions. This not only has consequences for philosophical theories of the role of visual consciousness but also for the role of experimental evidence in model testing in cognitive neuroscience.

1. Introduction

Milner and Goodale's theory of vision and action—the so-called Two Visual Systems Hypothesis (TVSH, for short)—has received a great deal of attention in cognitive neuroscience and philosophy of mind. TVSH plays a central role in philosophical discussions of the nature of perception and its content (Matthen, 2005;

Briscoe, 2008; Wu, 2014), the nature and function of consciousness (Clark, 2001), the nature of action and control (Wallhagen, 2007; Clark, 2007; Mole, 2009; Wu, 2013) and the nature and status of folk-psychology (Bermudez, 2006; Spencer, 2007; Grünbaum, 2012). For instance, Matthen (2005, e.g., p. 313) uses Milner and Goodale's TVSH to support the claim that visual object experience is characterised by a 'feeling of presence'. Clark (2001) has used TVSH to argue that visual consciousness is for high-level cognition and plays no role in 'on-line' control of action. And the view has also been used by Bermudez (2006) and Spencer (2007) to argue for the falsity of our folk-psychological explanations of object-oriented actions in terms of visual experience of objects and conscious intentions.

These philosophical debates have been preoccupied predominantly with the issue of consciousness—issues such as the role of consciousness in control of action and the neural correlates of consciousness (e.g. Kozuch, 2015). For the most part, they have been conducted as though the functional model of perception and action at the heart of TVSH (i.e. the functional perception-action model, or 'PAM' for short) had been uncontroversially specified. In fact, however, there is ongoing debate in the cognitive-psychological and neuroscientific literature about how best to specify PAM (Goodale, 2014; Schenk and McIntosh, 2010); moreover, it remains unclear how the specification of PAM relates to claims about consciousness. These ambiguities motivate the present paper. My aims are twofold: first and foremost, I identify a basic methodological problem inherent in studies designed to identify the correct version of PAM. Specifically, I demonstrate that scientific and philosophical discussions of PAM commonly assume a particular view of the individuation of computational mechanisms, and argue that if we assume this view of individuation, behavioural evidence cannot be used to adjudicate between competing formulations of PAM. Secondly, if the argument of this paper is sound, it has consequences for the use of Milner and Goodale's perception-action model to support strong claims about the architecture of the mind and the role of consciousness.

In what follows, I will, in section 2, briefly introduce Milner and Goodale's TVSH and argue that we should be careful to distinguish between a strong and a weak version of the functional perception-action

model (PAM). In sections 3 and 4, I argue that this ambiguity is due to a particular view of individuation of computational mechanisms (the 'received view'). In section 3, I sketch this common view of individuation, and, in section 4, I argue that given this view of individuation, we cannot adjudicate between weak and strong PAM based on behavioural evidence. In section 5, I consider whether the under-determination problem can be solved by using data from neuropsychology and brain imaging and, in section 6, I sketch a possible way to count the computational mechanisms involved in action and cognition that does not rely on the received view of individuation. In the final section, I return to the question of how the methodological problem concerning choice of PAM in cognitive neuroscience affects the philosophical debates about consciousness.

2. PAM, Strong and Weak

The Two Visual Systems Hypothesis (TVSH) (Milner and Goodale, 2006) consists of three parts:

1. An anatomical model, according to which visual processing pathways are anatomically divided into two separate streams: the dorsal stream projecting to the parietal cortex and the ventral stream projecting to the temporal cortex.

2. A functional perception-action model (strong PAM), according to which the mechanism computing the representations of object-features driving grasping is independent of the mechanism computing the representations of object-features driving cognition.

3. A set of claims about neural substrates of consciousness, according to which only ventral stream processing can produce conscious visual experience.

Roughly speaking, discussions in cognitive neuroscience about TVSH have mainly concerned PAM (TVSH part 2), whereas philosophers have been more concerned with consciousness (TVSH part 3). This division of labour has masked two important issues. First, it is unclear how to determine the correct functional perception-action model (PAM). Second, it is unclear how determination of the correct version of PAM

affects theories of consciousness. The main aim of this paper is to discuss the first issue (how should we distinguish empirically between strong and weak PAM?) and to add a number of remarks concerning the implications for the philosophical debates about TVSH and consciousness.

The general shape of Milner and Goodale's model of perception and action (PAM) is well known. The core elements of the model are captured nicely in the abstract of their seminal 1992 paper:

> ...the set of object descriptions that permit identification and recognition may be computed independently of the set of descriptions that allow an observer to shape the hand appropriately to pick up an object (Goodale and Milner, 1992).

By 'set of object descriptions', the authors mean set of representations of features such as the object's size, location and orientation. The general idea is that as visual information enters into the system through the retina, it is sent in almost raw form to early visual areas in the striate cortex. In the striate cortex, the analysis of the visual information begins. The story about this analysis is complex (Felleman and Van Essen, 1991) but one simple and attractive idea is that at some point this analysis divides into two functionally separate streams (Milner and Goodale, 2006). One stream, the dorsal stream, has the job of computing object descriptions that can be used in the control of manual action (in what follows, I will focus on grasping), whereas the other stream, the ventral stream, has the job of computing object descriptions that can be used in cognition (recognition, reasoning and planning).[1]

This idea that the computation of object descriptions (representations of size, location, orientation, etc.) is performed twice, by two independent computational mechanisms located in two anatomically separate neural streams, I call *strong PAM*.[2] Strong PAM can be contrasted with a much weaker version of

---

[1] In what follows, I will mostly use the determinable 'cognition' in contrast with 'grasping.' But, when appropriate, I will use more determinate concepts, such as 'recognition', to refer to more specific cognitive capacities. When the context demands a focus on the response modalities, I will contrast 'grasping' with 'reporting' or 'judging'.

[2] A recent very clear formulation of strong PAM is offered by Ganel, Chajut and Algom (2008). Discussing their finding that perceptual adjustments and estimations of the length of a target object but not maximal grip aperture in grasping

PAM. According to *weak PAM*, object descriptions are computed once and sent subsequently to two anatomically separate streams.[3] Both weak and strong PAM agree on the anatomical split into two different streams, one dorsal and the other ventral (that is, both accept TVSH part 1). They disagree on whether the mechanism for computing object descriptions from visual 'raw data' is located in the brain before or after this anatomical split. Strong PAM locates the computational mechanisms after the split. It therefore postulates two independent mechanisms, one in the dorsal stream for grasping and another in the ventral stream for recognition and judgement. Weak PAM locates the mechanism computing object descriptions from visual 'raw data' before the anatomical split. It therefore postulates only one mechanism, which feeds the same object descriptions into both streams for grasping (dorsal) and cognition (ventral).

Even if the distinction between strong and weak PAM (two separate or one common mechanism for computing object descriptions) seems clear, it is in fact more difficult to make the distinction—both theoretically and empirically—than one might expect. This seems to be the case even for Milner and Goodale (even if the official version of TVSH is committed to strong PAM). In the above quote from 1992, Goodale and Milner formulate a strong version of PAM, according to which object descriptions are

---

the object were in accordance with Weber's law, the authors write: 'These findings document a fundamental difference in the way that object size is computed for action and for perception and suggest that the visual coding for action is based on absolute metrics even at a very basic level of processing.' See also Ganel, Chajut, Tanzer and Algom, 2008: 'This dissociation further supports the conclusion that action and perception are sustained by qualitatively different computations.'

[3] Weak PAM has been defended by Franz *et al*., 2000; Franz and Gegenfurtner, 2008 and Christiansen *et al*., 2014. In a review of their own and others' data from studies of illusion effects on grasping, Franz and Gegenfurtner (2008) p. 944, conclude: 'Taken together, all this criticism might lead us to the conclusion that the brain is not subdivided in two parallel systems, which perform certain calculations twice. Specifically, the notion that object size is calculated twice, once in the ventral stream for perception (deceived by visual illusions and conscious) and once in the dorsal stream for action (veridical and unconscious), loses much of its strength if one follows our arguments on visual illusions and grasping.'

computed twice and in parallel by two independent computational mechanisms. In other places, however, the authors seem to formulate a different version of PAM. In a more recent commentary discussing the objection that some illusions appear to influence both grasping and judging, they write:

> …the two streams share common inputs from early retinotopic cortical areas (orientation, location, size, etc.), so that processing of these visual features is not the absolute province of one or other stream (Goodale and Milner, 2010, pp. 65–6).[4]

Here, they seem to suggest that some object descriptions ('orientation, location, size, etc.') are computed before the anatomical split.

Two issues might contribute to this difficulty in differentiating between the strong and weak versions of PAM. First, the anatomical and the functional segregation claims are not always clearly kept apart. Both perception-action models accept the anatomical split between separate neural structures for action (dorsal) and cognition (ventral) but only the strong model claims that the anatomical split is mirrored functionally in two separate systems for extracting and encoding object descriptions. To bring out the difference between anatomical and functional segregation, consider a case where there is no pull to suggest that the anatomical segregation is mirrored functionally in independent visual systems. Consider the case of a perfectly ambidextrous person who can grasp objects with both hands with an equally high level of dexterity. In this case, there would be no reason to suggest that left-hand-grasping is driven by different object descriptions than right-hand-grasping. Nevertheless, left-hand-grasping and right-hand-grasping would involve relatively independent neural systems. We could lesion one system without interfering with the other. So, in this case, a computational mechanism extracting representations of features from visual 'raw' information would feed its object descriptions to both systems (left-hand-grasping and right-hand-grasping).

Second, an important distinction between 'core computation' of object descriptions and 'transformations' is sometimes overlooked. To grasp the import of this distinction, consider again the case

---

[4] Similar statements are found in Dyde and Milner, 2002 and Milner and Dyde, 2003.

of the ambidextrous person. Even if the person's left-hand-grasping and right-hand-grasping are driven by a common representation of features (orientation, location, size, etc.), the actual grasping actions involve separate and distinct sensorimotor transformations. Grasping with the left hand involves steering the hand in one direction, rotating the wrist in one way, etc., whereas grasping with the right hand involves a different set of transformations specific to the right hand. These transformations translate between the representation of the object features, the motor goal, and the current location and orientation of the hand. Both strong and weak PAM is committed to there being transformational processes of this kind. Whether or not the model claims that grasping and cognition are driven by independently computed object descriptions, it must accept some distinction between the 'core computations' of object descriptions and transformational processes translating between these descriptions, the motor goal, and the current state of the effector. We will see the importance of these distinctions (anatomical versus functional segregation and 'core computation' versus 'transformation') as we go along.

Strong PAM is often taken to be supported by a wealth of data. There is no room here to review and discuss the data in an exhaustive manner. Let me only say that most of the evidence can be grouped into three types: neuropsychological dissociation, neuroimaging, and behavioural psychophysics data. This paper will primarily be concerned with behavioural data. I will return to a brief discussion of the two other types of data in section 5. For now, our main question is whether behavioural data can be used to distinguish between weak and strong PAM. And, relatedly: does behavioural data support strong PAM more than it does weak PAM? My answer to both of these questions is going to be a conditional 'no'. I will argue that, given a common understanding of how to individuate computational mechanisms, the behavioural data underdetermine the choice of model. The next step of my argument is, therefore, to sketch the core ingredients of a common way to understand the individuation of computational mechanisms.

3. Individuation of Computational Mechanisms

The difference between strong and weak PAM is a difference between a model according to which object descriptions for grasping and cognition are computed by two independent mechanisms and a model for which they are computed by one common mechanism. At the heart of this dispute is, therefore, a counting problem. Is there one common mechanism, or two independent mechanisms, for computing object descriptions? This leads to the question of how we can count these computational mechanisms.

The counting question can be interpreted in both metaphysical and epistemic ways. As a metaphysical question, it is a question concerning what determines the individuation and the actual number of computational mechanisms. Individuation of computational mechanisms involves how one would answer questions such as 'What makes a physical process a computation?', 'What is the difference between a computation and any other process?' and 'Under what conditions are two mechanisms the same or different?' The last question, in particular, is important in the present context. The present concern is what determines whether grasping and cognition are driven by one common mechanism for computing object descriptions or two independent mechanisms (one for grasping and one for cognition). As an epistemic question, the counting question concerns how we can know the number of computational mechanisms. How can, for example, cognitive psychologists determine the number of computational mechanisms? These are not easy questions. In the present paper, I am assuming that the answer to the epistemic counting question depends on the answer to the metaphysical question.

Let us return to the metaphysical problem of the individuation of computational mechanisms. The received view, in both philosophy and cognitive science, is that the individuation must rely on the fact that, to quote Sprevak (2010, p. 260), 'computation must involve representational content'. The received view is that the question of whether we are dealing with one or two computational mechanisms can only be answered in terms of representational content. A computational mechanism is what implements an algorithm bringing us from some representational input to some representational output. It is not entirely clear what it means to say that a computational mechanism is individuated by its representational content. But, as a rule of thumb, we can take it to involve the idea that differences in type of representational input,

type of representational output, or both, imply different computational mechanisms. There is no room here to engage with the deeper reasons for the received view of individuation. Let me just mention that one prominent reason is the assumption that if we do not individuate computational mechanisms by representational content—say, if we instead individuate them in terms of causal mappings (Putnam, 1967) or causal counterfactuals (Chalmers, 1996)—then we are confronted with a devastating problem of pan-computationalism (Searle, 1990 and 1992). Even walls and rocks would turn out to be computational mechanisms. This would render the counting of computational mechanisms senseless.[5]

Given this received view of individuation of computational mechanisms, answering our target individuation question (are the object descriptions driving grasping and cognition computed by one common, or two separate, computational mechanisms?) involves two parts: first, the computational mechanisms are individuated by type of representation; and second, the type of representation is determined in some way. What determines the type of representation?

One prominent proposal that fits well with current views and practices in cognitive sciences (Ramsey, 2007) is the claim that the job-description of input-output relations of a functional mapping determines the representational content. The idea is well known from Cummins' work on representations (Cummins, 1989, 1991, 1996). A very simplified version is that input to some functional mapping is representational insofar as the function is one that requires content. Take the function of adding two numbers implemented by a pocket calculator. The physical inputs are button pressings and the physical outputs are shapes on the display. We can only understand this mapping between physical input and output as implementing the function of adding if we understand the physical input and output as symbols (numerals) representing numbers. Similarly, in the case of cognitive functions, the physical input might be

---

[5] For an overview of the discussions, see Piccinini, 2015a, 2015b. Piccinini does not accept the received view, which I will here accept for the sake of argument. Think of it as a conditional acceptance, i.e.: 'if the received view is true, then …'

one neural state and the physical output a different neural state, but only if we understand these states as representing, say, brightness and edges can we understand the function as edge detection.

The received view of individuation would then be that computational mechanisms are individuated by the type of representational content, and the type of the representational content is determined by the job description of functional mapping between input and output. One worry at this point might be that this view of determining representations by functional job descriptions leads to a threatening form of relativism. If the type of representation is determined by the job description of the functional mapping between input and output, and this job description is relative to our explanatory purposes, it would appear that our explanatory purposes determine the types of representations and, hence, also the number of computational mechanisms.[6] To take a simplistic toy example: if one job is to identify words longer than two syllables and another job is to identify words of two syllables or one syllable, are there different representations and, hence, mechanisms for these two jobs? Or, if grasping for eating and grasping for placing can be described as embodying two different job descriptions, do we then have different representational formats and mechanisms?

This worry assumes, without argument, that job descriptions are completely relative to our explanatory purposes. That is, it is being assumed that we can freely choose how to describe the job of a functional mapping. There are a number of reasons for not accepting this type of relativism. First, descriptions of functions are constrained by their explanatory power and success. Take the pocket calculator. We could choose to describe the functional mappings in any number of ways (say, as having the job of randomly mapping clicking sounds to funny shapes) but the description of the mapping as the function of adding is a much better explanation. Similarly, for descriptions of mappings in the visual system, our choice of descriptions should depend on their explanatory power and success. Power and success do not merely depend on our goals. They depend on the place of the description in a more general

---

[6] For an argument in this spirit, see Searle, 1990.

computational model of cognitive architecture and systematic evidence produced by careful experimentation.

Second, the relativism about job descriptions is in conflict with a certain view of scientific realism. If we think that explanatory success in science is generally due to its tracking the truth, we have some reason to infer the structure of reality from well-accepted and successful models. Models in cognitive neuroscience describe operations at certain levels of abstraction, allowing for some generalisations but not others. Describing a function in terms of computing descriptions of object features is at an appropriate level of abstraction, whereas descriptions in terms of identification of one syllable versus two syllables, or grasping for eating versus grasping for placing, are not. One view would be that it is the structure of reality that in the end determines the right level of abstraction (our generalisations are explanatory and successful when we carve nature at its joints). In other words, there are reasons for assuming that job descriptions do not come for free.

We can now reformulate strong PAM along the following lines: the job descriptions of the grasping-function, on the one hand, and the cognitive functions, on the other, are such that these functions require different types of representational inputs. Grasping objects requires a control of movements that is attuned to shapes, sizes and orientation that constantly change as a function of change in the spatial relation between the object and the person. By contrast, the recognition of objects requires being able to recognise the same objects despite a continuous change in perspective. One way to describe these different requirements is to say that grasping is a function that requires the representational input to be in a viewer-relative (egocentric) format, whereas recognition is a function that requires its input to be in an object-relative (allocentric) format. This way of describing the functions, as requiring different types of representational inputs (egocentric versus allocentric), has 'upstream' consequences. The mechanisms for producing these different types of representation—that is, the mechanisms that have these representations as their output—must be two distinct mechanisms (at least, if the received view of individuation is true). Consequently, the functional mapping from visual 'raw' information to object

description is implemented by two independent computational mechanisms: one for extracting and encoding representations of features in an egocentric format, and one for doing it in an allocentric format. In other words, we have two separate computational mechanisms: one for computing the egocentric representational input that drives grasping, and one for computing the allocentric representational input that drives cognition.

Clark (2001) examplifies this view:

fine-grained action control requires the extraction and use of radically different kinds of information (from the incoming visual signal) from that required for recognition, recall, and reasoning. The former requires a constantly updated, egocentrically specified, exquisitely distance- and orientation-sensitive encoding of the visual array. The latter requires a degree of object-constancy, and the recognition of items by category and significance irrespective of the fine detail of location, viewpoint, and retinal image size. A computationally efficient coding for either task precludes the use of the very same encoding for the other (Clark, 2001, p. 501).[7]

_____

[7] Similar claims are found in Goodale and Haffenden,1998. They write: 'This division of labour in visual processing requires that different transformations be carried out on incoming visual information by the two streams. Consider first the task of the perceptual mechanisms in the ventral stream. To generate long-term representations of objects and their relationships, perceptual mechanisms must be "object-based" so that constancy of size, shape, color, lightness, and relative location can be maintained across different viewing conditions. […] However, whatever the particular coding might be, it is the identity of the object, not its disposition with respect to the observer that is of primary concern to the perceptual system. This is not the case for the visuomotor mechanisms in the dorsal stream, and other related structures, that support actions directed at that object. Here the underlying visuomotor transformations must be viewer-centered; in other words, both the location of the object and its disposition and motion must be encoded relative to the observer in egocentric coordinates, that is in retinocentric, head-centered, torso-centered, or shoulder-centered coordinates. […]To summarize then, while similar (but not identical) visual information about object shape, size, local orientation, and location is available to both systems, the transformational algorithms that are applied to these inputs are uniquely tailored to the function of each system. According to the

The basic idea can be expressed by the following two conditionals: first, if two cognitive functions are sufficiently different (*viz*. have sufficiently different job descriptions), then the efficient solution to each function requires a different representational format. Second, if efficient solutions for two functions require different representational formats, efficient solutions require two separate mechanisms computing the representations. We already knew that grasping and cognition involved separate and independent computational *transformational* mechanisms. After all, one terminates in the physical act of grasping and the other in an overt judgement. The new thing is that from the specification of representational formats, given the received view of individuation, something can now be inferred about the 'core computational' mechanisms. If computational mechanisms are individuated by their representational output, there are two distinct mechanisms for computing object descriptions: one for computing them in an egocentric format and one for computing them in an allocentric format.

The starting point of this chain of inferences is the claim that grasping and cognition are so radically different in terms of function that they require different types of representational inputs (egocentric versus allocentric). Justification of strong PAM seems to rest on this claim. This metaphysical account of individuation has consequences for the epistemic counting problem: how can we know the number of representational formats? No empirical science could rest content with the mere postulation of radically different functions. How can it be empirically established that the transformational processes involved in the control of grasping and cognition are so radically different that they must be driven by different types of input representations? One prominent method for answering this question would be to employ the standard procedure of cognitive psychology of looking for significant differences and similarities in response profiles across behavioural tasks. For example, similarities in the response profiles (here, reaction time) between the task of physically rotating a 3D-object and the task of making same-different

Milner and Goodale [2006] proposal, it is the nature of the functional requirements of perception and action that lies at the root of the division of labour in the ventral and dorsal visual projection systems of the primate cerebral cortex' (Goodale and Haffenden, 1998, pp.162–3). See also Foley *et al*., 2015.

judgements with respect to 2D-representations of 3D-objects have been taken to indicate that behaviour in the two tasks draw on similar cognitive operations and representations (Shepard and Metzler, 1971).

Correspondingly, in the case of grasping and cognition, differences in response profiles have been used to argue that behaviour in the two tasks is driven by different representations. This inference is only justified if the two tasks being compared are sufficiently alike. If they involve different levels of complexity and difficulty, and demand different levels of attentional resources, differences in the response profiles might not be caused by different types of representations. Behavioural experiments attempting to use dissociation between grasping and cognition to argue for independent computational mechanisms must align the two tasks sufficiently to eliminate these task confounds. In certain respects, grasping and judging have to be almost the same task.[8]

The inference from different response profiles of grasping and cognition to different cognitive representations thus seems to rely on the following conditional (call it the 'Task Conditional'):

If:

(1) a subject's 'cognitive' performance differs significantly from her 'grasping' performance and

(2) the experimental tasks are sufficiently similar in both 'cognitive' and 'grasping' conditions,

then:

(3) 'cognitive' and 'grasping' performance are most likely driven by separate types of representation.

If the Task Conditional can be used to establish that behaviour in the different conditions is driven by different representations, we can infer the presence of different computational mechanisms by using the individuation requirements. This experimental logic motivates a particular way of analysing the behavioural data. To put it crudely, data from trials in one response modality are lumped together and compared with data from trials in the other response modality. To illustrate with the Shepard and Metzler (1971) study, mean response times were calculated for trials with 2-D representations at an angle of

---

[8] For a discussion of this point, see Franz *et al.,* 2000.

rotation of 0°, 20°, 40°, etc. A linear function was then fitted to the mean reaction times and compared to what is assumed to be the linear function relating the angle of rotation and time to rotate in an actual rotation task. Similarly, in the grasping tasks, one common way to analyse the data is to compare mean maximum grip aperture in grasping trials to mean perceptual adjustment or estimation in 'judgement' trials. In this way, the general shape of the response profiles can be compared to each other. Admittedly, this is a very crude description of the often sophisticated statistical analysis involved. The aim here is simply to bring out the important difference between the usual experimental logic and the alternative type of paradigm sketched in section 6.

Both proponents and opponents of strong PAM usually accept the Task Conditional as unproblematic. Most arguments in the scientific literature against strong PAM have tried to deny that the behavioural data support the antecedent (that is, they deny the conjunction of (1) and (2)). These opponents to strong PAM have tried to establish that when the tasks are sufficiently similar, there is no significant difference in performance (see, for example, Bruno *et al*., 2010; Franz and Gegenfurtner, 2008; Franz *et al*., 2000; Schenk *et al.,* 2011). Thus, some of the most important scientific criticism of strong PAM has come from researchers who have used the conditional to establish that when grasping and judgement are sufficiently similar, grasping and judgement are most likely driven by a common representation.[9] From this, it is inferred that the object descriptions are computed by a common mechanism before the anatomical split.

In the next section, I will question this conditional and its use in determining representations and computational mechanisms. Before moving on, though, let us sum up the argument so far. Given the received view of individuation, without a way to infer from behavioural data the representational formats employed by the cognitive functions, it will be impossible to determine empirically whether grasping and cognition are driven by object descriptions computed by one common, or two separate, mechanisms. The

---

[9] That this is, in fact, the most important scientific criticism is supported by recent survey chapters and review articles written by proponents of strong PAM—for example, Goodale, 2014 and Westwood and Goodale, 2011.

rest of this paper is concerned with the following two questions: 1. Given the received view of individuation, how can we distinguish empirically between strong and weak PAM? 2. Are there ways in which we can empirically distinguish between the models that do not depend on the received view of individuation?

4. The Under-Determination Problem

Given the received view of individuation of computational mechanisms, strong PAM requires two inferential steps. First, the proponent of strong PAM has to infer the number of representational formats from systematic performance differences in experimental tasks. Second, the proponent has to infer the number of computational mechanisms from the number of representational formats. Often, it is assumed that grasping is a function of information that is encoded in an egocentric representational format, whereas judging is a function of information that is encoded in an allocentric representational format. We would consequently have separate mechanisms for computing egocentric and allocentric representations of length, size, orientation, etc.

Inferring representational formats from systematic performance differences in experimental tasks is no trivial achievement. Indeed, it is a recognised fact that this form of inference is confronted with a vicious under-determination problem. A cognitive function consists of some type of representation and some type of operation performed on the representation, resulting in the observed task performance. When inferring the type of representation from task performance, we are thus confronted with a problem with two unknowns: the representation and the operation. The structure of the under-determination problem is simple. Say that we know that the sum of two unknown numbers $(X + Y)$ is 5. There is no way in which we can infer the value of X just from our knowledge of the sum. Any change in X could always be compensated by a change in Y to yield the sum of 5. Analogously, in the case of inferring representational formats from systematic performance differences, we can never know whether observed performance differences are the result of changes to the representational formats or to the operations. To be sure, the

analogy is not perfect. Psychological processing is constrained in ways the simple equation is not. Whereas all numbers can be slotted into the equation, not all types of representation can be slotted into human psychology.

This point was made nearly 40 years ago by Anderson (1978). He argued that there was no way in which the debate concerning the representational format ('pictorial' versus 'propositional') of mental imagery could be settled by behavioural evidence. The debate concerned which of the models could best explain and predict data from experimental tasks assumed to require the use of mental imagery. Anderson showed that behavioural data (response time and accuracy) could be equally well explained and predicted by both models because any supposed change in representational format could always be compensated for by a change in the cognitive operations.[10]

Anderson intended his argument to apply not just to the mental imagery debate. It was a general challenge to cognitive psychology. In any case, the logic also has bite in the present context. As we saw in the preceding section, given the received view of individuation of computational mechanisms, the issue of whether there is one common, or two independent, mechanisms for computing object descriptions hinges on the determination of whether grasping and cognition are driven by one common representation or by two independent representations. For this inference to hold, it is essential that we can infer the representational format from the task performance. This is exactly what Anderson's argument is challenging. We cannot know, based on differences in behavioural performance of grasping and judging in illusion tasks, whether the differences should be explained by grasping using an egocentric format and judging using an allocentric format, or by grasping and judging using the same format but employing different transformational operations.

---

[10] In a recent review of the imagery literature, Ganis and Schendan (2011) write in a brief discussion of Anderson's 1978 paper: 'Eventually a consensus emerged that competing theories could not be sufficiently constrained by behavioral findings alone because they could be made to mimic each other by means of creative trade-offs between representation and processing assumptions.'(239-240)

But what about the different job descriptions for grasping and cognition that Clark and others have described? According to these descriptions, grasping needs to be tuned to the actual features (e.g. the size) of the object, whereas cognition needs to abstract away from the ever changing details in order to subsume the impressions under their general category or concept (it is a cup, no matter how it is oriented in space). This way of phrasing the difference between grasping and cognition is wildly misleading. Remember that in order to use the Task Conditional to infer anything about the similarity or differences of the underlying cognitive workings, the grasping task and cognitive task need to be sufficiently similar. Pitting grasping against object recognition (say, an object naming task) does not satisfy this condition. If we are interested in studying and comparing systems for processing of object features, then we have to pit grasping against some way of judging the object's size (or some other feature, not its category).

It is not obvious why judging an object's size should not be tuned to the object's actual size, just like grasping has to be. If that is the case, then it is not obvious why cognition could not be driven by the same type of egocentric representations as grasping. Without going deeper into the problem of how egocentric and allocentric representations are related to each other in motor control and spatial navigation, I will assume that it is possible that the representations driving the type of judgements involved in the tasks in question could have an egocentric format.[11]

If this possibility is accepted, then any significant difference between grasping and judging can be explained as a difference in the transformational operations rather than by a difference in representational formats. If there were systematic performance differences, they could occur because the ventral stream under certain conditions is not an optimal machine for processing egocentrically encoded representations. Maybe, under certain conditions, the ventral machine is less reliable, more resource-demanding and slower

---

[11] This possibility of transforming egocentric information into allocentric representations also seems to motivate observations made by Klatzky (1998). She points to results showing that 'path-integration' employing mainly allocentric representations of the animal's position can be computed from 'efference copy, vestibular signals, and optic flow' (p.8)—that is, from signals that are often taken to be egocentric.

than the dorsal machine. Conversely, with respect to dorsal stream processing, it could be argued that grasping employs the same visual representation as judging but that, in addition to visual representation of object size, grasping also uses other types of information (e.g. visual information about the hand and haptic information from touch) not available in judgement tasks (Franz, Hesse and Kollath, 2009; Schenk, 2012b). I will respond to a possible objection to this under-determination argument shortly. Before doing this, it is instructive to see how this dialectic has played itself out in the empirical literature.

The fact that behavioural data can either be explained by differences in representational formats or transformational operations is important in recent debates about Weber's law in relation to maximum grip aperture in grasping tasks. According to Weber's law, sensitivity to changes in the physical stimulus decreases as the stimulus intensity increases. In other words, the 'just noticeable difference' (JND) of change in the stimulus increases for larger stimulus magnitudes. Using standard deviation as a measure of JND (variance is a measure of the area where the participant is unable to tell the difference between the target stimulus and the comparison), Ganel, Chajut and Algom (2008) showed that variance in the maximum grip aperture was relatively constant over all object sizes tested in their experiment. When grasping objects of various sizes, the maximum grip aperture violates Weber's law. On this basis, Ganel, Chajut and Algom (2008) inferred that vision-for-grasping uses different computational principles than vision-for-cognition. In response, Utz *et al.* (2015) argue that Weber's law might be masked in the case of grasping by non-visual types of noise (e.g. haptic information or motor related information). On this basis, they speculate that if we could remove non-visual forms of noise from grasping, then maximum grip aperture might conform to Weber's law. Furthermore, Utz *et al.* observe that for larger object sizes there is in fact a decreasing variance. That is, in tasks with increasing object sizes they actually observed an inversion of Weber's law. The authors explain this effect by the fact that grasping is subject to biomechanical constraints. Less variance in grip aperture for larger object sizes is explained by the fact that the range of motion is constrained by the hand's ability to open up maximally. In other words, the findings reported by Ganel and colleagues are explained not by two distinct types of representation but by one

common representation plus non-visual forms of information and extra-sensory biomechanical factors involved in grasping.

We thus have two competing models: on the one hand, a strong model—according to which, grasping is driven by egocentric representations plus some operation Y_1 and cognition is driven by allocentric representations plus some operation Y_2. On the other hand, a weak model—according to which, grasping is driven by egocentric representations plus some operation Y_1 and cognition is driven by egocentric representations plus some operation Y_2. Behavioural data cannot distinguish between the two models. The second model can deliver the same behavioural output as the first model if only the invariance in the representational formats is compensated for by changes in the transformational operations and various kinds of transformational noise.

This has consequences for the counting of computational mechanisms, at least in so far as this counting relies on the received view of individuation. If behavioural data is unable to distinguish between the 'separate-representational-formats model' and the 'common-representational-format model', then behavioural data is unable to distinguish between strong and weak PAM. That is, given the received view of individuation of computational mechanisms by representational formats, if behavioural data cannot distinguish between the representational formats, then behavioural data cannot be used to count the number of mechanisms computing the representations in question.

It might be objected that even if I am correct about the determination of representational formats, we can still functionally distinguish between dorsal and ventral processing streams in terms of their radically different transformational operations. So, even if it is conceded that the apparent difference between representations may just be a difference between transformations, some people may respond by thinking that there are trivial and non-trivial differences between transformations, and the ventral/dorsal difference is non-trivial. Trivial differences between transformations exist between grasping an object with a two-finger precision grasp and whole-hand power grasp or grasping the object at one location and grasping it when located a little to the right. The differences between transformations in dorsal and ventral

stream are not trivial in this way, according to this objection. So, when I submit that different behavioural performance for perceptual versus action tasks might emerge because somehow the ventral system is less well equipped to process egocentric information, it might be objected that 'many scientists in the field would accept such evidence as valid evidence that vision for perception and action is in fact based on distinct computational mechanisms' (quoting an anonymous reviewer).

The methodological problem here concerns what can be inferred from systematic performance differences in separate response modalities, assuming that the tasks are sufficiently similar to rule out task differences (creating different attentional demands, allowing for different optimal processing strategies, etc.) as the explanation of performance differences. The objection concedes that we might not be able to distinguish between types of representational format on the basis of behavioural data but insists that significant differences in transformational processes are sufficient to distinguish between strong and weak PAM. A computational mechanism consists of a representation X and an operator Y. If we can know Y, this might suffice as valid evidence for strong PAM. This kind of reasoning is confronted with a number of serious problems.

First, according to this objection, we should individuate computational mechanisms not by their representational content but by their operations. On the one hand, if this is to be understood as a denial of the received view of individuation of computational mechanisms, then it is confronted with the problem of pan-computationalism. If it is not in virtue of operating on representational content but merely in virtue of its causal transactions, then everything instantiating the causal patterns in question would be a visual system. According to some philosophers, this would imply that everything is a visual system (Putnam, 1988; Searle, 1990). On the other hand, if this is to be understood as an acceptance that we need to look for other ways of counting computational mechanisms besides types of involved representational content, then I agree. I sketch such a suggestion in section 6.

Second, the debate between proponents of weak and strong PAM in the scientific literature is in terms of common or separate representations (and not in terms of different kinds of operations).[12] The scientific literature is thus preoccupied with the question of representations as a way to answer the question of the number of involved mechanisms.

Third, the debate about the applicability of Weber's law to grasping illustrates how hard it is to determine which operations or transformations count as the right ones. That is, which operations count as 'core computations' and which as extra-visual transformations? There are two aspects to this problem: first, the reason for thinking that some differences between transformations are 'trivial' is that they involve transformational operations on the same types of representational inputs. Left-hand-grasping versus right-hand-grasping and precision-grasping versus power-grasping are cases of 'trivial' differences in transformational operations because we are assuming that they operate on the same type of descriptions of object features. Remove the underlying assumption of types of representation, and we no longer have a distinction between trivial and non-trivial transformation differences. Second, we need some way to say that an operation is a visual operation and not a haptic one, say. If the proponent of strong PAM insists on not doing this, it is open to the supporter of weak PAM to say that the different response profile in grasping is due to other types of processing.[13] One standard way of drawing distinctions between types of processing is exactly in terms of representational content. In sum, in isolation from assumptions about

---

[12] See, for example, the following quote from Franz and Gegenfurtner (2008): 'If […] some methodological precautions are used, then these diverse perceptual results [i.e. results from psychophysical experiments on grasping under illusion] can be reconciled and point to a *single internal size estimate* that is used for perception and for grasping. This suggests that the Ebbinghaus illusion deceives *a common representation of object size* that is used by perception and action' (Franz and Gegenfurtner, 2008, p.920, my emphasis). See also Goodale and Haffenden, 1998 and Foley *et al.*, 2015.

[13] This is exactly what is argued by Utz *et al.*, 2015. They argue that Weber's law is masked in the grasping tasks by other kinds of sensory processing. A similar argument is found in Schenk, 2012b.

representational formats, there is no acceptable inference from different operations involved in grasping and cognition to independent and separate mechanisms for computing object descriptions.

If the line of reasoning presented in this paper is correct, the conclusion appears to be that available behavioural data are unable to distinguish between weak and strong PAM. Available data do not support the model that grasping and cognition are driven by separate forms of representations computed by two independent computational mechanisms any more than they support the model that grasping and cognition are driven by a common representation computed by one common computational mechanism. That is, without further substantial constraints, behavioural data do not allow us to determine whether systematic behavioural differences are driven by distinct types of representation (as claimed by Ganel *et al.,* 2008) or a common representation plus distinct transformational operations (as claimed by Utz *et al*., 2015).

5. Under-Determination and Converging Data

I have been focusing exclusively on behavioural data. Maybe the under-determination problem is only a problem concerning behavioural data? We have already seen that some additions of further processing assumptions (say, assumptions concerning Weber's law) to the perception-action models do not sufficiently constrain the choice of PAM. Maybe we can overcome the problem if we draw on other types of data? In particular, neuroimaging data and neuropsychological data have been absent from my discussion so far.

To return to the imagery debate—it is often assumed that although the type of representational format is under-determined by behavioural data alone, this might not be the case if we draw on data from neuroimaging experiments. In particular, neuroimaging experiments indicating that brain areas devoted to visual processing are active during imagery tasks have been thought to support the pictorial-representation model of mental imagery (Ganis, Thompson and Kosslyn, 2004; Ganis and Schendan, 2011). Similarly, in the case of grasping and cognition, perhaps the under-determination of the representational formats could be overcome by drawing on brain data.

Without producing anything like a general argument, in the case of PAM, this way of reducing the under-determination and constraining the model choice strikes me as unsuccessful. Let us first consider data from neuropsychological dissociations, since these have played an immensely important historical role in the generation of strong PAM (Goodale and Milner, 1992; Goodale *et al.*, 1991; Goodale *et al.*, 1994; Karnath *et al.*, 2009; Milner *et al.*, 1991; Milner and Goodale, 2006). The majority of the data stems from a few patients who, together, are thought to demonstrate a double dissociation between ventral stream and dorsal stream visual processing. On the one hand, we have patients who have suffered damage to the ventral stream (e.g. resulting in visual form agnosia), with the consequence that they can reach and grasp objects but cannot recognise them. On the other hand, we have patients with damage to the dorsal stream (e.g. resulting in optic ataxia), with the consequence that they can recognise objects and shapes but cannot reach and grasp them.

It is important to note that the existence of such doubly dissociative patterns does not establish the existence of two independent mechanisms computing object descriptions that feed into vision-for-action and vision-for-cognition respectively. There is no special deductive inference taking us from double dissociations to specific claims about the cognitive architecture (Davies, 2010). Double dissociation data might be strong evidence for the anatomical segregation into separate neural pathways but is not in itself evidence for the corresponding existence of two computational mechanisms for the extraction and encoding of visual features.[14] Consider a pair of perfectly ambidextrous persons, both grasping a rod with the left hand and grasping it with the right hand. We can imagine brain damage to one person, which selectively impairs her left-hand-grasping while sparing her right-hand-grasping, and brain damage to the second person with the reverse pattern of impairment and sparing. This would allow us to say that left-hand-grasping and right-hand-grasping are implemented in partially separate anatomical structures.

---

[14] Contrary to what is often claimed by proponents of strong PAM, e.g. Goodale *et al.,* 1994: 'This double dissociation lends strong support to the idea that the visual mechanisms mediating the perception of objects are functionally and neurally distinct from those mediating the control of skilled actions directed at those objects' (p. 604).

However, it does not allow us to infer that left hand grasping and right hand grasping are driven by two independent mechanisms for computing object descriptions. Behavioural dissociations could be explained by damage to independent 'transformational processing systems'.

Similarly, in the case of vision-for-grasping and vision-for-cognition, the neuropsychological dissociation data allow us to say that these two processing tasks (grasping and manual size estimation, say) are implemented in partially separate anatomical streams. However, they do not allow us to infer that vision-for-grasping and vision-for-cognition are driven by two independent mechanisms for computing object descriptions. The dissociation would only seem to support the claim that the competencies underlying performances in the two tasks (e.g. grasping versus manual size estimation) involve *some* modularity. This is still compatible with a substantial overlap between the competencies and their underlying neural systems. We could not from the double dissociations tell whether the best explanation is in terms of selective damage to two independent 'core mechanisms' for computing object descriptions or in terms of damage to independent 'transformational processing systems'.

This under-determination problem should not surprise us. Double dissociations do not embody a special kind of deductive logic (Davies, 2010). As empirical evidence playing a part in supporting an empirical hypothesis, data concerning double dissociations are part of a comprehensive net of data that can serve in various inferences to the best explanation. Data lend support to one model only in so far as they weaken the support for competing models (Sober, 1999, 2008). In the context of a discussion of strong against weak PAM, dissociative data would only strongly support the existence of two independent mechanisms for computing object descriptions (strong PAM) if they somehow pointed away from or conflicted with a weak model. But as we have seen, that is not the case. The dissociations could be explained by damage to 'transformational processing systems'. Summing up, neuropsychological data

concerning double dissociations do not by themselves establish strong PAM. They are equally probable given weak PAM.[15]

It should be noted that I am not implying that lesion data can never constrain computational modelling. All I am claiming is that in the choice between strong and weak PAM, neuropsychological data do not seem decisive for the simple reason that both versions of PAM accept the same underlying anatomical 'two-pathways' model. To be sure, strong and weak PAM do not embody the exact same anatomical commitments—after all, one model postulates two neural mechanisms and the other only one. So, at a fine-grained level, they make different anatomical predictions. At a more coarse-grained level, the level of human neuropsychology, they make the same predictions.

This brings us to data from neuroimaging experiments. Perhaps this kind of data can sufficiently constrain and justify our choice of model.[16] A lot has been written recently on the epistemic relationship

---

[15] A further problem confronting the inference from neuropsychological dissociation to strong PAM concerns the fact that grasping might be guided by many types of information—for example, haptic information, visual information, etc. Impairment of one informational source can be compensated for by use of another informational source. Hence, it can be difficult to rule out the possibility that spared grasping performance is due not to spared visual processing but to haptic processing, say. This was recently argued by Schenk (2012a). Schenk demonstrated that DF's performance in grasping tasks was abnormal (similar to her performance in judgement tasks, such as manual size estimation) when haptic feedback was removed. That is, in the original Milner and Goodale studies with DF (Milner and Goodale, 2006), judging tasks and grasping tasks were not sufficiently similar. Schenk (2012a) demonstrates that if they are sufficiently similar, systematic differences disappear. See, however, Whitwell *et al.*, 2014 and Hesse and Schenk, 2014, for important qualifications.

[16] This type of data is thought by some researchers to deliver the strongest form of evidence for strong PAM. This is suggested by Goodale in a recent review of the fMRI literature: '…perhaps some of the most convincing evidence for the perception-action proposal has come from functional magnetic resonance imaging (fMRI) studies of the dorsal and ventral streams in the human brain'(Goodale, 2014, p. 280).

between cognitive models and neuroimaging data.[17] One thing that seems to result from these debates is that imaging data seriously under-determine the choice of cognitive model. In particular, there are reasons for thinking that if the existence of neuropsychological double dissociations does not support a two-mechanisms model over a one-mechanisms model, then neither will neuroimaging data. First, if the arguments of Henson (2005) are on the right track, the inference pattern of most fMRI paradigms is similar to the inference pattern lying behind double dissociations. Just as double dissociations are used in support of dual-route models (e.g. dual-route models of long term memory, where the working memory route can be damaged while sparing the implicit route, and vice versa—see Coltheart, 2001), neuroimaging data indicating differential patterns of activation are used to support similar architectural claims (such as a dual process model of recognition memory and familiarity). Second, if the arguments of Machery (2012) are correct, then fMRI data has a much poorer epistemic standing than double dissociations. The reason is that the number of possible confounding factors is always higher in neuroimaging dissociations than in double dissociations in neuropsychology. There are, therefore, reasons for thinking that if double dissociations do not support strong over weak PAM, then neither will fMRI data.

Admittedly, claims in this area of cognitive science should be made with extreme caution. Imaging techniques are rapidly evolving and new paradigms are constantly being developed that allow for the simultaneous use of multiple measurement techniques. Maybe new techniques and paradigms can be developed to obtain data at a sufficiently fine level of anatomical grain to constrain and justify the choice between strong and weak PAM. The challenge is that at present, it is difficult to see that strong and weak PAM embody seriously different anatomical commitments.[18]

---

[17] See Coltheart, 2006a, 2006b, 2010; Henson, 2005, 2006; Machery, 2012, 2013; Mole and Klein, 2010, and Poldrack, 2006, 2010.

[18] One promising avenue of research might be the use of traditional behavioural measures coupled with the simultaneous use of TMS and EEG. One such study is by Wokke *et al*. (2014). Using figure-ground discrimination tasks, the authors provided strong evidence for collaborative interaction (inhibitory and facilitatory functions) between

Summing up this section, neuropsychological dissociations and data from neuroimaging studies do not provide us with data that can be used to overcome the under-determination of weak and strong PAM.

6. Sceptical Ruin of Cognitive Psychology?

Anderson (1978) argued that the under-determination of cognitive models by behavioural data is a general problem for cognitive psychology. You might think that if he were right, it would mean the sceptical ruin of cognitive psychology. After all, cognitive psychology's primary *modus operandi* is to infer processing models from behavioural experiments. I will restrict myself to the choice between strong and weak PAM. An implication of this general sceptical conclusion would be that it would be impossible to make an empirically informed rational choice between the two perception-action models.

We can avoid these sceptical consequences. Recall that the under-determination problem for the two-mechanism or one-mechanism model is premised on the assumption of the received view of individuation of computational mechanisms. The inferences go from behavioural experiments to types of representational formats to number of computational mechanisms. If there were other ways of counting computational mechanisms besides counting types of representational formats, there might be ways to determine which of the models is most likely to be correct. One option is to investigate whether the computational mechanisms qua information processors have statistical-informational properties that could allow us to distinguish between strong and weak PAM.

---

ventral and dorsal processing streams. Such interaction might be more easily accommodated by weak PAM for the following reason. According to weak PAM, visual processing in dorsal and ventral streams is driven by a common representation. This makes cross-talk between the streams computationally plausible and relatively easy to understand. If, by contrast, the streams were speaking different computational tongues (strong PAM), cross-talk would plausibly involve some computational translation—if the interaction between the streams were to be more than brute causal influence. See, however, Goodale and Milner, 2010.

To see what I have in mind, consider the following analogy. Imagine an army intelligence agency. At one end, information from the field enters into the agency and, at the other end, actions are executed—say, a drone is sent to execute a supposed terrorist at a set of coordinates and a report describing the whereabouts of the terrorist is sent to the Secretary of Defence. On a given trial, we can observe a batch of data entering the agency and the results of its actions but we do not know which of the following two models of the internal workings of the agency is the correct one: according to the first model (strong Army-Intelligence-Agency Model, AIAM for short), raw information is analysed by two separate and independent analysts, each of whom sends her analyses to her own decision-maker, who instructs her own executor (one executor sends a drone and the other a report). According to the second model (weak AIAM), raw information is analysed by only one analyst, who sends her analysis to two separate and independent decision-makers, each of whom instructs her own executor (one sends a drone, the other a report). So, according to strong AIAM, information is analysed twice and in parallel by two independent analysts, whereas, according to weak AIAM, information is analysed once and sent to two independent decision makers.

How can we know whether weak or strong AIAM is the true description of the agency? One suggestion would be the following. As long as the raw information entering the agency and analysed by the analyst is clear and unambiguous, we cannot distinguish between the two models. When the information has this high-fi quality, two independent analysts would (on a trial-by-trial basis) tend to arrive at the same analysis (supposing they follow the same agency protocols). The effects of two analysts would tend to be indistinguishable from the effects of one. Strong AIAM and weak AIAM would predict the same responses. If, by contrast, the information were degraded, ambiguous, or of a very low quality, it would make a difference whether information were analysed by one or two analysts. If there were two analysts, they would (on a trial-by-trial basis) have a high risk of arriving at different analyses because they would have a tendency to interpret very degraded and ambiguous information in different ways. The divergent analyses thereby generated would be sent to the decision-makers, who would then send their equally diverging

instructions to their executors. As a result, we would (on a trial-by-trial basis) tend to observe a misalignment of target coordinates for the drone and the description of the whereabouts of the terrorist. By contrast (still with degraded information), were there only one analyst computing only one analysis, there would, of course, be no difference between the results sent to the decision-makers (since they would be the results of one and the same analysis), and we would tend to observe a (trial-by-trial) alignment of responses—i.e. on a given trial, the target coordinates for the drone would match the verbal descriptions of the terrorist's whereabouts. In other words, when in-coming information has poor low-fi quality, on a trial-by-trial basis, strong AIAM would predict a tendency toward misaligned responses, whereas weak AIAM would predict a tendency toward aligned responses. Notice that nothing here depends on the determination of the representational format of the analyses. Counting the number of mechanisms for computing the analysis (the number of analysts) depends only on measuring the probability of alignment of the two executors on a given trial.

A similar logic can be applied to the case of grasping and judging. In fact, Christiansen *et al.* (2014) applied just this type of logic in a recent study where participants were instructed (in the same trial) to both grasp and judge (length and orientation) rectangles on a computer monitor. As in the case of the army intelligence agency, the goal of the study was to determine whether object descriptions used to drive the two responses (grasping and judgement) are computed by one common or two independent mechanisms. As in the case of the agency, the authors argued that we should not expect to observe (on any given trial) any significant difference between the two responses when the visual stimulus has high-fi quality (high contrast, long presentation time, etc.). However, when presented with a degraded stimulus, it should become possible to distinguish between strong and weak models. Interestingly, the results demonstrated that, on a trial-by-trial basis, grasping and judging responses to length and orientation of a rectangle are in agreement significantly above chance level. On trials where participants successfully grasp the object, they also make correct judgements, whereas on trials where they do not succeed in grasping the object, their judgements are wrong. Furthermore, the authors showed that the accuracy of grasping and judging

responses, on a trial-by-trial basis, was very similar for both length and orientation. According to the authors, these findings are best explained by assuming that grasping and judging responses are driven by feature information processed by a common mechanism. If the authors are correct, counting mechanisms for computing object descriptions need not depend on the determination of representational formats. Counting the number of mechanisms for computing the representations of features can depend only on measuring the probability of alignment on a trial-by-trial basis of the grasping and judging performance.[19]

One might wonder exactly how this is supposed to be different from the use of response biases to infer the number of representational formats and computational mechanisms. After all, is not the supposed alternative I have sketched also inferring the number of computational mechanisms from differences and similarity in response biases? In answering this question, it is important to notice that the analysis associated with this alternative paradigm is looking at trial-by-trial variance. That is, the authors were interested in analysing the probability that on a given trial grasping and judging were the same (both correct or both incorrect) or different (one correct and the other incorrect). The results excluded the possibility that sameness of response on a given trial was due to chance and to the response in one modality (grasping) driving the response in the other modality (judging) (Christiansen *et al*., 2014). This trial-by-trial focus is different from the standard way of comparing response profiles for grasping tasks and judgement tasks.

This example goes some way towards demonstrating that Anderson's under-determination problem has no catastrophic consequences for cognitive psychology. General scepticism about scientific psychology ensues only if we assume the received view of individuation of computational mechanisms and adopt the research strategy of determining representational formats. If we accept that there might be other ways to count computational mechanisms, debates like the one concerning strong and weak PAM could still be empirically settled.

---

[19] A similar logic and statistical analysis is employed by Gegenfurtner and Franz, 2007.

7. Concluding Remarks

In this paper, I have argued that we should be sceptical of uncritical reliance on the strong version of Milner and Goodale's perception-action model (PAM). Most of the data produced and used in the discussions of PAM do not support any strong conclusions about the cognitive architecture of the mind. If that is the case, we should also be suspicious of any attempt to use strong PAM in support of philosophical claims about the nature of mind. Let me round off by first summing up the main argument and then relating the results to the debate about consciousness in philosophy of mind (concerned with the third part of TVSH).

My target in this paper has been an on-going debate in the cognitive neurosciences: is strong or weak PAM best supported by available data? According to the strong perception-action model, object descriptions used in guidance of grasping and object descriptions used in various cognitive performances are extracted from visual information and encoded by two independent computational mechanisms. According to the weak model, object descriptions used in grasping and cognition are computed by one common mechanism. So, according to the first model, size is computed in parallel by two independent mechanisms, whereas according to the latter model, size is computed only once by one common mechanism.

How do we determine whether there are one or two computational mechanisms? I argued that if we accept the assumptions of a standard view of individuation of computational mechanisms, these mechanisms are individuated by their representational content. The number of computational mechanisms is determined by the types of representations serving as input or output to these mechanisms. On the assumption that types of representations are determined by the representational format, the important task for psychologists and cognitive neuroscientists is to determine the formats of representation that are supposed to drive behaviour in various types of behavioural tasks. By looking at systematic performance differences and similarities across different task situations, it is assumed that we can infer whether grasping and cognition are guided by representations with the same or with different representational formats.

Given the received view of individuation, only if we can determine the number of representational formats guiding performance can we infer the number of computational mechanisms.

I argue that this chain of inferences is obstructed by a vicious under-determination problem. Systematic performance differences do not support a two-representational-formats model more than they support a common-representational-format model. The number of representational formats is under-determined by behavioural data, and the use of data from neuropsychology and neuroimaging studies does not seem to fix this problem – hence the data under-determine the choice between strong and weak PAM.

This reliance on the received view of individuation of computational mechanisms is crucial. The under-determination problem is a problem for determining representational formats and only indirectly a problem for counting mechanisms. Setting aside the issue of representations, maybe computational mechanisms have other properties that can be measured? Maybe some of these properties can be used to count the mechanisms? These are open questions. I sketched one possible avenue for further research with respect to PAM, building on the idea that the computational mechanisms in question are information processors: they have the job of extracting and encoding information about object features from the visual data. If two independent mechanisms are doing the extraction and encoding, under informationally noisy conditions they should tend to arrive at different results (i.e. under certain conditions, arriving at the same result should be a matter of mere chance). By contrast, if only one mechanism is doing the extraction and encoding, it should tend to arrive at the same result on a trial-by-trial basis. Under these conditions, we might have measurable differences between the two models.[20]

Now, how does this under-determination problem relate to the philosophical discussion of the implications of the Two Visual Systems Hypothesis (TVSH) for theories of phenomenal consciousness? As I

---

[20] This alternative counting proposal is consistent with the received view of individuation of computational mechanisms. I am not proposing an alternative theory of individuation. I am simply suggesting that qua information processors these computational mechanisms might have additional non-representational properties that would allow us to count them.

remarked in section 2, TVSH consists of three parts: (1) an anatomical model, (2) a functional perception-action model (strong PAM), and (3) a theory about the substrate of visual consciousness. Philosophers have predominantly been preoccupied with the third part, concerning the substrate of visual consciousness, and they have primarily based their analyses upon a particular version of (2), namely strong PAM. But it is important to notice that strong and weak PAM are not only both consistent with the received TVSH view that visual consciousness is the result of ventral stream processing; they are also both consistent with the denial of the received TVSH view on visual consciousness. In other words, part 2 and 3 of TVSH are logically independent of each other. This has a number of consequences.

First, even if weak PAM is logically consistent with the received TVSH story about the ventral stream substrate of visual consciousness, weak PAM undermines some of the motivation for this TVSH view. Some of the attraction of the official TVSH view of visual consciousness comes from the idea of the action control system as a 'zombie' within the brain, out of touch with consciously controlled processing and deliberation (Koch and Crick, 2001). This idea is fuelled by strong PAM and by the claim that action systems and perception systems are independent of each other. This latter claim is inconsistent with weak PAM, according to which action and perception are driven by the same computational mechanism. So, even if the dorsal stream were not a substrate of consciousness, it would not be strangely decoupled from perception.

Second, there are also implications for certain philosophical arguments against TVSH. Given the logical independence, proponents of TVSH have a possible reply to the objection that since some dorsal stream processing is conscious, TVSH must be false (Briscoe, 2008; Wu, 2014). They could hold on to strong PAM and give up the idea that TVSH is a theory of visual consciousness. Given the complexities involved in determining what counts as phenomenally conscious, this might be a natural response. It has already been acknowledged that ventral stream processing is often nonconscious—and who really knows how to

determine the contents of 'overflowing' phenomenal consciousness (see Block, 2007; Kouider *et al*. 2010)?[21]

The argument of this paper provides us with good reasons for changing the orientation of current empirical research on the functional architecture of vision for cognition and action. Specifically, it indicates that we should stop trying to determine types of representations, and instead start trying to look for other ways in which to determine the nature and number of involved computational mechanisms. In other words, the commitment to a representationalist individuation of computational mechanisms turns out to have consequences for model testing in cognitive neuroscience.

Section of Philosophy

University of Copenhagen

References

Anderson, J.R. 1978: Arguments concerning representations for mental imagery. *Psychological Review*, *85*(4), 249.

Bermúdez, J.L. 2006: Arguing for eliminativism. In B. Keeley (ed.), *Paul Churchland: Contemporary Philosophy in Focus*. Cambridge: Cambridge University Press, 32–66.

Block, N. 2007: Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*(5–6), 481–99.

Briscoe, R. 2008: Another look at the two visual systems hypothesis: the argument from illusion studies. *Journal of Consciousness Studies*, *15*(8), 35–62.

---

[21] This retreat to strong PAM seems to be the view of Clark (2009, p. 1467): 'On those increasingly elusive questions concerning the nature and neural underpinnings of conscious visual experience itself, the jury (it seems to me) remains out. But whatever the outcome, there is no doubting the value and impact of the dual visual systems model itself'.

Final draft. To be published in *Mind & Language*, September 2017.

Bruno, N., Knox, P.V. and de Grave, D.D. 2010: A metanalysis of the effect of the Mueller-Lyer illusion on

saccadic eye movements: no general support for a dissociation of perception and oculomotor action.

Vision Research, 50, 2671–82.

Chalmers, D. J. 1996: Does a rock implement every finite-state automaton? *Synthese*, *108*(3), 309–33.

Christiansen, J.H., Christensen, J., Grünbaum, T. and Kyllingsbæk, S. 2014: A common representation of

spatial features drives action and perception: grasping and judging object features within trials. *PloS*

one, *9*(5).

Clark, A. 2001: Visual experience and motor action: are the bonds too tight? *Philosophical Review*, *110*(4),

495–519.

Clark, A. 2007: What reaching teaches: consciousness, control, and the inner zombie. *The British Journal for*

the Philosophy of Science, *58*(3), 563–94.

Clark, A. 2009: Perception, action, and experience: unraveling the golden braid. *Neuropsychologia*, *47*(6),

1460–68.

Coltheart, M. 2001: Assumptions and methods in cognitive neuropsychology. In B. Rapp (ed.), *The*

Handbook of Cognitive Neuropsychology: What Deficits Reveal About the Human Mind. Hove, East

Sussex: Psychology Press, 3–21.

Coltheart, M. 2006a: What has functional neuroimaging told us about the mind (so far)? *Cortex*, 42, 323–

31.

Coltheart, M. 2006b: Perhaps functional neuroimaging has not told us anything about the mind (so far).

Cortex, 42, 422–7.

Coltheart, M. 2010: What is functional neuroimaging for? In S.J. Hanson and M. Bunzl (eds.), *Foundational*

Issues in Human Brain Mapping. Cambridge, MA: MIT Press, 263–72.

Cummins, R. 1989: *Meaning and Mental Representation*. Cambridge, MA: MIT Press.

Cummins, R. 1991: The role of representation in connectionist explanations of cognitive capacities. In W.

   Ramsey, S. Stich and D. Rumelhart (eds.), *Philosophy and Connectionist Theory*. Hillsdale, NJ: Lawrence

   Erlbaum, 91–114.

Cummins, R. 1996: *Representations, Targets, and Attitudes*. Cambridge, MA: MIT Press.

Davies, M. 2010: Double dissociation: understanding its role in cognitive neuropsychology. *Mind and

   Language*, 25(5), 500–40.

Dyde, R.T. and Milner, A.D. 2002: Two illusions of perceived orientation: one fools all of the people some of

   the time; the other fools all of the people all of the time. *Experimental Brain Research*, 144(4), 518–27.

Felleman, D.J. and Van Essen, D.C. 1991: Distributed hierarchical processing in the primate cerebral cortex.

   Cerebral Cortex, 1(1), 1–47.

Foley, R.T., Whitwell, R.L. and Goodale, M.A. 2015: The two-visual-systems hypothesis and the perspectival

   features of visual experience. *Consciousness and Cognition*.

Franz, V. and Gegenfurtner, K. 2008: Grasping visual illusions: consistent data and no dissociation. *Cognitive

   Neuropsychology*, 25, 920–50.

Franz, V.H., Gegenfurtner, K.R., Bülthoff, H.H. and Fahle, M. 2000: Grasping visual illusions: no evidence for

   a dissociation between perception and action. *Psychological Science*, 11, 20–25.

Franz, V.H., Hesse, C. and Kollath, S. 2009: Visual illusions, delayed grasping, and memory: no shift from

   dorsal to ventral control. *Neuropsychologia*, 47(6), 1518–31.

Ganel, T., Chajut, E. and Algom, D. 2008: Visual coding for action violates fundamental psychophysical

   principles. *Current Biology*, 18(14), R599–R601.

Ganel, T., Chajut, E., Tanzer, M. and Algom, D. 2008: Response: when does grasping escape Weber's law?

   Current Biology, 18(23), R1090–1.

Ganis, G. and Schendan, H.E. 2011: Visual imagery. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(3),

   239–52.

Ganis, G., Thompson, W.L. and Kosslyn, S.M. 2004: Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, 20, 226–41.

Gegenfurtner, K.R. and Franz, V.H. 2007: A comparison of localization judgments and pointing precision. *Journal of Vision*, 7(5), 11.

Goodale, M.A. 2014: Visual control of action. In K. N. Ochsner and S. M. Kosslyn (eds.), *The Oxford Handbook of Cognitive Neuroscience, Vol. 1*. Oxford: Oxford University Press, 273–95.

Goodale, M.A. and Haffenden, A. 1998: Frames of reference for perception and action in the human visual system. *Neuroscience and Biobehavioral Reviews*, 22(2), 161–72.

Goodale, M.A., Meenan, J.P., Bülthoff, H.H., Nicolle, D.A., Murphy, K.J. and Racicot, C.I. 1994: Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology*, 4(7), 604–10.

Goodale, M.A., Milner, A.D., Jakobson, L.S. and Carey, D.P. 1991: A neurological dissociation between perceiving objects and grasping them. *Nature*, 349, 154–6.

Goodale, M.A. and Milner, A.D. 1992: Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–5.

Goodale, M.A. and Milner, A.D. 2010: Two visual streams: interconnections do not imply duplication of function. *Cognitive Neuroscience*, 1, 65–8.

Grünbaum, T. 2012: Commonsense psychology, dual visual streams, and the individuation of action. *Philosophical Psychology*, 25(1), 25–47.

Henson, R. 2005: What can functional neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology*, 58, 193–233.

Henson, R. 2006: Forward inference using functional neuroimaging: dissociations versus associations. *Trends in Cognitive Sciences*, 10(2), 64–9.

Hesse, C. and Schenk, T. 2014: Delayed action does not always require the ventral stream: a study on a patient with visual form agnosia. *Cortex*, 54, 77–91.

Final draft. To be published in *Mind & Language*, September 2017.

Karnath, H.O., Rüter, J., Mandler, A. and Himmelbach, M. 2009: The anatomy of object recognition—visual form agnosia caused by medial occipitotemporal stroke. *The Journal of Neuroscience*, 29(18), 5854–62.

Klatzky, R.L. 1998: Allocentric and egocentric spatial representations: definitions, distinctions, and interconnections. In C. Freska, C. Habel and K.F. Wender (eds.), *Spatial Cognition, An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*. Berlin: Springer Verlag.

Koch, C. and Crick, F. 2001: The zombie within. *Nature*, 411(6840), 893.

Kouider, S., De Gardelle, V., Sackur, J. and Dupoux, E. 2010: How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences*, 14(7), 301–7.

Kozuch, B. 2015: Dislocation, not dissociation: the neuroanatomical argument against visual experience driving motor action. *Mind and Language*, 30(5), 572–602.

Machery, E. 2012: Dissociations in neuropsychology and cognitive neuroscience. *Philosophy of Science*, 79(4), 490–518.

Machery, E. 2013: In defense of reverse inference. *The British Journal for the Philosophy of Science*.

Matthen, M. 2005: *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception*. Oxford: Clarendon Press.

Milner, D. and Dyde, R. 2003: Why do some perceptual illusions affect visually guided action, when others don't? *Trends in Cognitive Sciences*, 7(1), 10–11.

Milner, A.D. and Goodale, M.A. 2006: *The Visual Brain in Action* (2nd ed.) Oxford: Oxford University Press.

Milner, A.D., Perrett, D.I., Johnston, R.S., Benson, P.J., Jordan, T.R. and Heeley, D.W. 1991: Perception and action in 'visual form agnosia'. *Brain*, 114, 405–28.

Mole, C. 2009: Illusions, demonstratives, and the zombie action hypothesis. *Mind*, 118(472), 995–1011.

Mole, C. and Klein, C. 2010: Confirmation, refutation, and the evidence of fMRI. In S.J. Hanson and M. Bunzl (eds.), *Foundational Issues in Human Brain Mapping.* Cambridge, MA: MIT Press.

Piccinini, G. 2015a: Computation in physical systems. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition)

<http://plato.stanford.edu/archives/sum2015/entries/computation-physicalsystems/>.

Piccinini, G. 2015b: *Physical Computation: A Mechanistic Account*, Oxford: Oxford University Press

Poldrack, R.A. 2006: Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10, 59–63.

Poldrack, R.A. 2010: Mapping mental function to brain structure: how can cognitive neuroimaging succeed? *Perspectives on Psychological Science*, 5, 753–61.

Putnam, H. 1967: The nature of mental states. In W.H. Capitan and D.D. Merrill (eds.), *Art, Mind, and Religion*. Pittsburgh University Press.

Putnam, H. 1988: *Representation and Reality*. Cambridge, MA: MIT Press.

Ramsey, W.M. 2007: *Representation Reconsidered*. Cambridge University Press.

Schenk, T. 2012a: No dissociation between perception and action in patient DF when haptic feedback is withdrawn. *The Journal of Neuroscience*, 32(6), 2013–17.

Schenk, T. 2012b: Response to Milner et al.: grasping uses vision and haptic feedback. *Trends in Cognitive Sciences*, 16(5), 258.

Schenk, T., Franz, V. and 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), 812–18.

Schenk, T. and McIntosh, R.D. 2010: Do we have independent visual streams for perception and action? *Cognitive Neuroscience*, 1(1), 52–62.

Searle, J.R. 1990: Is the brain a digital computer? In *Proceedings and Addresses of the American Philosophical Association,* American Philosophical Association, 21–37.

Searle, J.R. 1992: *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Shepard, R.N. and Metzler, J. 1971: Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–3.

Final draft. To be published in *Mind & Language*, September 2017.

Sober, E. 1999: Testability. In *Proceedings and Addresses of the American Philosophical Association*, *Vol. 73, No. 2*, American Philosophical Association, 47–76.

Sober, E. 2008: *Evidence and Evolution: The Logic Behind the Science*. Cambridge University Press.

Spencer, C. 2007: Unconscious vision and the platitudes of folk psychology. *Philosophical Psychology,* 20, 309–27.

Sprevak, M. 2010: Computation, individuation, and the received view on representation. *Studies in History and Philosophy of Science Part A*, 41(3), 260–70.

Utz, K.S., Hesse, C., Aschenneller, N. and Schenk, T. 2015: Biomechanical factors may explain why grasping violates Weber's law. *Vision Research*, 111, 22–30.

Wallhagen, M. 2007: Consciousness and action: does cognitive science support (mild) epiphenomenalism? *The British Journal for the Philosophy of Science*, 58(3), 539–61.

Westwood, D.A. and Goodale, M.A. 2011: Converging evidence for diverging pathways: neuropsychology and psychophysics tell the same story. *Vision Research*, 51(8), 804–11.

Whitwell, R.L., Milner, A.D., Cavina-Pratesi, C., Byrne, C.M. and Goodale, M.A. 2014: DF's visual brain in action: the role of tactile cues. *Neuropsychologia*, *55*, 41–50.

Wokke, M.E., Scholte, H.S. and Lamme, V.A. 2014: Opposing dorsal/ventral stream dynamics during figure-ground segregation. *Journal of Cognitive Neuroscience*, 26(2), 365–79.

Wu, W. 2013: The case for zombie agency. *Mind*, 122(485), 217–30.

Wu, W. 2014: Against division: consciousness, information and the visual streams. *Mind and Language*, 29(4), 383–406.