Australasian Journal of Philosophy

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rajp20

Species of Realization and the Free Energy Principle

Michael D. Kirchhoff[a]

[a] University of Wollongong

Published online: 17 Dec 2014.

To cite this article: Michael D. Kirchhoff (2014): Species of Realization and the Free Energy Principle, Australasian Journal of Philosophy, DOI: 10.1080/00048402.2014.992446

To link to this article: http://dx.doi.org/10.1080/00048402.2014.992446

Australasian Journal of Philosophy, 2014
http://dx.doi.org/10.1080/00048402.2014.992446

Routledge
Taylor & Francis Group

SPECIES OF REALIZATION AND THE FREE ENERGY PRINCIPLE

Michael D. Kirchhoff

This paper examines, for the first time, the relationship between realization relations and the free energy principle in cognitive neuroscience. I argue, firstly, that the free energy principle has ramifications for the wide versus narrow realization distinction: if the free energy principle is correct, then organismic realizers are insufficient for realizing free energy minimization. I argue, secondly, that the free energy principle has implications for synchronic realization relations, because free energy minimization is realized in dynamical agent-environment couplings embedded at multiple time scales.

Keywords: free energy, realization, dynamical systems

1. Introduction

My topic is the potential confluence between work on the metaphysics of realization in philosophy [Wilson 2001, 2004a, 2004b; Gillett 2002, 2003, 2007; Shapiro 2004; Polger 2007, 2010; and Aizawa and Gillett 2009a, 2009b] and research on the free energy principle in cognitive neuroscience [Friston 2002, 2003, 2010, 2011; Hohwy *et al.* 2008; Friston *et al.* 2012; and Clark 2013]. The free energy principle is a variation of the predictive brain hypothesis, which states that the brain is constantly making predictions about potential future events. It is the growing consensus in theoretical and systems neuroscience that a fundamental feature of neural computation is that the brain is always trying to reduce prediction error.

In the philosophical literature, no work has yet been done to examine what the implications of the free energy principle are for realization relations. Yet, as we shall see, certain notions of realization are fundamentally inconsistent with the free energy principle. The free energy principle starts from the assumption that biological systems resist a natural tendency for disorder by minimising free energy. Free energy is an upper bound on 'surprise'—the average surprise being entropy. In realization terms, the free energy principle is the view that free energy minimization is realized by physical systems—e.g. the brain—in a way that requires free energy to be a property of physical brain states—or by states in an extensive system including neural structures but not limited to the brain [Bruineberg and Rietveld 2014]. The dynamics inherent in free energy minimization suggest a temporal extensiveness into realization relations that—in dynamical systems theory—is associated with reciprocal, circular, causation. Unlike 'standard' accounts of realization, generally thought to be synchronic relations, I will discuss realization of free energy minimization in terms of

diachronic realization. In addition to this, certain versions of realization view the properties realized in a one-to-one fashion as being such that the realized properties an entity possesses are intrinsic (internal) to that entity. Yet this notion of realization is incompatible in an important way with free energy minimization. Consider that free energy is itself conditioned on the external environment within which the system is embedded. As we will see, free energy is an upper bound on surprise, and the state of a specific system—in terms of its surprise—depends upon the state of the world that is 'surprising'. This has implications for the nature of realization and, in turn, for the mapping between the states of a system and the properties realized by those states.

Before examining these consequences of the free energy principle for realization relations, I start by introducing the state of play in discussions of realization.

2. Species of Realization

In his classic paper 'Minds and Machines' [1960], Putnam introduced the conception of realization, formulating the relationship between the mental and the physical as one of realization. The argument given by Putnam was based on an analogy with the relation between the physical arrangement of matter and the abstract operations of a Turing machine implemented by that physical arrangement of matter. Thus, Putnam made the distinction between the logical description of a Turing machine and the physical states realizing the states referred to by that logical description, the idea being that mental states are realized by brain states in just this sense [Wilson 2004a: 101]. Despite Putnam's contribution to the debate on realization, only recently have philosophers started to scrutinise the metaphysics of the realization relation [Wilson 2001; Gillett 2002, 2003, 2007; Polger 2004, 2010; Shapiro 2004; and Aizawa and Gillett 2009a]. Unsurprisingly, this work has not produced unanimity about the correct theory of realization. Despite this lack of consensus, reflection on a couple of recent treatments—Wilson [2001, 2004b, 2004b], Gillett [2002, 2007], and Aizawa and Gillett [2009a]—will help to sketch several issues about realization that will be discussed.

First, the realization relation is thought to be a relation of ontological dependence [Gillett 2007: 166]. The received view amongst realization theorists, regardless of their different persuasions, is that this form of dependency is *synchronic* (i.e. holds at an instant *t*). Bennett specifies this assumption by saying the following [2011: 93–4]:

> Building relations do not unfold over time. If property *P* realizes property *Q*, it does so at some time *t* … Causation, in contrast, is paradigmatically *diachronic*, and that idea is frequently invoked to distinguish causation from relations like [realization].

If so, realization is a non-temporal kind of dependence, whereas causation, for example, is a temporal dependence relation.

Second, consider the distinction between what Gillett calls the *flat* view and the *dimensioned* view of realization. The flat view, which Gillett also dubs the standard view, is a one-to-one mapping between realizer/realized, while the dimensioned view is a many-to-one relation between qualitatively distinct properties. Gillett defines the flat view as the conjunction of two theses. The first thesis concerns *sufficiency*, such that property P realizes property Q only if the causal powers of Q are collectively a subset of P. Gillett calls this the *metaphysical sufficiency thesis* [2007: 174]:

(1) Property instances P1−Pn are realizers of property instance F, at time t, *if and only if* P1−Pn are a minimal combination of property instances which are together metaphysically sufficient for an instance of F at t (MS-thesis).

The second thesis concerns the individuals in which the properties are instantiated. Gillett frames this idea accordingly: 'A property instance X realizes a property instance Y *only if* X and Y are instantiated in the same individual' [2002: 317].

Gillett thinks that the flat view is inadequate and provides an alternative dimensioned view. The dimensioned view is a many-to-one mapping between different realizer/realized properties. Take Gillett's example [2002] of a cut diamond, S^*. S^* instantiates the property of being very hard, H. Suppose H is composed by carbon atoms $S1-Sn$, and that $S1-Sn$ have the properties of being bonded, $B1-Bn$, and of being aligned, $A1-An$. While S^* has H, it is not the case that S^* has $A1-An$ and $B1-Bn$. Similarly for the carbon atoms, which have $A1-An$ and $B1-Bn$ but not H. The dimensioned view is (on the face of it) compatible with the hierarchical composition of the free energy principle, such that the free energy of constituents combines the free energy of an ensemble.

The distinction between the flat view and the dimensioned view, together with the fact that realization is understood as a synchronic relation, gives us some idea of the different species of realization and their properties.

Yet even this formulation of realization remains incomplete. According to Wilson [2001, 2004a, 2004b; Wilson and Clark 2009], there are cases of the realizer/realized relation in which the realizers of some realized property P extend beyond the boundary of the individual bearer, IB, who has P. This formulation rejects the second thesis of the flat view (outlined above). As Wilson states: 'wide realizations ... extend beyond the physical boundary of the individual, they are not exhaustively constituted by the intrinsic, physical properties of the individual subject' [2001: 12]. However, the wide realization view accepts the metaphysical sufficiency thesis, because it is only the physical properties constituting a total realization *together with the appropriate background conditions* that suffice for P. This is consistent with the free energy formulation, where the appropriate environmental realizers of free energy minimization are associated with external states. These external states are commonly referred to as hidden causes in the environment, where surprise—in the free energy formulation—is conditionally dependent on

predictive representations (or posterior beliefs) about the causes.[1] Consolidating the free energy principle with the view of wide realizations gives us the following: only properties instantiated *within* the individual *together with* properties instantiated beyond that individual's biological boundary will metaphysically suffice for a certain realized property.

With this overview of realization, I turn to developing my arguments that have been outlined above.

3. Argument #1: Realization, Wide Realization, and the Free Energy Principle

The first argument is as follows. The flat view of realization states that both the realizer and the realized are instantiated within one and the same individual. But this is fundamentally inconsistent with the free energy principle, where the minimization of free energy and its physical realizers are instantiated in different individuals.

As such, that argument supports Gillett's critique of the flat view. It poses a challenge to Gillett's account of the dimensioned view if his account fails to allow for the fact that the realizers of free energy minimization go beyond the individual organism. For example, advocates of extended and enactive accounts of cognition state that there is a brain-body-world system that realizes cognitive activity [Noë 2004; Clark 2008]. This is in line with the free energy principle, which states that free energy minimization is realized in a dynamical organism-environment coupling [Friston 2011]. Any relevant properties attributed to the organism itself are relational properties and they fail to be wholly realized within the skin-and-skull. This brings the dimensioned view and the free energy principle into contact with wide realization.

3.1. Argument from Thermodynamics

The free energy principle states that all physical systems (in order to survive) must actively resist a natural tendency for disorder (Friston 2003, 2009, 2010, 2011; see also Ashby [1952] and Haken [1983]). This is the thermodynamic starting point of the free energy principle, and it brings the free energy principle into alignment with principles of dynamical systems. The central premise of dynamical systems is that physical systems in general, and biological systems in particular, belong 'to a class of systems that are both *complex* and that exist *far from thermodynamic equilibrium*' [Thelen and Smith 1994: 51]. Biological systems are complex in the sense that they consist of multiple components, and these components tend to be different with disparate properties and causal powers. Biological systems exist in far-from-thermodynamic equilibrium, because such systems contravene the second law of

[1]Despite the current unpopularity of notions such as 'representation' and 'mental state' in dynamical and enactive approaches to cognitive science, I continue to use these terms here. I do this because the aim of this paper is not to engage in discussions of the status of representational analysis in cognitive science. However, unlike the predictive processing accounts considered here, most of which allow for representational talk, there is no *in principle* reason to think that free energy minimization *must* involve representations. Whether or not the nested and hierarchical architecture realizing free energy minimization is representational is an important question to consider—a task for another occasion.

thermodynamics. The second law of thermodynamics states that entropy (i.e. a measure of disorder) of closed systems increases over time [Friston 2010: 127].

Organisms are capable of maintaining reduced levels of entropy in the face of fluctuations and increasing levels of entropic disorder in the external environment. In his discussion of which kinds of properties and components realize this capacity, Kemp [1982] mentions, among other things, the role of the temperature regulating system in sustaining appropriate levels of internal temperature. But this is not possible without some sort of blood filter (i.e. circulatory system), the property of which is to filtrate and pump blood throughout the body [Cosmelli and Thompson 2010]. We do not need to add additional components and properties to this example in order to establish the following: an organism is able to maintain an upper bound of entropy, and this ability is realized by different physical components in a system.

But some philosophers have tried to resist this conclusion by postulating certain combinations of the components and suggesting that it is this combination, rather than the single parts, that is the individual that realizes a certain higher-level property. Indeed, if a combination of components is taken as the putative realizer of a certain property, it is not implausible that this combination realizes some property in a way that is compatible with the flat view. Following Gillett [2002], let us call this proposed combination COMBO. In his discussion of the flat account, Gillett [2002] provides a promising argument against the success of this response. As he says, 'we can quickly show that the response simply relocates the problems facing the Flat view without ameliorating them' [2002: 320]. According to Gillett's argument, our best physics informs us that, at the most fundamental level, individuals are quarks and leptons, and these particles have properties such as change, charm, spin, and so on. If we return to Gillett's example of a cut diamond S^*, with the realized property of being very hard, H, it follows by transitivity that if COMBO (now made up at the level of the diamond) realizes H, and COMBO (ultimately) is realized by certain fundamental microphysical properties/relations, *MPs*, then H is realized by the *MPs*, which are instantiated in individuals different from S^*.

Presumably, if COMBO were raised against the property of entropy regulation, it would face similar problems as those pointed to by Gillett. Indeed, if Gillett is correct to insist that components in the realization relation are individuals, it follows that the property of entropy reduction is realized in different individuals.

If this is true, we have an argument against the assumption of the flat view: that the realization relation is a one-to-one relation that holds within one and only one individual. *Prima facie*, at least, this supports the dimensioned view. However, I shall now show that Gillett's presupposition—that the components, whose properties enter into relations of realization, are spatially contained within the individual associated with the composed entity—is problematic. If the world is as stated by the free energy principle, then the systemic parts and their properties that realize free energy minimization are not wholly contained within an individual organism, but they include

(necessarily so) components and properties of that individual's extra-neural and/or extra-bodily environment.

In contrast to closed systems, biological systems are open systems. Open systems are dissipative systems, i.e. such systems preserve their order while immersed in a dynamical environment by exchanging energy or matter with that milieu. Combining the thermodynamics of the free energy principle with realization leaves us with the following. Call H the property of self-maintenance, X the process of drawing energy from the environment, P the process of manipulating some energy source, and R the process of dissipation. We may then state, more precisely, that H is realized by processes X, P, and R. However, here X can at best be a *partial* realization of H—and similarly for P and R. A partial realization is what Shoemaker [1981] calls a *core* realization, which is a particular component of the central nervous system, say, that is identifiable as performing a core role in bringing about H. If this turns out to be correct, partial realizations alone will not satisfy as being metaphysically sufficient for H. This does not yet provide us with an argument against the flat view. But that core realizers are insufficient for H identifies the need for something extra. According to Shoemaker, when considering the relation between some realized state or process such as H, and the system, S, in which H is realized, one must distinguish between *core* realizations and *total* realizations. In his discussion of Shoemaker's account, Wilson provides the following definition of core- and total-realization [2001: 8]:

(a) *core* realization of H: a state of the specific part of S that is most readily identifiable as playing a crucial role in producing or sustaining H.

(b) *total* realization of H: a state of S, containing any given core realization as a proper part, that is metaphysically sufficient for H.

Wilson does not discuss the case of self-maintenance, even though he uses the placeholder 'H' in his definitions. With this clarified, consider that total realizations of H include X, P, and R. In this sense, total realizations are *complete* realizations. But if the free energy principle is correct, then even if total realizations are complete they are still metaphysically insufficient for H, since—as Wilson would say—the total realization of H 'excludes the *background conditions* that are necessary for there to be the appropriate, functioning system' [2001: 9].

If we consider the thermodynamic formulation of the free energy principle, it becomes apparent why it is only the physical states that make up the total realization in conjunction with the appropriate extra-bodily properties that will suffice—metaphysically—for realizing H. A *total* realization of H includes X, P, and R. However, excluded from the total realization of H is the necessary fact that the environment itself 'unfolds in a thermodynamically structured and lawful way' [Friston and Stephan 2007: 422], which is necessary for the system S to function the way it does. Call these environmental realizers of H, ER. Importantly, ER's are not part of the total realizations of H, instantiated in S, since the ERs are not properties of S—the

individual within which the total realizations of H are contained. Thus, if H is a realized property, the realizers of H are not wholly contained within the individual S, where S is the individual instantiating H.

There is at least one reason to believe that this outcome (against the dimensioned view) is premature. For example, one could argue that this is a challenge to Gillett's view only if he cannot say that the individual realizing free energy minimization is, in fact, something larger than the individual organism. Gillett could attempt to argue that there is a brain-body-world system, and that it is this extensive system that realizes free energy minimization. But, in his [2007] paper, Gillett provides a critique of the appeal to extra-organismic (external) entities as actual physical realizers, due to the fact that such an appeal turns on the view that 'external' entities are elements of a metaphysical sufficiency condition for the realized entity. On Gillett's view, however, metaphysical sufficiency 'leads to scientific hyper-extension ... by placing realizers, and parts, beyond the normal scientific limits and understanding' [2007: 176]. Gillett argues, that, in scientific examples, realizers are not metaphysically sufficient for realized properties, because—strictly speaking—only realizers together with entities that function as background conditions are sufficient for the realized properties. Gillett's view is based on what he terms '[our] well-confirmed scientific theories' [2007: 176], which he finds in his analysis of examples from chemistry and biology.

Even if Gillett is correct in what he takes to be our well-formed scientific theories in chemistry and biology, it is important to mention that the free energy minimization formulation is premised on the fact 'that the environment unfolds in a thermodynamically structured and lawful way and biological systems embed these laws into their anatomy' [Friston and Stephan 2007: 422]. Here is a scientific theory that takes it as an integral fact that certain nonneural and nonbodily properties of the environment are necessarily part of the realization base of an organism's ability to reverse an increase in entropy over time. Furthermore, Gillett's appeal to 'spatial containment', as he specifies that parts and their properties be spatially bounded within the individual that is associated with the entity composed, does not find any corresponding image in physics. Indeed, as Ross and Ladyman state: 'The types of particles which physical theory describes do not have spatiotemporal boundaries in anything like what common sense takes for granted in conceptualizing everyday objects, and in that respect are not classical individuals' [2010: 156].

3.2. Introducing Predictive Processing in the Free Energy Principle

Similar problems with both the flat view and the dimensioned view arise when we consider the information-theoretic perspective of the free energy principle. Before looking at the supporting evidence for this claim, I need to introduce some important concepts and distinctions.

Free energy 'bounds surprise, conceived as the difference between an organism's predictions about its sensory inputs ... and the sensations it actually encounters' [Friston *et al.* 2012: 1]. The information-theoretic formulation of the free energy principle states the following: 'Organisms that

succeed [in minimising free energy] do so by minimizing their tendency to enter into this special kind of surprising (that is, non-anticipated) state' [Friston *et al.* 2012: 1]. The brain does this by enacting processes whose function is to minimize 'prediction error'. Prediction error refers to the difference between sensory input from the environment and active predictions of such input in virtue of certain internal states of an organism or system [Friston and Stephan 2007; Friston 2010]. If there is a mapping from the internal states to the sensory input, the former are said to predict the hidden causes of the sensory inputs—where hidden causes are the causes of the sensory impressions on the organism. This is the Bayesian view of prediction error reduction. In the brain, the prediction errors are linked to superficial pyramidal cells located in the upper layers of the cerebral cortex [Hohwy *et al.* 2008], while top-down predictions, required to form prediction errors in the sensory cortex, are realized by activity in deep pyramidal cells [Brown *et al.* 2011].

The (neural) dynamics involved in free energy minimization include recurrent information passing between the superficial pyramidal cells and the deep pyramidal cells at multiple cortical levels [Friston 2003; Clark 2013]. This is the hierarchical organization of the predictive architecture. Briefly, recurrent information processing is mediated by bottom-up prediction errors and top-down predictions [Howhy *et al.* 2008]. Bottom-up prediction errors—through recurrent processing loops—optimize the top-down predictions, thus (eventually) cancelling the prediction error itself. This is also known as 'explaining away' [Clark 2013]. In information processing terms, via a cascade of top-down predictions, the sensory input is explained away, leaving only the prediction error to be passed on in the system.

3.2.1. Argument from Predictive Processing in the Brain

If this account of the brain's information-theoretic operations is correct, then the flat view of realization is inconsistent with the free energy principle. In contrast to the flat view, which states that the realizers are intrinsic (internal) to the individual whose states or properties they are, free energy is realized only in relation to external states and the dependencies on those states by the system in question.

Consider, initially, what Rao and Ballard say about the bidirectional connectivity in predictive hierarchical architectures [1999: 80]: '[Prediction] and error-correction cycles occur concurrently throughout the hierarchy, so top-down information influences lower-level estimates, and bottom-up information influences higher-level estimates of the input signal.' An example of this in operation is the phenomenon of binocular rivalry.

Binocular rivalry is a form of subjective visual experience that occurs, in a special experimental setup, when one stimulus is shown to one eye and another stimulus is shown to the other eye. For example, when an image of a house is presented to the right eye and an image of a face to the left eye, the subjective experience tends to unfold in a bi-stable manner by alternating between the house and the face. This is what is known as binocular rivalry.

As Hohwy *et al.* mention, in order to account for binocular rivalry two parts need explanation. First, there is the *selection problem*: 'why is there a perceptual decision to select one stimulus for perception rather than the other, and, further, why is one of the two stimuli selected rather than some conjunction or blend of them?' [2008: 690]. Second, there is the *alternation problem*: 'why does perceptual inference alternate between the two stimuli rather than stick with the selected one?' [2008: 690]

From the perspective of Bayesian inference, if a subject is currently experiencing an image of a face, F, why then does the F hypothesis have the highest probability, given that F and H have an equal likelihood? This is the selection problem. The alternation problem is to explain why the system (the brain), having selected F, say, after only a few seconds deselects in favour of H. Note that, for my present purposes, discussion of both the selection and the alternation problem is unnecessary, so I shall restrict my attention here to the selection problem. According to Hohwy *et al.* the predictive processing framework posits a hierarchical inversion of generative models of how inputs are caused to explain the selection problem [2008: 691]:

> At the higher, hypothesis-generating level only the currently best hypothesis is allowed to generate predictions. It seems plausible that inhibition will be lateral, in relation to other hypotheses at the same level. This gives high activity for the winning hypothesis with the highest posterior and thus for the dominant percept, and lower activity for other hypotheses at that level. At the lower level there is the opposite pattern: the bottom-up driving signal for the dominating percept is explained away by good predictions, meaning the prediction error for the dominant hypothesis is suppressed. Conversely, the bottom-up error signal for the currently suppressed stimulus is not.

As with the property of *self-maintenance*, free energy minimization is ineliminably relational, with all areas (thalamo-cortical, cortico-cortical) at work simultaneously, yet at different temporal frequencies. Order is maintained in the overall processing, Friston and Stephan explain, through *synchronous* activity in the various top-down and bottom-up loops in the hierarchical architecture [2007: 443].

In dichoptic viewing conditions, when F is viewed by one eye and H by the other, the hypothesis with the highest *prior probability* (how probable the prediction was before the input) could be considered as a core realization, viz. a specific part of S (the system) that is identifiable as playing a crucial role in producing the realized property. But selected hypotheses are metaphysically context-sensitive, in the sense that they will have a visual experience of F, say, only in relation to their activity and location within a generative hierarchical organization [Friston 2002].

While total realizations are said to be complete realizations, the assumption of the flat view that realized properties are realized by physical intrinsic realizers is inconsistent with the free energy principle, because the mechanism integrating the parts in the overall processing architecture is *temporal synchrony* [Varela *et al.* 2001; Engel *et al.* 2001; Friston 2003; Engel 2010]. In

contrast to the idea that properties are intrinsically realized, temporally synchronous patterns are extrinsic properties of a dynamical system such as the brain. So, if the free energy principle is indeed the right way to understand neuro-dynamical functioning, then the flat view of realization is inadequate, since the physical realizers are themselves entirely relational.

3.2.2. Argument from Wide Predictive Processing

What I call the argument from wide predictive processing shows that the reference to spatial containment in the dimensioned view is problematic. The argument is as follows. If active manipulation of extraneural resources, embedded in ongoing loops of action and perception, afford extra-bodily circuitry for minimization of prediction error, the spatial containment condition is itself inconsistent with the free energy principle. This lends support to a wide conception of realization.

Roepstorff *et al.* [2010] suggest that the brain is a hierarchically organized predictive machine, which attempts to anticipate its sensory inputs based on empirical priors of causes in the environment—empirical priors are also known as hidden causes. That idea finds its fullest expression in the *patterned practice approach* in social anthropology and social neuroscience. The idea is that, just as top-down predictions modulate bottom-up input, so too can socially embedded and culturally transmitted practices be understood as modulatory. Roepstorff and Frith [2004] provide evidence for this, as they consider the concept of 'top-top' modulatory control of action in a study of the 'Wisconsin card-sorting task' (WCST). Based on brain imaging experiments, Roepstorff and Frith argue that the state-oriented (here-and-now time perspective) 'top' in 'top-down' driving and modulatory control of action, rather than being conceived of as *internal* to the experimental participant, is in fact socially distributed across the experimenter and experimental participant in cognitive experiments. Roepstorff and Frith focus on several experiments. I shall focus on one of these, namely a study of the cross-species neural correlates of action conducted by Nakahara *et al.* [2002].

Nakahara *et al.* had two macaque monkeys perform a version of the WCST. The WCST consists of four cards and 128 response cards with geometric figures that vary according to perceptual dimensions such as colour, form, or number. The experimental participant is presented with cards that display specific symbols of one of the three perceptual dimensions, such as three green circles, or three yellow triangles, etc. The task requires the participant to find the correct classification rule, viz. sorting criteria. During the task, the participant is given feedback related to the correctness of their sorting. Once the participant chooses the correct rule, they must maintain the use of this rule, irrespective of the fact that the stimulus changes. After a certain number of correct matches, the experimenter changes the sorting criteria without warning, demanding that the participant discover the new classification rule.

During the task, Nakahara *et al.* had the two monkeys perform a computerized version of the WCST, where the monkeys had to select one of three cards, relative to the classification rule in use at the time of sorting.

Nakahara *et al.* also had 10 human subjects perform the same task. To perform well in the WCST, the participant must enact and modify a particular cognitive set, 'which can be used as a template for acting in the world' [Roepstorff and Frith 2004: 191]. According to Roepstorff and Frith, this is a clear indication of a top-down control of action in both an anatomical sense (from prefrontal to lower brain areas) and in the predictive processing sense (denoting a form of hypothesis or prediction-driven processing). Adding to the central conclusion by Nakahara *et al.* that there is evidence of cross-species neural correlates of action, Roepstorff and Frith provide an alternative interpretation of the experimental outcomes, one that is based on a combination of patterned practices and the different developmental and social trajectories between the macaque monkeys, on the one hand, and the humans, on the other.

The important thing to note is that whereas it took Nakahara *et al.* up to one full year of training to get the monkeys to perform the WCST in an MRI scanner, it took the human participants only 30–60 minutes of verbal instruction to perform equally well. Thus, despite displaying similar patterns of behaviour and brain activation, the learning trajectory between the two species is quite different.[2] In the human case, Roepstorff and Frith stress, the internal top-down story breaks down. On the standard view, bottom-up effects are driven through sensory inputs established 'from the outside', whereas top-down predictions are generated 'from the inside', e.g. via knowledge-driven predictions about the causes of the sensorium. But, as Roepstorff and Frith argue [2004: 192],

> [the] 'verbal instructions' that enable the human volunteers to perform well in the task, fail to fit this scheme. The instructions are clearly coming 'from the outside' and are mediated via the senses, i.e. bottom up, and yet their main purpose is to allow for the very rapid establishment of a consistent model of how the participants are to interpret and respond in the situation, i.e. top-down.

The main result to which Roepstorff and Frith point is that, given this breakdown of the conventional model of the 'top' in top-down processing, 'the origin of the "executive top" employed in the WCST is outside the brain of the participant, namely [socially mediated by the] experimenter' [2004: 192]. They refer to this socially mediated form of interaction between the experimenter and the participant as a 'top-top exchange of scripts' [2004: 192]. Scripts are 'shared representations' enacted in situated practices, where shared representations concern top-level aspects of control (that is, the goal of the task) instead of low-level processes concerning, for example, how specific movements should be made. According to Roepstorff *et al.* [2010: 1056]:

> From the inside of a practice, certain models [i.e. certain ways of interacting with one another] of expectancy come to be established, and the patterns, which over time emerge from these practices, guide perception as well as action.

[2]Whether this is entirely accurate is questionable. That is, if there is a difference in 'learning trajectories', then—all things being equal—there must be a difference in both 'patterns of behaviour' and 'brain activity'. If that is not the case, how could we account for the difference in learning histories? Thanks to an anonymous referee for this point of clarification.

As with predictive processing in the brain, one property of dynamical processes that regulates the coherence and resonance between patterns of expectancy in the brain and patterns of expectancy unfolding in the social context is *temporal dynamics*. For instance, forward neural connections mediate their post-synaptic effects over very fast timescales, ranging from 1.5–6 ms decay time, while backward neural connections are mediated by slower dynamics, with ~50 ms decay time [Friston 2003: 1328].

According to Friston, slower neural dynamics mediate contextually enduring effects, which is why backward neural connections can modulate forward neural connections. This difference between forward and backward neural connections is referred to by Friston as 'functional asymmetry', to emphasize the difference in functional role between those neural connections. In the WCST case, the proposal is (among other things) that top-down predictions—in the context of culturally mediated practices—take the form of socially situated top-top interaction in patterned practices. The interactions between experimenter and participant display temporal dynamics that are much slower (ranging from 30–60 minutes) than bottom-up and top-down neural connections. If slower evolving dynamics mediated contextually relevant information, situated practices may display modulatory effects.

Recall that the dimensioned view assumes that components and their properties are spatially contained within the individual associated with the composed entity. On Gillett's view, if we treat cultural practices and other extra-bodily components and properties as *physical realizers* of some property, we fail to discriminate between physical realizers and the *background conditions* (i.e. causal conditions) necessary for those physical realizers [2007: 175]. Yet what if some properties of information processing simply could not be realized in the absence of particular ways of interacting in the world? Then we would have reason to believe that specific situations and cultural practices are not merely causally enabling or necessary background conditions for predictive processing, but in fact are realizers. This is the impetus behind Wilson's account of *wide realization* [2001; Wilson and Clark 2009; Clark 2013].

We usually distinguish between realizers and background conditions, through an analysis of which parts and other components are plausibly entities that *play the most salient causal role* or simply *play the role* of the composed entity. In other words, realizers are entities whose productive 'causal function' results in productive 'causal functions' of the composed entity. By contrast, entities that are merely background conditions do not 'play the role' of the composed entity.

From a patterned practice approach in social neuroscience [Roepstorff and Frith 2004; Roepstorff *et al.* 2010], it would seem that the patterned interactions between the experimenter and the participant could not simply be 'screened off' as background conditions for predictive processing, precisely because those interactions guide as well as modulate perception and action. Indeed, those *patterns of expectancy* play *part of* the role of the realized process of free energy minimization. If correct, this backs a wide realization base for free energy minimization.

4. Argument #2: Realization, Synchronicity, and the Free Energy Principle

The second argument turns on the idea that realization is a synchronic relation. If free energy minimization is a realized property, the relation between the realizer/realized is such that there is a synchronic relation between the realized property and its physical realizers. The synchronic claim is that the realizers of some property exist at the same time as the realized property. There is no temporal delay, say, between the occurrence of the physical realizers and the realized. But free energy minimization involves couplings between dynamic activities at different temporal scales. In the context of realization mappings, this temporally unfolding activity—at multiple levels in the processing hierarchy—may be better thought of as diachronic realization.

The starting point for the argument is that it is coherent to distinguish between a conceptual argument for realization and an empirical argument for realization. That is, if there is a relation of realization between realizer/realized, then that relation must hold synchronically. This is a conceptual argument for the claim that realization is synchronic. We can outline the argument as follows: (i) the synchronic nature of realization serves to distinguish it from causation; (ii) causation is a diachronic relation; (iii) therefore, realization is not a diachronic relation. The inference from premises (i) and (ii) to the conclusion (iii) is valid. The problem with the argument is with the evidence for premise (i): that the synchronic nature of realization serves to distinguish it from the relation of causation.

No doubt, many readers are at this point eager to make the following objection—that I have misunderstood the synchronicity condition embedded in realization, since the synchronicity of realization is simply the claim that, no matter what property realizes the property of free energy minimization, that property (or those properties) must be occurring simultaneously with the realized property. But if free energy minimization is realized in the first place, it is not realized by a set of additional properties; rather, certain processes realize it. Strictly speaking, the property of free energy minimization comes to be instantiated in a process that emerges over time. Crucially, in dynamical systems such processes are embedded in nonlinear bottom-up and top-down processing. But if this is correct, whatever property a dynamical system exemplifies at a particular time t is, in part at least, a function of temporally prior states of the system. Realization of free energy minimization—in a dynamical system—at a certain time t is dependent on temporal dynamics at multiple scales in the entire system.

Suppose we agree that dynamical systems fail to instantiate the property of free energy minimization without that property being embedded in temporally unfolding and integrated ensembles of neuronal assemblies over time [Engel *et al.* 2001; Engel 2010]. We thus agree that the instantiation of free energy minimization necessitates the unfolding of complex and global temporal dynamics in the brain [Friston 2010]. Specifically, the rates of change within, the time-course of, and any time-dependent synchronization of individual neurons or neuronal ensembles are nontrivially part of the realization of free energy minimization. Indeed, if the free energy view is true,

the realization mapping between physical realizers and minimization of free energy cannot be understood to be synchronic—instead it must to thought of as diachronic. If this is correct, we thus have two conceptions of realization: synchronic versus diachronic realization. Although this is an important implication, it is not part of this paper to give a treatment of diachronic realization; the aim is only to point to its possibility—a full treatment of diachronic realization is a task for another occasion.

Consider the following illustration in Friston and Stephan [2007], depicting the quantities that define free energy:

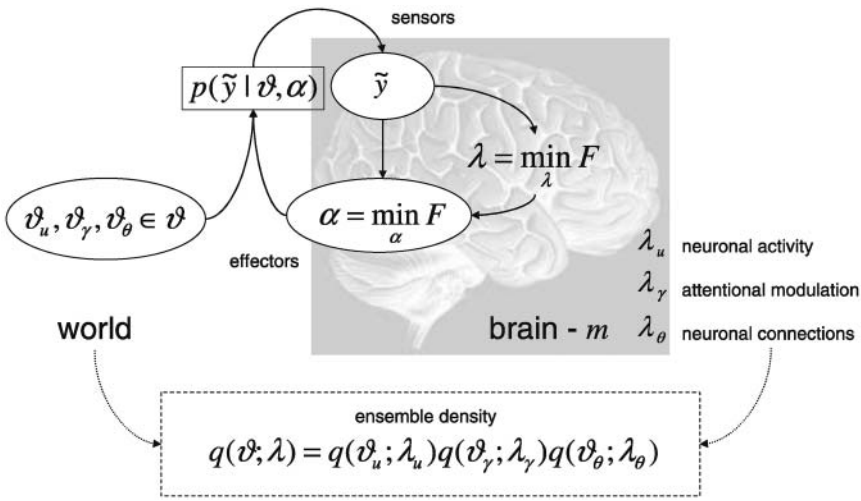

Figure 1: Illustration detailing the quantities that define free energy [Friston and Stephan 2007: 424; used with kind permission from Karl Friston].

This illustration informs us of two things. It describes the quantities of a system's (m's) interaction (or exchange) with the environment. And it indicates that all of the quantities that change do so in order to minimize free energy. The latter feature is shown in the formulation, $\lambda = \min F$ (λ refers to systemic quantities, whereas F denotes free energy). $\hat{y}$ characterises effects of the environment on the system; a is the effects of the system on the environment. According to Friston and Stephan, biological systems can minimize free energy by changing the two quantities on which free energy depends: (i) a system can act on the environment (a), thus changing the sensory input $\hat{y}$; (ii) or a system can change its posterior beliefs by changing its internal states, λ. The first path by which biological systems can minimize free energy can be represented as a conditional probability formulation of the form $p(\hat{y}) \rightarrow p(\hat{y} \mid a)$. In general, embodied manipulation of the environment can change the sensory input of m, given the general form of the probability condition, i.e. the probability of an effect e occurring, given that f occurs. Simply put, a system can minimize free energy by acting on the world, thus optimizing the accuracy of its own predictions by actively sampling and sculpting the environment. More formally, '($p(\hat{y} \mid \upsilon, a)$ is the conditional probability of

sensory input given its causes, υ, and the state of its effectors (i.e. action)' [Friston and Stephan 2007: 424] The remaining pathway by which a system can minimize free energy is by changing its internal states (λ). While portrayed as analytically separable, both pathways have functionally and structurally convergent dynamics. As Friston notes, '[i]nternal brain states and action minimize free energy … which is a function of sensory input and a probabilistic representation … of its causes' [2010: 128]. Thus, all the quantities that can change do so to minimize free energy.

One interpretation of figure 1, although arguably incorrect, is that free energy minimization is physically realized by quantities $q(\upsilon, \lambda) = q(\upsilon_u; \lambda_u)q(\upsilon_y; \lambda_y)q(\upsilon_\theta; \lambda_\theta)$ such that free energy minimization is realized at an instant t. The assumption that free energy minimization is realized at a single moment in time is an assumption that turns, I suspect, on our tendency to represent it spatially or pictorially, as in figure 1. But the spatial representation is misleading, in the sense that the spatial (inert) representation is not analogous to the temporal dynamics through which free energy minimization is realized.

We already know that all quantities involved in free energy minimization change to minimize free energy. Consider that the quantities describing both environmental (or hidden) causes, υ, and the quantities describing neural states unfold and change on a timescale of milliseconds, seconds, and minutes. As Friston and Stephan remind us, environmental causes could be large and heterogeneous in number. Recall that it is these environmental realizers—or hidden causes—on which free energy minimization is (in part) conditionally dependent.

Friston and Stephan point out that a 'key difference among them is the timescales over which they change' [2007: 429]. In figure 1, these environmental causes are partitioned into three sets, $\upsilon = \upsilon_u, \upsilon_y, \upsilon_\theta$, indicating change on a timescale of milliseconds, seconds, and minutes. According to Friston and Stephan, this 'induces a partitioning of the system's parameters into $\lambda = \lambda_u, \lambda_y, \lambda_\theta$ that encode time-varying marginals of the ensemble density' [2007: 429]. As they specify [2007: 429],

> [t]he first, λ_u, are system quantities that change rapidly. These could correspond to neuronal activity or electromagnetic states of the brain that change with a timescale of milliseconds. The causes υ_u they encode correspond to evolving environmental states, for example, changes in the environment caused by structural instabilities or other organisms. The second partition λ_y changes more slowly, over seconds. These could correspond to the kinetics of molecular signaling in neurons; for example calcium-dependent mechanisms underlying short-term changes in synaptic efficacy and classical neuromodulatory effects. … Finally, λ_θ represent system quantities that change slowly; for example long-term changes in synaptic connections during experience-dependent plasticity, or the deployment of axons that change on a neurodevelopmental timescale.

All of these quantities are part of the physical machinery realizing free energy minimization, and all of these quantities change to do so in virtue of evolving over time. If the quantities responsible for the realization of free energy minimization *change* to minimize free energy, and if these quantities

change differently across a timescale of milliseconds, seconds, and minutes, then this excludes the possibility of free energy minimization being synchronically realized.

If this is correct, and if one were to insist on free energy minimization having a realization base, then it seems that the best we could do is to say the following: *during that period of time* (however long or how brief it is), free energy minimization was realized by the quantities specified in figure 1. But it does not follow that, during this period, that minimization of free energy was synchronically realized by the quantities specified in figure 1, because synchronicity is thought to be non-temporal.

Moreover, if the world is as stated by the free energy principle, then another problem with invoking synchronic realization reveals itself by considering that a common strategy by which to identify what 'constitutes' a realization base of a certain realized property is that of appealing to whatever plays the most salient causal role(s) with regards to the instantiation of the realized property [Cosmelli and Thompson 2010: 364]. However, free energy minimization is instantiated in nonlinear dynamical systems. According to Cosmelli and Thompson [2010: 365]: 'In dense nonlinear systems in which all state variables interact with each other, any change in an individual variable becomes inseparable from the state of the entire system.'

If we accept that nonlinear dynamics is one fundamental property due to which the patterns of spatiotemporal neuronal assemblies minimize free energy—e.g. by constantly creating predictions about forthcoming sensory events [Engel 2010]—it seems a small step to accept that we cannot identify what 'constitutes' the realization base of free energy minimization at a synchronic point in time. Crucially, this shows that the minimization of free energy is an inherently temporal phenomenon.

Suppose that free energy minimization is dependent on the integration of multiple different ensembles and that the integration of such ensembles requires very precise temporal dynamics. Would that make a difference for how to assess the relationship between realization and free energy minimization? If realization is synchronically defined, and if free energy minimization is realized in temporal dynamic couplings at different 'levels', then this denies the possibility of synchronic realization in the context of free energy minimization.

There is ample evidence to suggest that the integration of neuronal assemblies involved in top-down processing is dependent upon extremely fast and synchronous activation, both in cortico-cortical networks and in cortico-thalamic networks (e.g. Engel *et al.* [2001]; Varela *et al.* [2001]; Friston and Stephan [2007]). Engel *et al.* [2001] go as far as to suggest that 'top-down factors can lead to states of "expectancy" or "anticipation" that can be expressed in the temporal structure of activity patterns before the appearance of stimuli' [2001: 710]. Other studies suggest that not only changes in discharge rate of neurons or neuronal ensembles, but also changes in neuronal synchrony, can be predictive in nature (e.g. Riehle *et al.* [2000]). This might be so if the brain utilizes a so-called temporal binding mechanism through which large-scale neuronal assemblies coordinate their activity in synchrony, as suggested by Engel *et al.* [2001].

In short, the dynamics (neuronal and extraneuronal) inherent in the minimization of free energy introduce a temporal dimension into realization relations that—in the terms of dynamical systems theory—is understood as reciprocal, circular, causation. Yet in the context of realization relationships, this may be better explained, as well as understood, in terms of diachronic realization—a conception of realization that takes into account the coupling between different temporal scales.

5. Conclusion

In this paper, I have argued that the free energy principle has implications for the wide versus narrow realization distinction: if the free energy principle is correct, then internal organismic realizers are insufficient for realizing free energy minimization, because the free energy principle itself attributes minimization of free energy to an organism-environment coupling. I then argued that, if the free energy principle is correct, it has implications for the view that realization is synchronic. The reason is that free energy minimization is realized (in part) by dynamical processes that are inherently temporal in nature. If this is true, we have two coherent views of realization—synchronic realization, on the one hand, and diachronic realization, on the other—yet only one that underlies free energy minimization.

University of Wollongong

References

Aizawa, Ken and Carl Gillett 2009a. The (Multiple) Realization of Psychological and Other Properties in the Sciences, *Mind and Language* 24/2: 181–208.

Aizawa, Ken and Carl Gillett 2009b. Levels, Individual Variation and Massive Multiple Realization in Neurobiology, in *Oxford Handbook of Philosophy and Neuroscience*, ed. John Bickle, Oxford: Oxford University Press: 539–82.

Ashby, William Ross 1952. *Design for a Brain*, London: Chapman & Hall.

Bennett, Karen 2011. Construction Area (No Hard Hat Required), *Philosophical Studies* 154/1: 79–104.

Brown, Harriet, Karl Friston, and Sven Bestmann 2011. Active Inference, Attention, and Motor Preparation, *Frontiers in Psychology* 2: 1–10.

Bruineberg, Jelle and Erik Rietveld 2014. Self-Organization, Free Energy Minimization, and Optimal Grip on a Field of Affordances, *Frontiers in Human Neuroscience* 8: 1–14.

Clark, Andy 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford: Oxford University Press.

Clark, Andy 2013. Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science, *Behavioral and Brain Sciences* 36/3: 181–204.

Cosmelli, Diego and Evan Thompson 2010. Embodiment or Envatment?: Reflections on the Bodily Basis of Consciousness, in *Enaction: Towards a New Paradigm for Cognitive Science*, ed. John Stewart, Olivier Gapenne, and Ezequiel A. Di Paolo, Cambridge, MA: The MIT Press: 361–86.

Engel, Andreas 2010. Directive Minds: How Dynamics Shapes Cognition, in *Enaction: Towards a New Paradigm for Cognitive Science*, ed. John Stewart, Olivier Gapenne and Ezequiel A. Di Paolo, Cambridge, MA: The MIT Press: 219–44.

Engel, Andreas, Pascal Fries, and Wolf Singer 2001. Dynamic Predictions: Oscillations and Synchrony in Top-Down Processing, *Nature Reviews Neuroscience* 2/10: 704–16.

Friston, Karl 2002. Beyond Phrenology: What Can Neuroimaging Tell Us About Distributed Circuitry? *Annual Review of Neuroscience* 25: 221–50.

Friston, Karl 2003. Learning and Inference in the Brain, *Neural Networks* 16/9: 1325–52.

Friston, Karl 2009. The Free-Energy Principle: A Rough Guide to the Brain? *Trends in Cognitive Sciences* 13/7: 293–301.

Friston, Karl 2010. The Free-Energy Principle: A Unified Brain Theory? *Nature Reviews Neuroscience* 11/2: 127–38.

Friston, Karl 2011. Embodied Inference: Or 'I Think Therefore I Am, If I Am What I Think', in *The Implications of Embodiment (Cognition and Communication*), ed. Wolfgang Tschacher and Claudia Bergomi, Exeter: Imprint Academic: 89–125.
Friston, Karl and Klaas E. Stephan 2007. Free-Energy and the Brain, *Synthese* 159/3: 417–58.
Friston, Karl, Chris Thornton, and Andy Clark 2012. Free-energy Minimization and the Dark-Room Problem, *Frontiers in Psychology* 3/130: 1–7.
Gillett, Carl 2002. The Dimensions of Realization: A Critique of the Standard View, *Analysis* 62/4: 316–23.
Gillett, Carl 2003. The Metaphysics of Realization, Multiple Realizability, and the Special Sciences, *The Journal of Philosophy* 100/11: 591–603.
Gillett, Carl 2007. Hyper-Extending the Mind? Setting Boundaries in the Special Sciences, *Philosophical Topics* 35/1–2: 161–87.
Haken, Hermann 1983. *Synergetics: An Introduction: Non-Equilibrium Phase Transition and Self-Organization in Physics, Chemistry, and Biology*, 3rd edn, Berlin: Springer.
Hohwy, Jakob, Andreas Roepstorff, and Karl Friston 2008. Predictive Coding Explains Binocular Rivalry: An Epistemological Review, *Cognition* 108/3: 687–701.
Kemp, Thomas 1982. *Mammal-Like Reptiles and the Origin of Mammals*, New York: Academic Press.
Nakahara, Kiyoshi, Toshihiro Hayashi, Seiki Konishi, and Yasushi Miyashita 2002. Functional MRI of Macaque Monkeys Performing a Cognitive Set-Shifting Task, *Science* 295/5559: 1532–6.
Noë, Alva 2004. *Action in Perception*, Cambridge, MA: The MIT Press.
Polger, Tom 2007. Realization and the Metaphysics of Mind, *Australasian Journal of Philosophy* 85/2: 233–59.
Polger, Tom 2010. Mechanisms and Explanatory Realization Relations, *Synthese* 177/2: 193–212.
Putnam, Hilary 1960. Mind and Machines, in *Dimensions of Mind*, ed. Sidney Hook, New York: New York University Press: 138–64.
Rao, Rajesh P. N, and, Dana H. Ballard 1999. Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-Field Effects, *Nature Neuroscience* 2/1: 79–87.
Riehle, Alexa, Franck Grammont, Markus Diesmann, and Sonja Grün 2000. Dynamical Changes and Temporal Precision of Synchronized Spiking Activity in Monkey Motor Cortex During Movement Preparation, *Journal of Physiology* 94/5–6: 569–82.
Roepstorff, Andreas, Jörg Niewöhner, and Stefan Beck 2010. Enculturating Brains Through Patterned Practices, *Neural Networks* 23/8–9: 1051–9.
Roepstorff, Andreas and Chris Frith 2004. What's at the Top in the Top-Down Control of Action? Script-Sharing and 'Top-Top' Control of Action in Cognitive Experiments, *Psychological Research* 68/2–3: 189–98.
Ross, Don, and James Ladyman 2010. The Alleged Coupling-Constitution Fallacy and the Mature Sciences, in *The Extended Mind*, ed. Richard Menary, Cambridge, MA: The MIT Press: 155–66.
Shapiro, Lawrence 2004. *The Mind Incarnate*. Cambridge, MA: The MIT Press.
Shoemaker, Sydney 1981. Some Varieties of Functionalism, *Philosophical Topics* 12/1: 93–119.
Varela, Francisco, Jean-Philippe Lachaux, Eugenio Rodriguez, and Jacque Martinerie 2001. The Brainweb: Phase Synchronization and Large-Scale Integration. *Nature Reviews Neuroscience* 2/4: 229–39.
Wilson, Robert A 2001. Two Views of Realization, *Philosophical Studies* 104/1: 1–31.
Wilson, Robert A 2004a. *Boundaries of the Mind: The Individual in the Fragile Sciences*, Cambridge: Cambridge University Press.
Wilson, Robert A 2004b. Realization: Metaphysics, Mind, and Science, *Philosophy of Science* 71/5: 985–96.
Wilson, Robert A, and Andy Clark 2009. How to Situate Cognition: Letting Nature Take its Course, in *The Cambridge Handbook of Situated Cognition*, ed. Philip Robbins and Murat Aydede, Cambridge: Cambridge University Press: 55–77.