Modularity and the Predictive Mind

Zoe Drayson

Modular approaches to the architecture of the mind claim that some mental mechanisms, such as sensory input processes, operate in special-purpose sub-systems that are functionally independent from the rest of the mind. This assumption of modularity seems to be in tension with recent claims that the mind has a predictive architecture. Predictive approaches propose that both sensory processing and higher-level processing are part of the same Bayesian information-processing hierarchy, with no clear boundary between perception and cognition. Furthermore, it is not clear how any part of the predictive architecture could be functionally independent, given that each level of the hierarchy is influenced by the level above. Both the assumption of continuity across the predictive architecture and the seeming non-isolability of parts of the predictive architecture seem to be at odds with the modular approach. I explore and ultimately reject the predictive approach's apparent commitments to continuity and non-isolation. I argue that predictive architectures can be modular architectures, and that we should in fact expect predictive architectures to exhibit some form of modularity.

1   Introduction

The aim of this paper is to explore the relationship between two approaches to the architecture of the mind: modular architectures, in which at least some mental mechanisms operate in functionally-isolated special-purpose modules; and predictive architectures, in which mental mechanisms are integrated into a hierarchy of Bayesian computational processes. The two approaches are often discussed as though they are in tension with each other, but I aim to show that predictive approaches are consistent with modular approaches, and that predictive architectures may in fact entail some form of modularity.

Claims about the modularity of mind take various forms, but all proponents of modular architectures agree that there are special-purpose sensory subsystems which are functionally isolated from other mental processes. The more modest versions of modularity propose that only this subset of mental processes are modular, while amodal reasoning and thought are non-modular. Proponents of 'massive modularity', on the other hand, propose that all mental processes are modular. In this paper, I'll be focusing on a commitment that is shared by proponents of both modest and massive modularity: the claim that there is some stage of sensory processing which is isolated from amodal thought processes. This seems to be the feature of modular architectures that is most in tension with the predictive approach to the mind.

Predictive approaches to the mind propose that our mental architecture should be understood as a hierarchy of Bayesian computational mechanisms. Cognitive information higher in the hierarchy is used to generate predictions about lower-level sensory information, and information about prediction

errors is passed back up the hierarchy. Proponents of the predictive approach, most notably Clark and Hohwy, claim that there is no obvious boundary or distinction to be drawn between the kinds of processing found at different levels of the hierarchy, which seems to be at odds with the idea of modular decomposition. Additionally, the Bayesian priors for each level in the hierarchy are provided by the level above, which makes it difficult to see how a lower-level sensory process could be informationally isolated from higher-level cognition. These features of the predictive approach are thus *prima facie* in tension with the modular approach to the architecture of the mind.

In this paper, I explore the apparent tensions between the predictive and modular approaches to the mind, and suggest that they can be reconciled. I start by introducing modular architectures (Section 2) and predictive architectures (Section 3). I outline the predictive approach's commitments to the continuity of perception and cognition, and to the non-isolation of sensory processing from higher level informational influence (Section 4). I then challenge these commitments (Section 5). I argue that continuity at one level of a mental architecture is compatible with discontinuity at a different level, and that continuity alone cannot provide an argument against modularity. I also argue that the predictive approach's commitment to top-down influences on lower-level processes does not entail the lack of functional isolability that would be inconsistent with modularity. Furthermore, I show that there are ways of understanding modularity on which predictive approaches seem to require a sort of modularity.

2   The Modular Mind

To claim that the mind is modular is to posit a functional distinction between different kinds of mental processes. The modular approach to the mind claims that certain of our mental processes are functionally distinct, in the sense of being independent or isolable from other mental processes. Sensory input processes are often cited as an example of modules in virtue of exhibiting the relevant sort of functional independence: if the data on one's relevant sensory receptors are processed without drawing on information elsewhere in the system, such as data at one's other sensory receptors or one's beliefs more generally, then that sensory process is functionally independent from those other mental processes. To the extent that a sensory process relies only on its own proprietary store of information, it operates in functional isolation from other mental processes. There is much debate over precisely how to specify the sort of independence and isolation that matter for modularity, some of which I'll touch on later in this section. It is important to notice, however, that these notions of independence and isolation associated with modular architectures are functional rather than structural. In particular, modular approaches to the mind are not claims about neural localization: whether a functionally-isolated information process is also neurally isolated is a matter for empirical investigation.

Modular architectures come in different strengths. 'Modestly modular' views, such as those of Fodor (Fodor 1983), propose that modularity is largely a feature of sensory input processes (e.g. vision, audition), and sometimes extend this claim to include mental processes such as language comprehension. Each of these modules, however, is assumed to output into the same central cognitive system: a set of non-modular processes which combine the products of the modules with stored beliefs and memories, and which is responsible for amodal thought processes such as general reasoning, belief-revision, planning, and decision-making.

'Massively modular' views, such as those of Peter Carruthers (Carruthers 2006) reject the idea of a central cognitive system in favour of further modular processing. Proponents of massive modularity tend to deny, therefore, that we have general-purpose capacity for reasoning. Instead, they propose that there are distinct modules for the kinds of reasoning related to distinct domains: the reasoning processes that allow us to detect unfair social behavior, for example, might be functionally independent from the reasoning processes that we go through in assessing potential mates.

Whether the architecture in question is modestly or massively modular, the most characteristic feature of modular processes is their functional independence or isolation from other information processes. Beyond that, there is much disagreement over the exact definition of modularity which need not concern us here. And even specifying the precise kind of functional independence exhibited by modules is difficult. One way to characterize it is in terms of information encapsulation, whereby a process is informationally encapsulated to the extent that it lacks access to information stored elsewhere in the architecture (Fodor 1983). Information encapsulation seems to explain why a mental process can be tractable and fast, in virtue of not being able to take the mind's vast amounts of information into consideration. But to use this as the defining feature of a module is somewhat restrictive: we might have reasons to think that a process is functionally independent in some interesting respect even where it does not meet the condition of being informationally encapsulated. As an example, consider sensory input processes such as vision. Visual processes have long been considered to be informationally encapsulated, on the assumption that each different feature of the environment (shape, colour, etc.) is extracted by a dedicated and functionally independent process. Empirical evidence, however, shows interactions within and across sensory modalities to an extent that challenge their claim that informational encapsulation. But visual processes still seem to exhibit degrees of functional independence, in the sense that they are functionally detachable or seperable from other mental processes. Daniel Burnston and Jonathan Cohen, for example, argue that visual processes still exhibit distinct perceptual strategies: while these strategies interact with each other, each perceptual strategy is sensitive to a restricted range of informational parameters. This sensitivity means that each process interfaces with other mental processes in a limited number of ways. This, they suggest, gives us a different way to characterize the sort of functional independence involved in modular processes:

> [W]hat makes modular processes modular, detachable, and in some sense separable from the rest of mentation is that they interface with other aspects of mental processing in a circumscribed number of ways. That is, modular processes are modular just because, and in so far as, there is a delimited range of parameters to which their processing is sensitive. (Burnston and Cohen 2015, p. 132)

In what follows, I won't assume that modular processes need to be informationally encapsulated by definition. I will assume that there needs to be some form of functional independence and isolation, and I'll come back to Burnston and Cohen's characterization of modularity later in the paper.

3   The Predictive Mind

The predictive approach to the architecture of the mind goes by a number of names, such as 'the hierarchical predictive processing perspective' (Clark 2013) and the 'the prediction error minimization framework' (Hohwy 2013). There are two fundamental ideas on which it rests: a hierarchical architecture of Bayesian computational processes, and a data-compression strategy called 'predictive coding'. I'll outline each out of these commitments in turn.

Proponents of predictive mental architectures propose that the brain implements Bayesian computational processes: it generates hypotheses or expectations about the world and updates these in light of new evidence in accordance with Bayes' Rule. Bayes' Rule states that the probability of a hypothesis, given the evidence, is updated by considering the product of the likelihood (the probability of the evidence given the hypothesis) and the prior probability of the hypothesis. In the case of visual perception, for example, the claim is that the visual system generates hypotheses about the external world on the basis of its existing information and estimates their prior probability: the probability of their being true before the sensory data has been taken into account. If I walk into my office, for example, the hypothesis that the furniture is located where I left it will generally have a higher prior probability than the hypotheses that the furniture is on the ceiling. The system also has existing information about

what sort of sensory input is caused by certain objects, e.g. what sort of retinal data are associated with viewing certain items of furniture from certain angles. This is what enables the system to calculate the likelihood of the hypothesis once the actual sensory data are known: the probability of that particular sensory input, given the hypothesis in question. The product of the hypothesis' likelihood and its prior probability is what generates the posterior probability of the hypothesis. According to the predictive approach, the hypothesis with the highest posterior probability will determine what I perceive.

This example is oversimplified. In fact, the predictive approach posits many hypothesis-testing computational processes, arranged hierarchically. The hypotheses and predictions at lower levels of the hierarchy tend to be spatially and temporally precise, while those at higher levels are more abstract. The higher-level hypotheses act as Bayesian priors for lower-level processes: each level tries to predict the input to the level below, and then updates the hypothesis accordingly. The predictive approach adopts an empirical Bayes method, on which the priors are estimated from the data rather than fixed pre-observationally, so they are shaped over time from the sensory data. This allows the predictive approach to account for priors without circularity, because the hypotheses determining perceptual experience are not themselves based directly on perceptual experiences but extracted indirectly from higher-level hypotheses. For more on empirical Bayes and the predictive approach, see (Hohwy 2013, p. 33) and (Clark 2013, p. 185).

Predictive mental architectures combine this hierarchy of Bayesian processing with the second key element of the approach: the computational framework of predictive coding. Predictive coding is a way of maximizing the efficiency of an information system, by ensuring that it doesn't process any more information than it needs to. Predictive coding was developed as a data compression strategy by computer scientists to allow more efficient storage and transmission of large files, and works by using the information system's existing information to predict its own expected inputs, so that the system only needs to account for deviations from its predictions. This means that more of its resources can be allocated to novel information. Neuroscientists have suggested that the brain might use a similar strategy to deal with the massive amounts of sensory information it receives: if it can correctly predict at least some of the information received by the sense organs, then it can direct its resources to accounting for novel or unexpected sensory information. (Further details of predictive coding as a neurocomputational strategy can be found in Friston and Stephan 2007.)

Proponents of predictive mental architectures combine the Bayesian computational hierarchy with the predictive coding strategy to argue that the brain doesn't first process all the sensory input available, then construct hypotheses about the probable worldly causes of the sensory input. Instead, hypotheses generated at higher levels of the hierarchy are used to predict the inputs received from the level below. At the lowest level of the hierarchy, the hypotheses are used to predict the data on the sense receptors. If a prediction is correct, there is no need to update the corresponding hypothesis: the input is explained away by it. The corresponding hypothesis only needs to be updated if its prediction is incorrect, that is, if there are discrepancies between the generated prediction and the actual input at a level. In this way, the predictive architecture only has to account for unexpected or novel inputs.

4   The Apparent Tension between Modular and Predictive Architectures

There are two features of predictive architectures of the mind that seem to be directly in tension with modular architectures of the mind. First, the predictive approach is often described as lacking any obvious boundary between cognitive processes and perceptual processes in the predictive hierarchy, and therefore as committed to the continuity of cognition and perception. Second, information-processing at each level of the predictive hierarchy incorporates information from the level above, which suggests that no level of processing is isolated from the information elsewhere in the system. In this section, I'll explore these features in greater detail, and show where the *prima facie* tension is supposed to arise.

4.1 The Continuity Claim

The continuity claim is the claim that there is no clear divide between cognitive processes and non-cognitive (particularly perceptual) processes in the Bayesian hierarchy. The claim is that the entire predictive hierarchy uses the same Bayesian computational processes, so there is no distinction to be made between the kinds of processes that underlie cognition, on the one hand, and the kinds of processes that underlie perception, on the other hand.

Clark makes the continuity claim when he proposes that predictive architectures depict perception and cognition as "profoundly unified and, in important respects, continuous" (Clark 2013, p. 187). Predictive architectures, he claims, "appear to dissolve [...] the superficially clean distinction between perception and knowledge/belief [...] we discover no stable or well-specified interface or interfaces between cognition and perception" (Clark 2013, p. 190). Predictive architecture "makes the lines between perception and cognition fuzzy, perhaps even vanishing" (Clark 2013, p. 190).

Hohwy makes the continuity claim when he observes that the predictive approach to mental architecture seems to deny our standard distinction between cognition (understood as involving conceptual thoughts such as beliefs) and perception (understood as involving the processing of sensory experience):

> It [the predictive framework] seems to incorporate concepts and thinking under a broader perceptual inference scheme. [...] On this view, concepts and beliefs are fundamentally the same as percepts and experiences, namely expectations (Hohwy 2013, p. 73).

Proponents of predictive architectures thus suggest that our standard tendency to distinguish between believing and perceiving is not supported by their account of mental architecture. They generally allow that processes lower in the hierarchy are doing something akin to sensory processing, in the sense that their predictions are often spatially and temporally precise, and may even refer to these as perceptual processes. They likewise allow that the predictions of higher-level processes are increasingly abstract and more akin to amodal reasoning, and may even refer to these processes as cognitive. But the core of the continuity claim is that there are many levels in the middle of the hierarchy that are not recognizable as standard cases of either perception or cognition: there is no point in the hierarchy at which the processes stop being perceptual and start being cognitive (see also Vetter and Newen 2014).

There is a *prima facie* tension between the continuity claim associated with predictive architectures and the commitments of modularity. Modular architectures seem to respect our standard distinction between perceiving and believing: perception and cognition are assumed to be different kinds of processes (often relying on different forms of representation) with a clear boundary between them. In particular, at least some part of perceptual processing is assumed to be functionally isolated and independent from cognitive processing.

4.2 The Non-Isolation Claim

The second relevant commitment of the predictive approach is the non-isolation claim: the claim that there is no part of perceptual processing that is informationally isolated from higher-level cognitive processing. This commitment seems to arise from the hierarchical structure of the Bayesian computational mechanisms, in which the priors at each level in the hierarchy are provided by the level above. This suggests that there is top-down influence at every level. (Recall that the lower down the hierarchy, the more spatiotemporally precise the predictions are; the higher up the hierarchy, the more abstract the predictions are.)

Proponents of predictive architectures seem to assume that this role of the Bayesian priors within the hierarchy entails that no part of the hierarchy is isolated from higher-level information. Hohwy, for example, claims that there is "no theoretical or anatomical border preventing top-down projec-

tions from high to low levels of the perceptual hierarchy" (Hohwy 2013, p. 122). Clark emphasizes that the sorts of abstract predictions made at higher levels can influence the more spatiotemporally precise predictions at lower levels: he claims that perception is "theory-laden" and "knowledge-driven" on the predictive approach, and that "[t]o perceive the world just is to use what you know to explain away the sensory signal" (Clark 2013, p. 190).

In philosophy, perception that is subject to top-down effects is often characterized as 'cognitively penetrable'. Predictive architectures are sometimes described in these terms: Hohwy, for example, suggests that predictive architectures "must induce penetrability of some kind" (Hohwy 2013, p. 120); while Gary Lupyan claims that "[p]redictive systems are penetrable systems" (Lupyan 2015, p. 547) and that we should expect penetrability whenever we have higher-level information processing making predictions about lower-level information processing; and Petra Vetter and Albert Newen claim that the top-down influences on perception suggest that "[p]redictive coding can thus be regarded as an extreme form of cognitive penetration" (Vetter and Newen 2014, p. 72).

If predictive architectures are committed to the cognitive penetration of perception, this seems to put them in tension with modular architectures. Proponents of modular architectures claim that at least some informational processing is isolated from other informational processing. In particular, the relevant claim here is that sensory processing (at least in its early stages) is not influenced by cognitive information in the form of beliefs or memories. Modularity thus provides a mechanism for avoiding cognitive penetration. If proponents of predictive architectures are right that higher-level processes are able to influence lower-level processes in the relevant way, then this suggests that predictive architectures are not modular architectures.

5 Predictive Architectures as Modular Architectures

When proponents of predictive architectures propose versions of the continuity claim and the non-isolation claim, they don't always draw an explicit contrast between their approach and modular approaches. But it is often assumed that predictive architectures are non-modular architectures: Jona Vance and Dustin Stokes, for example, describe the predictive approach as involved in "the development of non-modular mental architectures" (Vance and Stokes forthcoming). I want to demonstrate that neither the continuity claim nor the non-isolation claim entail that predictive architectures are non-modular.

First, I'll argue that the continuity claim does not show that there is no distinction between perception and cognition. Then I'll argue against the non-isolation claim, and demonstrate that there is a kind of isolation on the predictive approach that is consistent with some form of modularity.

5.1 Challenging the Continuity Claim

As already demonstrated, proponents of predictive architectures argue for the continuity claim on the grounds that the same kinds of Bayesian computational processes are in play throughout the processing hierarchy. This means that at each level in the hierarchy, the same methods of hypothesis testing are used, whether the hypotheses in question are at the spatiotemporally precise or more abstract end of the scale.

This fact in itself, however, does not distinguish predictive architectures from modular architectures. Proponents of modularity can allow that perceptual processes and cognitive processes use the same kind of computational mechanisms: Fodor, for example, thinks that both perceptual and cognitive processes use classical computational inference involving rules and representations (Fodor 1983). Whether perceptual processes use precisely the same kind of format as cognitive processes, e.g. the same 'language of thought', is an empirical matter and thus isn't ruled out simply by adopting a modular architecture (Aydede 2015). When modular approaches distinguish between cognitive processes and perceptual processes, therefore, this distinction can't rely on the claim that the relevant processes

use different computational mechanisms. Since it is consistent with modularity to claim that perception and cognition use the same kinds of computational processes, the continuity of the Bayesian hierarchy on the predictive approach does not demonstrate that predictive architectures are non-modular.

It is true that modular architectures generally distinguish between higher-level cognitive processing and a certain stage of perceptual processing. But this distinction need not be found in the finer-grained details of the computational processes. By way of example, consider the difference between classical computational theories which posit a syntactic process of symbol manipulation, and connectionist theories which posit processes of activation through a network, mediated by connection weights. Connectionist networks may look non-modular in the sense that there are no obvious boundaries between one part of the network and any other part. But a system that is non-modular at one level of description can still be modular at a different level, as Martin Davies has demonstrated: once we observe the connectionist network performing a task, discontinuities can appear and functions can become dissociated. Davies suggests that we employ the idea of coarser and finer grains of modularity, to respect the fact that "[w]hat we really have is a hierarchy of levels of coarser and more detailed interpreted descriptions of the way in which the task is carried out" (Davies 1989, p. 547).

The moral here is that modularity is a matter of grain: a computational system can be modular when viewed at one level of abstraction but not when viewed at another: continuity in the fine-grained details of the information-processing is compatible with discontinuity at a coarser-grained perspective. When proponents of predictive architectures make the continuity claim, therefore, this does not rule out that the system is a modular one. And as long as modularity is not ruled out, it is possible that perceptual processes are distinct from cognitive processes at some appropriately coarse-grained level of description.

Furthermore, even if the predictive approach could demonstrate that there is no clear boundary between perceptual processes and cognitive processes, this would not entail that there is no difference between the two. The lack of clear boundary might suggest that the distinction between perception and cognitive is non-exclusive, for example, such that processes in the middle of the hierarchy are best classified as both perceptual and cognitive; or it might suggest that the distinction between perception and cognition is not exhaustive, such that processes in the middle of the hierarchy are neither perceptual nor cognitive. Both of these approaches are consistent with the claim that there are clear cases of perceptual processing which is not cognitive at the lower end of the hierarchy, and clear cases of cognitive processing which is not perceptual at the higher end of the hierarchy.

5.2   Challenging the Non-Isolation Claim

The non-isolation claim is the claim that there is no part of perceptual processing that is informationally isolated from cognitive processing. As already demonstrated, proponents of predictive architectures argue for the non-isolation claim on the grounds that each level in the predictive Bayesian hierarchy has its priors provided by the level above, suggesting that lower-level processes can't be wholly isolated from the influence of higher-level processes. This aspect of predictive archictectures has been interpreted by some as entailing the cognitive penetration of perception.

Care is needed when talking about cognitive penetrability, because some people use the term 'cognitive penetration' to refer to all top-down influences on perception – regardless of which stages of perceptual processes are influenced, which kinds of cognitive processes are doing the influencing, and what the type of influence is. But in philosophy in particular, the label is often reserved for a certain kind of top-down influence on perception: the phenomenon whereby particular kinds of higher-level mental states (notably beliefs and memories) exert a direct influence on the contents of conscious perceptual experience. (See Macpherson forthcoming for an overview of the different varieties of cognitive penetration.)

It is far from clear whether predictive architectures result in cognitive penetration of this sort. On the predictive approach, conscious perceptual experience is the product of the entire prediction minimization process: it is determined by the interactions between top-down and bottom-up information flow within the entire hierarchy, rather than being associated with a particular level in the Bayesian hierarchy (cf. Clark 2013, p. 185). While this might suggest that there generally will be top-down effects on perceptual experience, it doesn't entail that these top-down influences will be of the right kind to constitute cognitive penetration. As an epistemologically interesting phenomenon, cognitive penetration requires that the cognitive states involved are traditionally doxastic: they are the beliefs of the person rather than merely information represented in the cognitive system. (For further elaboration on the personal/subpersonal and doxastic/subdoxastic distinctions, see Drayson 2012, Drayson 2014.) Proponents of the predictive approach can interpret the architecture as including these doxastic states, but they also have the option of adopting an eliminativist take on beliefs so construed (cf. Dewhurst 2017). There is also a question of whether the influence from top-down processing associated with predictive architectures is appropriately direct. As a result, the predictive approach does not necessarily entail cognitive penetration. In a longer discussion of some of these issues, Fiona Macpherson reaches the similar conclusion that "mere acceptance of the predictive coding approach to perception does not determine whether one should think that cognitive penetration exists" (Macpherson forthcoming, p. 10).

If we leave talk of cognitive penetration out of the picture, however, there is still the question of top-down influence more generally on perceptual processing. The non-isolation claim merely proposes that no part of perceptual processing is isolated from cognitive processing. If we allow that the labels 'perceptual' and 'cognitive' refer respectively to the most spatiotemporally precise predictive processes and the most abstract predictive processes, as discussed with relation to the continuity claim, then predictive architecures look like clear instances of non-isolation. Predictive architectures are committed to the claim that slightly more abstract and less spatiotemporally precise hypotheses act as the priors for slightly less abstract and more spatiotemporally precise hypotheses. Since this is the case at every level in the hierarchy, then doesn't it follow that cognitive processes influence perceptual processes?

There are, I suggest, ways to avoid this conclusion. Notice that this way of reasoning seems to assume that the 'influencing' relation between two levels has the logical property of transitivity: if Level *A + 1* influences Level *A*, and Level *A* influences Level *A − 1*, then Level *A + 1* influences Level *A − 1*. When Hohwy suggests that the top-down processing associated with predictive architectures "is not a free-for-all situation" (Hohwy 2013, p. 155), his argument seems to involve rejecting the transitivity of the 'influencing' relation between two levels. He claims that each pair of levels form a "functional unit", with the higher level passing down predictions and the lower level passing up prediction errors (Hohwy 2013, p. 153), and that each pair of levels is evidentially insulated:

> In this sense the upper level in each pair of levels only 'knows' its own expectations and is told how these expectations are wrong, and is never told directly what the level below 'knows'. […] For this reason, the right kind of horizontal evidential insulation comes naturally with the hierarchy. (Hohwy 2013, p. 153)

Hohwy argues that this evidential insulation prevents high-level processes from influencing low-level processes, except where there is uncertainty and noisy input. He seems to be suggesting that we shouldn't expect the 'influence' relation between any two levels to exhibit transitivity, because the relation should be understood in terms of one level 'providing evidence for' or 'justifying' another. Notice that this relies on a strongly epistemic reading of the Bayesian hierarchy in terms of knowledge and evidence, which Hohwy unpacks in terms of the confidence of a system in its judgments. Ultimately this comes down to the precision of the system's predictions, and its expectations about how good its perceptual inferences are in particular situations. If lower-level predictions are highly precise,

then it is more difficult for them to be influenced by higher-level predictions. (For an argument against Hohwy's approach, see Vance and Stokes forthcoming.)

I propose an alternative (not necessarily incompatible) way to explain why we shouldn't expect the 'influencing' relation between levels to be transitive: I suggest that the relation between any two levels is one of probabilistic causal influence. Bayesian networks are simply maps of probabilistic dependence, and probabilistic dependence is transitive: if Level $A - 1$ probabilistically depends on Level A, and Level A probabilistically depends on Level $A + 1$, then Level $A - 1$ probabilistically depends on Level $A + 1$ (Korb et al. 2009). But predictive architectures use Bayesian computation: mechanisms that implement *causal* Bayesian networks. In causal Bayesian networks, the probabilistic dependences between variables are the result of causal processes between those variables. Metereological models used to forecast the weather provide a good example of problems that probabilistic causal models raise for the transitivity of the causal relation between two events:

> For instance, given our very coarse and only probabilistic metereological models, each day's weather may be granted to causally influence the next day's weather. But does the weather, say, at the turn of last century still influence today's weather? It does not seem so; somewhere in between the influence has faded completely, even though it may be difficult to tell precisely when or where. (Spohn 2009, p. 59).

Proponents of probabilistic causal models tend to argue that the causal influence between states of a network is weak, and that such weak causal influences will not be preserved over long causal chains (Spohn 2009). As a result, probabilistic causation is widely acknowledged to result in failures of transitivity of the causal relation (Suppes 1970). If we apply this thinking to predictive architectures, we can explain why, when it is true that Level $A + 1$ causally influences Level A, and Level A causally influences Level $A - 1$, we need not expect it to be true that Level $A + 1$ causally influences Level $A - 1$. The further apart the levels in the hierarchy are, the less likely there is to be causal influence from the higher level to the lower level. In this way, we can accept that each level in the predictive hierarchy is causally influenced by (i.e. gets its priors from) the level above, without having to accept that each level in the hierarchy causally influences all the levels below it, or that each level is causally influenced by all the levels above it. And so it remains plausible that there are perceptual processes (lower-level processes involved in spatiotemporally precise predictions) which are isolated from cognitive processes (higher-level processes involved in abstract predictions) in the sense that the former are not causally influenced by the latter. The non-isolation claim associated with the predictive approach seems to rely on the assumption that the relation of influence between any two levels in the predictive hierarchy is a transitive one. I have argue that the relation of causal dependence between two events fails to be a transitive relation when the causal dependence in question is probabilistic. And since Bayesian computational mechanisms are probabilistic causal networks, we shold not expect the relation of causal influence between levels to be transitive. As a result, the non-isolation claim is not entailed by predictive mental architectures.

5.3   Modularity or Something Like It

I have argued that, despite the appearance of tension between predictive architectures and modular architectures, adopting the predictive approach does not entail rejecting modularity. Claims about the continuity of perception and cognition seem to be compatible with modularity, and the commitment to top-down information processing doesn't entail that no part of perceptual processing is isolated from cognition.

I want to go further, and suggest that predictive architectures themselves possess a kind of modularity. Against the non-isolation claim, I argued that the causal influence of top-down information

processing is unlikely to penetrate from the higher levels of the Bayesian hierarchy to the lower levels, due to the causal intransitivity of probabilistic causal networks such as those found in a Bayesian computational hierarchy (Suppes 1970, Korb et al. 2009). Just as the causal influence of today's weather on future weather diminishes the further into the future we consider, so the causal influence of higher-level processing in the predictive hierarchy diminishes the further down the hierarchy we look: in probabilistic causal models, causal influence is not preserved over long causal chains (Spohn 2009).

As a result, we can acknowledge that a low-level process is influenced by the levels immediately above, but deny that much higher-level processes have any causal influence on it. Such a low-level process would function independently and in isolation from those high-level processes. In other words, it would posess the sort of functional features that we associate with modular architectures. The top-down effects from the level immediately above would presumably count as 'within-module' effects, and this might extend to further levels above. The important point is that not every higher level would necessarily exert a causal influence merely in virtue of being a higher level, because transitivity cannot be expected in causal probabilistic networks.

The kind of modularity involved, however, looks somewhat different from traditional approaches to the modularity of mind. It is not clear, for example, that there is genuine information encapsulation going on here: while it is very unlikely that high-level processes could extend their causal influence all the way down to low-level processes, it is still possible. On traditional approaches to modularity, encapsulated processes are generally portrayed as informationally isolated in principle, rather than merely in practice. A related concern is that, on the view I'm suggested, the boundaries of modules wouldn't be clearly defined – or at least wouldn't retain fixed boundaries over time. It even seems possible for the boundaries of modules to overlap on this view, which is not a feature of traditional modularity.

I propose that these worries should not lead us to think of the architectures in question is non-modular. Some of these features are simply the result of using Bayesian computation: probabilistic mechanisms won't yield the same clean distinctions as classically computational mechanisms. This should prompt us to explore the nature of modularity in probabilistic systems, rather than to reject the useful notion of modules as inapplicable to such architectures. At least some of the modular aspects of predictive architectures, I suggest, can be understood as informationally encapsulated in the Fodorian sense: the processes in question can access less than all of the information available to the organism as a whole (Fodor 1983). But even if it could be argued that there is no information encapsulation in predictive architectures, we can retain the claim that predictive architectures are modular architectures by returning to an alternative characterization of modules introduced earlier. Recall Burnston and Cohen's proposal that a process is modular to the extent that there is a delimited range of parameters to which its processing is sensitive; i.e. insofar as there is a circumscribed number of ways that the process interfaces with other aspects of mental processing. The processes at each level of the Bayesian hierarchy are highly circumscribed in the sense that they are generating hypotheses at different spatiotemporal grains: more precise hypotheses towards the lower levels, and more abstract hypotheses towards the higher levels. Each process is sensitive to a limited range of parameters: generally those of the processes in the levels immediately above and below. And notice that Burnston and Cohen's approach to modularity actually predicts the existence of overlapping modules. They argue that if we individuate modules by the delimited range of parameters to which their processing is sensitive, then it is likely that modular systems will significantly overlap in their associated ranges of parameters (Burnston and Cohen 2015).

My conclusion, that predictive architectures are modular architectures, is similar to Hohwy's conclusion that predictive architecture is "a kind of partially segregated architecture" (Hohwy 2013, 152). (Note that Hohwy's partially segregated architecture posits horizontal evidential insulation in addition to the vertical evidential insulation that is relevant to cognitive penetration. For further details, see Hohwy 2013, pp. 152-155.) Hohwy's argument against widespread cognitive penetration also draws

on the difference in spatiotemporal grain between hypotheses at different level: he claims that higher levels can't influence much lower levels because of the difference in abstractness of their predictions (Hohwy 2013). But it is unclear to me how Hohwy can use this difference to argue against top-down influences on sensory processing without first explaining why we should not expect the relation between levels in the predictive hierarchy to be transitive. His argument for evidential insulation between levels relies on a strong epistemological interpretation of Bayesian processing, as previously discussed, and a courtroom metaphor. My argument can be read as a way of taking Hohwy's argument (from the differences in spatiotemporal grain to the unlikelihood of cognitive penetration) and fleshing it out with the addition of two further claims: first, that causal influence in probabilistic causal networks is intransitive; and second, that the circumscribed nature of the hierarchical processes allows us to individuate them as modules.

6 Conclusion

Proponents of predictive architectures often emphasise the way in which their approach constitutes "a genuine departure from many of our previous ways of thinking about perception, cognition, and the human cognitive architecture" (Clark 2013, p. 187). In particular, they tend to highlight the continuity and integration of the Bayesian computational hierarchy, suggesting that it is at odds with the sorts of functionally decomposed or compartmentalized architectures that are associated with modular approaches to the mind.

I have argued that whatever continuity and integration we find in predictive architectures is consistent with modularity. There is no reason to think that the continuity of processing in the predictive hierarchy forces us to deny the distinction between perception and cognition, for example. And while the Bayesian computational processes get their priors from the level above, the limited reach of causal influences in Bayesian mechanisms prevents this from resulting in problematic cognitive penetration.

Furthermore, I have argued that the causal probabilistic networks employed by predictive approaches will actually result in the existence of sensory processes that are functionally isolated from high-level cognition. The causal intransitivity of the probabilistic computational mechanisms ensures that the top-down causal influence diminishes at each level, and suggests that there will be low-level processes which operate entirely independently of the influence from high-level processing. In other words, I have suggested that the predictive approach to the mind, in virtue of its Bayesian computational processes, leads us to expect a modular architecture – albeit perhaps a non-traditional version of modularity.

That predictive architectures are modular architectures should perhaps come as no surprise, if we reflect on the general motivations for modularity. An entirely integrated non-modular computational system, in which every process has access to all the information in the system, would be inefficient, slow, and potentially intractable. Modularity provides a way to make computational processes more efficient and fast by restricting the information to which certain processes have access. Predictive architectures operate in exactly the same way: each level of the hierarchy is restricted in the information it accesses: it uses information from the level above to predict its inputs from the level below. One way to interpret this fact is to claim that predictive architectures are rivals to modular architectures, using different processes to achieve similar results. I would suggest, to the contrary, that predictive and non-predictive architectures achieve similar results by organizing their information in a modular way. This is not to deny that there are differences between kinds of computational architectures. But modularity is a higher-level feature that can be shared by distinct computational architectures, and which allows us to categorize systems by the way they organize their information.

The focus on Bayesian computational hierarchies and predictive coding makes for a new and interesting approach to mental architecture, which is ripe for exploration and development. And while there are aspects of the continuity and integration of these architectures that deserve to be empha-

sized, we must not lose sight of the fact that they ultimately have to explain a wide array of very different skills and capacities. The appeal of the predictive approach is not that it shows our minds to be continuous and integrated, but that it shows how a set of continuous and integrated computational processes can be organized in such a way as to give rise to distinct mental capacities. It does so, I suggest, by organizing its information processes in a modular way. Moreover, it provides us with a new way to understand modularity as a flexible and dynamic feature of architectures, and to appreciate that predictive architectures are modular architecture.

References

Aydede, M. (2015). The language of thought hypothesis. In E. N. Zalta (Ed.) *The Stanford encyclopedia of philosophy* Metaphysics Research Lab, Stanford University.

Burnston, D. C. & Cohen, J. (2015). Perceptual integration, modularity, and cognitive penetration. In A. Raftopoulos & J. Zeimbekis (Eds.) *Cognitive influences on perception: Implications for philosophy of mind, epistemology, and philosophy of action.* Oxford University Press.

Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought.* Oxford: Clarendon Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36* (3), 181–204.

Davies, M. (1989). Connectionism, modularity and tacit knowledge. *British Journal for the Philosophy of Science, 40* (December), 541–55. Taylor & Francis.

Dewhurst, J. (2017). Folk psychology and the Bayesian brain. In T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing.* Frankfurt am Main: MIND Group.

Drayson, Z. (2012). The uses and abuses of the personal/subpersonal distinction. *Philosophical Perspectives, 26* (1), 1–18.

———(2014). The personal/subpersonal distinction. *Philosophy Compass, 9* (5), 338–346.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Friston, K. J. & Stephan, K. E. (2007). Free-energy and the brain. *Synthese, 159* (3), 417–458.

Hohwy, J. (2013). *The predictive mind.* Oxford University Press.

Korb, K. B., Hope, L. R. & Nyberg, E. P. (2009). In F. Emmert-Streib & M. Dehmer (Eds.) *Information-theoretic causal power* (pp. 231–265). Boston, MA: Springer US.

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology, 6* (4), 547–569.

Macpherson, F. (forthcoming). The relationship between cognitive penetration and predictive coding. *Consciousness and Cognition.*

Spohn, W. (2009). The difficulties with indirect causation. *Causation, Coherence and Concepts: A Collection of Essays* (pp. 57–65). Springer.

Suppes, P. (1970). *A probabilistic theory of causality.* Amsterdam, North-Holland Pub. Co.

Vance, J. & Stokes, D. (forthcoming). Noise, uncertainty, and interest: Predictive coding and cognitive penetration. *Consciousness and Cognition.*

Vetter, P. & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition, 27*, 62–75.