Why do you fear the bogeyman? An *embodied predictive coding* model of perceptual inference

Giovanni Pezzulo

Abstract Why are we scared by nonperceptual entities such as the bogeyman, and why does the bogeyman only visit us during the night? Why does hearing a window squeaking in the night suggest to us the unlikely idea of a thief or a killer? And why is this more likely to happen after watching a horror movie? To answer these and similar questions, we need to put mind and body together again and consider the embodied nature of perceptual and cognitive inference. Predictive coding provides a general framework for perceptual inference; I propose to extend it by including interoceptive and bodily information. The resulting *embodied predictive coding* inference permits one to compare alternative hypotheses (e.g., is the sound I hear generated by a thief or the wind?) using the same inferential scheme as in predictive coding, but using both sensory and interoceptive information as evidence, rather than just considering sensory events. If you hear a window squeaking in the night after watching a horror movie, you may consider plausible a very unlikely hypothesis (e.g., a thief, or even the bogeyman) because it explains both what you sense (e.g., the window squeaking in the night) and how you feel (e.g., your high heart rate). The good news is that the inference that I propose is fully rational and gives minds and bodies equal dignity. The bad news is that it also gives an embodiment to the bogeyman, and a reason to fear it.

Keywords Embodied predictive coding · Perceptual inference · Decision-making · Interoception

> "There is more wisdom in your body than in your deepest philosophy."
> –Friedrich Nietzsche

G. Pezzulo (✉)
National Research Council, Institute of Cognitive Sciences and Technologies, Via S. Martino della Battaglia, 44, 00185 Rome, Italy
e-mail: giovanni.pezzulo@istc.cnr.it

It is a windy night. You go to sleep a bit shocked because, say, you had a small car accident or just watched a shark attack horror movie. During the night, you hear a window squeaking.

In normal conditions, you would attribute this noise to the windy night. But this night, the idea that a thief or even a killer is entering your house jumps into your mind. Normally you would have immediately dismissed this hypothesis, but now it seems quite believable, despite the fact that there have been no thefts in your town in the last few years; suddenly, you find yourself expecting a thief to come out of the shadows. How is this possible?

According to the *predictive coding* theory (Clark, 2013; Friston, 2005; Rao & Ballard, 1999), the perceptual system is a hierarchical generative model that performs a Bayesian form of inference from the available sensory data (say, the sound of a window squeaking) to perceptual and cognitive hypotheses that represent the most likely causes of the data (say, the wind or a thief).

At higher levels, the competing perceptual hypotheses correspond to possible explanations of the sensory stimuli. Let's assume, for simplicity, only two mutually exclusive hypotheses: "It is the wind" and "It is a thief." Because they are mutually exclusive, the probability of the two hypotheses sums to 1 [e.g., if P(wind) = .8, then P(thief) = .2].

These hypotheses compete on the basis of how well they explain the sensory evidence, which in our example is the sound of the window squeaking. We can consider the three-level predictive coding hierarchy shown in Fig. 1. The arrows indicate that wind and thief can be considered likely causes of the window squeaking, which in turn can be considered the cause of the heard sound. Intuitively, this corresponds to one of my two hypotheses (wind vs. thief) causing the window to squeak, which in turn causes the sound I hear.

Let's assume, once again for simplicity, that I can unequivocally attribute the sound that I hear to a window squeaking, so we can simplify the problem with the two-level hierarchy

Wind vs. Thief

$\downarrow$

window squeaking

$\downarrow$

heard sound
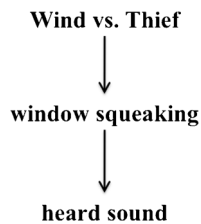
Fig. 1 A "predictive coding" hierarchy. See the main text for an explanation

shown in Fig. 2. Essentially, the predictive coding framework implements the idea of Helmholtz (1866/1962) that perception is an unconscious inference of the causes of sensation. In this framework, the arrows indicate causality; inference is done in the reverse direction, because its objective is inferring the most likely *cause* (wind vs. thief) by treating its sensed consequences (the sound of the window squeaking) as

evidence. Mechanistically, the higher-level hypotheses generate sensory predictions (say, one predicts the sensory Event A, and another predicts the sensory Event B); these predictions are propagated in a top-down way and compared against the sensory measurements (e.g., the sound of a window squeaking). A sensory prediction error is thus generated and propagated in a bottom-up way that helps revise the initial hypotheses. The hypothesis that generates less prediction error is strengthened, and its probability increases. This process is typically iterated until it settles to stable values; the iterations permit a valid inference, despite the noise in the process (e.g., the initial sensory measurements could be wrong).

If we disregard the step-by-step dynamics, predictive-coding inference selects the hypothesis (wind or thief) that better explains the sensory data (sound of window squeaking) by using Bayes's rule:

$$P\left(\text{wind}\middle|\text{evidence}\right) = \left\{ \frac{P(\text{wind})P\left(\text{evidence}\middle|\text{wind}\right)}{P(\text{wind})P\left(\text{evidence}\middle|\text{wind}\right) + P(\text{thief})P\left(\text{evidence}\middle|\text{thief}\right)} \right\}$$

The first element of Bayes's rule is $P(\text{wind} \mid \text{evidence})$. It reads "the probability of the wind hypothesis given the sensory evidence," and is the probability that we want to calculate; note that, because wind and thief are mutually exclusive, the probability of the thief hypothesis is simply one minus the probability of the wind hypothesis.

The second element of Bayes's rule is $P(\text{wind})$. It is called the "prior probability" of the wind hypothesis. Indeed, even before perceiving the sensory data, the initial hypotheses (wind vs. thief) can be assigned a priori probabilities (i.e., acquired independently of the current sensory evidence). Some examples are prior knowledge of how likely a theft is in this town, how windy the night was when I went to sleep, and so forth.

The third element of Bayes's rule is $P(\text{evidence} \mid \text{wind})$. It reads "the probability of the sensory evidence given the wind hypothesis," and is called the *likelihood* of the hypothesis. Let's think of the likelihood as a form of counterfactual reasoning that tells how likely the heard sound would be if we were to assume the wind (or the thief). In other words, the likelihood is the support that some evidence (e.g., the sound of a window squeaking) gives to a hypothesis (e.g., wind). Note

that although I only mentioned the sound of the squeaking window as evidence, several other sensory events count as evidence and should be included in the likelihood calculations. For example, my dog is not barking, and I see nothing moving in the shadows. We can now ask how likely is all of this sensory evidence, given the competing hypotheses. If we assume that a thief is in the house, a window squeaking is somewhat likely, but perhaps my dog should have barked and I should have seen a silhouette moving in the shadows or heard the sound of steps. If we assume that the wind is strong, a window squeaking is quite likely, and my dog not barking is also quite plausible. So, overall, the wind hypothesis explains quite well not only the sound of the window squeaking, but also the other evidence that I have. The thief hypothesis explains only some of them.

Overall, according to the predictive coding framework, we have rich theories of wind and thieves as *causes* of multiple sensory events. We know the prior probabilities of wind and thieves; we know how the world would look in the case of either the wind (e.g., a window squeaking) or a thief (e.g., a window squeaking and my dog barking); and we can combine this information in a statistical way to obtain a robust estimation of which hypothesis (wind vs. thief) is correct.

The last element of Bayes's rule is $P(\text{wind}) P(\text{evidence} \mid \text{wind}) + P(\text{thief}) P(\text{evidence} \mid \text{thief})$. It is a normalization factor that ensures that all of the probabilities sum to 1. This normalization factor is known as the *model evidence* and

Wind vs. Thief

$\downarrow$

sound of window squeaking
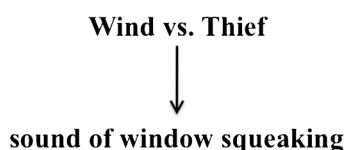
Fig. 2 Simplified predictive coding hierarchy

reports the probability of data under a particular model. Here, "models" pertain to the hypothesis space or the number of alternative explanations being entertained. For example, the model that I have been considering is that the hidden events in the world are caused either by the wind or by a thief (another model could consider three rather than two alternative hypotheses—namely, the wind, the thief, and my cat returning from a nocturnal adventure). One then adjudicates between the different models using model evidence in a hierarchical fashion; this is known as *Bayesian model comparison*, and is something that I will return to later.

A concrete quantitative example can help understand Bayes's rule at work. Normally, given the prior knowledge that the night is windy, the sound of a window squeaking would have an easy explanation. Let's assume that the prior probability of wind is .8 (you know it is windy) and the prior probability of a thief is .2 (very few thefts reported recently in your town). Let's also assume that the wind hypothesis explains the sensory events slightly better (your window squeaking, no barking, etc.) than the thief hypothesis; accordingly, we can set the likelihoods of the two hypotheses as P(sensory evidence | wind) = .6 and P(sensory evidence | thief) = .5. By applying Bayes's rule, the probability of wind (= .8276) is much higher than the probability of the thief (= .1724). From a slightly different angle, one could say that we should invoke Occam's razor, so the fact that one hypothesis (wind) explains the data so well reduces the need to invoke alternative causes.

So, you should quite unequivocally attribute the sound of window squeaking to the wind. You should stand up and close the window before it breaks, and not freeze in your bed as you are doing right now! What is wrong with this picture?

Intuitively, this has something to do with the fact that you just watched a horror movie or had a small car accident, although these events are apparently not causally related with the inference. One possibility is that your inference suffers from a *cognitive bias* from prior events; for example, if rather than a shark movie you watched a zombie movie, the idea of a thief or killer could have been somewhat "primed," and given an unlikely high prior probability. A similar priming effect could be hypothesized if you just heard of a friend robbed or of a murder in a nearby town on the TV. However, none of these events happened, and the cognitive bias effect seems unlikely. So, what is missing?

Embodied predictive coding: Using both sensory and interoceptive flows as evidence

I argue that the predictive coding account that I have described up to now misses a crucial ingredient: *interoceptive* information. I only mentioned external sensory events, such as the sound of a window or the sight of moving shadows. However,

the rich "theories" that we have about wind and thieves can also include interoceptive information, such as the fact that my heart rate would be higher if a thief was present but the wind was not.

Embodied predictive coding extends the standard predictive coding scheme by incorporating interoceptive information and the perception of the physiological condition of the body. Specifically, higher-level hypotheses would predict both sensory evidence provided by brain sensory pathways and interoceptive information linked to the autonomic system and the sympathetic and parasympathetic brain pathways; see Fig. 3.

After a horror movie or a small car accident, your body state can be altered, and interoceptive signals may report a high heart rate or sweating. If you also consider this interoceptive information in the aforementioned predictive coding scheme, then the calculations become quite different, because now the "thief" hypothesis explains more evidence (sensory plus interoceptive) than the "wind" hypothesis: It also explains why your hearth rate is high and why you are sweating; see Fig. 4.

The reader may have noted that this picture is slightly odd, because the interoceptive states (high heart rate and sweat) were caused by the horror movie or the car accident, not by the (putative) thief. Let's skip this point for the moment (see below) and consider that in the ongoing competition between "thief" and "wind," all of the available interoceptive information is considered as evidence to be explained.

So, let's redo the calculations. The prior information has not changed, but the likelihood of the events to be explained (sensory plus interoceptive) has changed drastically, because now the thief can account much better for the available evidence. Let's rename the full available evidence (sensory plus interoceptive) as E, which intuitively corresponds to "all I sense and all I feel." We can now assign $P(E \mid \text{wind}) = .3$ and $P(E \mid \text{thief}) = .8$. Your (posterior) belief in the thief hypothesis now rises to .4 (so the belief in the wind hypothesis is .6). Still, this is not sufficient to explain why the thief hypothesis sounds so convincing to you.

Let's introduce another element of the predictive coding picture that I have skipped up to now: the *uncertainty* of the data. Not all data are treated equally in the inference, because some are more uncertain than others. For example, you can be a bit uncertain of the fact that you really heard the sound of a window squeaking; maybe you had a bad dream or simply heard your neighbors' TV. So, overall, what you heard is a bit uncertain and not totally reliable. Visual evidence is even less certain, since there is no light, so the fact that you see nothing

hypotheses

↓

sensory & interoceptive evidence
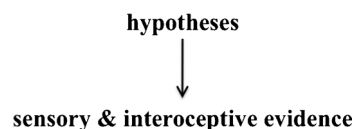
Fig. 3 Embodied predictive coding considers both sensory and interoceptive evidence

Wind vs. Thief

↓

Sound of window squeaking, high heart rate, sweat

Fig. 4 An embodied predictive coding view of the wind-versus-thief inference

moving in the shadow should have minor influence on the inference. On the contrary, in most cases interoceptive information is quite certain, so it has a greater influence on the inference. We will see later that this influence is, technically, proportional to the *precision* of the evidence at hand.

One way to formulate this idea is by using *Bayesian multisensory integration* (Ernst & Bülthoff, 2004). In this framework, perception is multisensory. The sources of evidence (e.g., vision or touch) can be unreliable, so to provide a robust estimate of the multisensory evidence E, all of the available information has to be integrated and weighted by the relative uncertainty of the sources. For example, visual and auditory streams can both provide information on the location of an object that is simultaneously seen and heard, but in most cases the visual information has to be weighted more, because (at least for location judgments) it is more reliable. The multimodal integration can proceed, for example, by using the following maximum likelihood estimation (MLE) equation:

$$\widehat{E} = \sum_i w_i \widehat{E}_i.$$

In the equation, the estimation performed using MLE is called $\hat{E}$ (rather than E, as before) to emphasize that it is more than just a sensory measurement: It is a weighted sum (with weights w_i) of the individual estimates $\hat{E}_i$. All of the weights sum to 1, as is specified in the following equation:

$$\sum_i w_i = 1.$$

The weights have to be proportional to the reliability of the information sources. One way to model this in a probabilistic framework is by making the weight proportional to the *precision* or *inverse variance* of the individual estimates. Intuitively, the more uncertain the information source (e.g., visual or auditory), the less its weight, as expressed by the following equation (where σ_i^2 is the variance):

$$w_j = \left\{ \frac{1/\sigma_i^2}{\sum\limits_{i=1...,j...N} 1/\sigma_i^2} \right\}.$$

Using the same logic in the wind-versus-thief example, we assume that the estimate $\hat{E}$ of the evidence integrates and weights multiple information sources (in our case,

multisensory and interoceptive) using the aforementioned equations. In principle, it is also possible to include a prior probability to the MLE in order to make the integration more robust (Ernst & Bülthoff, 2004). However, a quantitative specification of $\hat{E}$ is not important here; what is important is that $\hat{E}$ is a multimodal state that emphasizes (and assigns more weight to) interoceptive information, rather than auditory and visual information.

Let's redo the calculations again and consider how well the competing hypotheses (wind vs. thief) explain your new estimate of the evidence $\hat{E}$. Now the picture is different from before, because not only does the thief hypothesis explain more evidence, it also explains better the evidence coming from the most reliable source (interoception), which is emphasized in $\hat{E}$. On the contrary, the wind hypothesis explains only the less reliable evidence (auditory and visual). For the sake of simplicity, I assume that this maps onto a decrease of the likelihood of the wind hypothesis and an increase of the likelihood of the thief hypothesis, and I assign $P(\hat{E} \mid \text{wind}) = .15$ and $P(\hat{E} \mid \text{thief}) = .9$. If we again apply Bayes's rule, your belief in the thief hypothesis now rises to .6, and yes, you should start worrying about the thief!

The takeaway from this simple example is that we should put mind and body together again in predictive coding models of perception and cognition. If we consider interoceptive information as a source of evidence in the same way that we consider sensory events, it naturally emerges that it can influence perceptual inference, belief formation, and choice. Importantly, in some cases, such as during the night, interoceptive information can be more reliable than sensory (e.g., visual) stimuli. Because predictive coding weights information sources according to their reliability and precision, the prediction errors associated with interoceptive states will enjoy great functional efficacy. Perhaps this is why the bogeyman only visits us in the dark.

Note that this example is not explained well by cognitive bias (e.g., a priming of the thief hypothesis after watching a zombie movie, which can be modeled as a change in the prior probability). Rather, the way that the horror movie or the small car accident influences the current (wind vs. thief) inference is through the body state. The mechanism that I propose is an *embodied predictive coding* that gives sensory and interoceptive information equal dignity.

Predictive coding hierarchies

I mentioned that predictive coding hierarchies can include heterogeneous sources of evidence, say sensory and interoceptive. But in addition, predictive coding hierarchies can go well beyond two or three levels and can combine heterogeneous elements at the different levels. At the lowest levels, perceptual hypotheses that are close to the sensory and

interoceptive events are considered; at the highest levels more profound regularities can be represented, and the hierarchy can include, at least in principle, long-term beliefs that are increasingly more removed from sensorimotor events and that are mainly acquired through cultural learning (but that still remain grounded through the linkage with lower-level events).

Scaling up the predictive coding inference to multiple levels requires replicating the prediction-error minimization process for each pair of consecutive levels. The winning hypothesis is the one producing the least total error in the whole hierarchy. This hierarchical inference can be described as a *free energy minimization*, which is discussed in detail by Friston (2005, 2010).

The fact that hierarchical predictive coding can use multifarious information leads us again to the problem of how to select the right evidence for a given inference. For example, why use your high heart rate as evidence for the wind-versus-thief competition, given that it is due to the car accident or the horror movie? Although this specific example might seem straightforward, even in this simple case, establishing the right causal structure of a given problem is hard. One reason is that the interoceptive flow can have a long duration, and body states tend to change more slowly than sensory events. Evidence has indicated that subjective emotional responses tend to persist longer than the emotional stimulation periods (Garrett & Maddock, 2001). Similarly, a horror movie can generate an arousal state that persists after the end of the movie, and this makes more complex the attribution of a body state that you sense now to an event (the horror movie) that ended hours before. In general, estimating the right causal relations between hypotheses and sources of evidence is far from trivial; it can be considered a central problem of cognitive development and cognitive processing (Tenenbaum, Kemp, Griffiths, & Goodman, 2011).
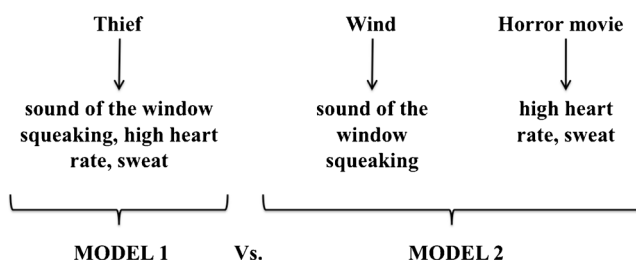
Here I will cast the problem as one of Bayesian *model selection* (Koller & Friedman, 2009); see Fig. 5. In this view, two *models* compete, not only two hypotheses. In Model 1, the same hypothesis (the thief) explains both your sensory state (the sound of the window squeaking) and your interoceptive state (your heart rate). In Model 2, one hypothesis (the wind) explains your sensory state but not your interoceptive state, and another hypothesis (the horror movie) explains your

interoceptive state but not your sensory state; so, to make sense of what is happening, you need to jointly maintain two hypotheses.

It is at this point that we can see the relevance of the *model evidence* in adjudicating between different models. Model evidence can be expressed as "accuracy" minus "complexity." This means that the model with the greatest evidence is not necessarily the most accurate one; it also has to have the minimum complexity (i.e., Occam's razor). Above, we have seen that although Model 2 may provide a more accurate explanation for the sensations, it has more degrees of freedom, and is therefore more complex. It is entirely possible that you might have entertained both models when accounting for the squeaking window and still have preferred Model 1, even if it did not explain the sensory (exteroceptive and interoceptive) input as fully as Model 2. This leads to the interesting notion that our brains may do Bayesian model comparison and model selection in order to choose categorical hypotheses about states of the world.

It is worth noting that the inference that I have described is not instantaneous, but takes time to complete and has rich internal dynamics; for example, mental states can oscillate and can include partially conflicting hypotheses that are reconciled at some later point in time (Kiebel, Daunizeau, & Friston, 2008; see also Spivey, 2007). Consider the case in which one hypothesis is highly plausible (i.e., generates low prediction error) at certain levels of the hierarchy, but not at other levels. In our example, at some (higher) level of the hierarchy, the thief hypothesis might sound very unlikely, but at other (lower) levels, it might sound so plausible that you freeze in your bed or get up and turn the light on. An iterative hypothesis-testing process guarantees that at some point the prediction error will be minimized at all levels, but the conflict between levels might require time to settle and give rise to instabilities; see Hohwy, Roepstorff, and Friston (2008) for an illustration of instability in the perceptual dynamics of predictive coding, and Kiebel et al. (2008) for a proof of principle using dynamical instabilities to model the recognition of sensory sequences.

Furthermore, the inference can produce "changes of mind" (say, from Model 1 one to Model 2 of Fig. 5), especially when not all of the evidence is available from the beginning, or more generally when the process is *nonstationary* (i.e., different intervals show stronger and weaker evidence for each of the competing hypotheses). In our example, interoceptive information might be available immediately that points to Model 1; subsequently, additional evidence might be gathered that points to Model 2 (e.g., after a while, you remember that the last time you watched a horror movie you had similar bad dreams). As a result, you might initially consider Model 1 to be more parsimonious, and thus more likely. Successively, when additional information becomes available and "percolates" through the predictive coding hierarchy, you might

```
    Thief              Wind         Horror movie
      |                 |                |
      v                 v                v
sound of the window  sound of the    high heart
squeaking, high heart   window       rate, sweat
   rate, sweat        squeaking

    MODEL 1      Vs.        MODEL 2
```

Fig. 5 Model selection. See the main text for explanations

change your mind and prefer Model 2. This idea suggests that changes of mind can arise within a single inference process in which different kinds of evidence (say, sensory vs. conceptual) are available at different time intervals; this is something that I will return to in the discussion of *dual-process* theories in the Conclusions section.

Overall, nothing during the predicting-coding inference can be considered as a fact carved in stone. Every piece of evidence is evaluated and weighted by its uncertainty. Even the strongest hypotheses continuously compete and can be revised in the light of new evidence, although, of course, knowledge at the higher levels (e.g., one's own *core beliefs*) are very hard to change.

Interoceptive information and motivation

Interoceptive information can come in different varieties. Up to now I have focused on body states such as heart rate. But also, visceral and autonomic states such as hunger or thirst are monitored by sympathetic and parasympathetic flows. This implies that perceptual processing can be modulated by the motivational conditions (e.g., neutral, hungry, thirst). Along these lines, Montague and King-Casas (2007) argued that

A sated and comfortable lioness looking at two antelopes sees two unthreatening creatures against the normal backdrop of the temperate savannah. . . . The same lioness, when hungry, sees only one thing—the most immediate prey. . . . In another circumstance, in which the lioness may be inordinately hot, the distant, shaded tree becomes the prominent visual object in the field of view. (p. 519)

Montague and King-Casas went even beyond this, and proposed that motivational states modulate perceptual saliency:

[T]he mismatch between the internal need (to stay at comfortable temperature) and the external signals (it is hot outside) changes the importance of the visual signals. (p. 519)

This implies that the weight of evidence should be modulated by its behavioral significance or salience, and not only by the uncertainty of its information source. In keeping with this idea, evidence exists that the coding strategies of sensory neurons are influenced by the saliency of the stimuli, and the behaviorally relevant events are emphasized (Machens, Gollisch, Kolesnikova, & Herz, 2005). In generalized predictive coding schemes, behaviorally relevant or salient events are emphasized by virtue of having more precision. The top-down control of precision has been considered in terms of attention (Feldman & Friston, 2010) and highlights

the fact that not only do hypotheses have to be optimized, but the confidence in these hypotheses has to evaluated. Future research will be needed to elucidate whether visceral and autonomic states can influence perceptual and cognitive inference by modulating the precision of the motivationally relevant stimuli, and whether this influence can be described in terms of attention.

Interoceptive information, feeling, and emotion

I have proposed that in embodied predictive coding, interoceptive states modulate inferences. However, the converse is also true: Inference dynamics can modify or produce new interoceptive states. Consider that predictive coding architectures are *generative*, and during the inference they almost literally "synthesize" the predicted sensory states via top-down links (as is summarized by a nice article title: "To recognize shapes, first learn to generate images"; Hinton, 2007b). This mechanism is useful because the "synthesized" sensory expectations can be directly compared to the sensory reality so as to assess the plausibility of a hypothesis. As a consequence, during the inference, the idea of a thief seems so vivid that you can almost see a silhouette moving in the shadows. Similarly, during inference, interoceptive states can be "synthesized" by the generative dynamics of embodied predictive coding, and this modifies your body and interoceptive state, as well as what you feel.

And there are further consequences. Up until now, we have considered predictive coding from a purely perceptual perspective. In other words, we have regarded perception as minimizing prediction errors in the exteroceptive and interoceptive domains, to arrive at the best explanation or hypothesis for multimodal sensations. A generalization of predictive coding known as *active inference* (Friston, Daunizeau, & Kiebel, 2009) considers that proprioceptive prediction errors can be minimized through action, which reduces to engaging reflex arcs. If we now extend the same argument to interoceptive signals, we have a mechanism for autonomic control by engaging autonomic reflexes. This introduces a *circular causality*, in which sympathetic arousal (for example) can be both a cause and consequence of emotional or salient perceptual states. Previously, we had considered that an elevated heart rate was the residual consequence of some prior experience. In the more general setting of active inference, it is possible that the belief that a thief has entered the home produces predictions about elevated heart rates that are fulfilled automatically, through sympathetic reflexes. These, then, may reinforce the embodied predictive coding of uncertain auditory cues and produce cascading effects that ultimately reinforce the bogeyman belief.

To sum up, the dynamics of generative processes and active inference suggest a bidirectional link between perceptual and

cognitive inference, on the one hand, and feelings and emotions, on the other hand, which deserves empirical investigation.

A further link between our framework and emotion is provided by the *interoceptive predictive coding* account of conscious presence and its disturbances. Seth, Suzuki, and Critchley (2012) defended a view of feelings that can be traced back at least to James (1890) and that has been recently revitalized by Damasio (2000). In this framework, feelings and emotions depend on the perception of changes in the body. According to the James–Lange model (James, 1890), aspects of felt emotion involve the perception of our own bodily (visceral, interoceptive) states, and changes in the body (e.g., visceral events) should be considered the *causes* of feelings, and not their consequences. In other words, I feel fear because I can sense my viscera "moving," rather than vice versa (as some might assume). The predictive coding account is well suited to explain this cause–effect relation. Seth & Critchley (2012) established this link to argue that "subjective feeling states are constituted by continually updated predictions of the causes of interoceptive input" (p. 228).

The interoceptive predictive coding idea can explain how perceptual and cognitive inference produce new feelings and emotions. Interoceptive states that are "synthesized" as part of perceptual and cognitive inference enter the interoceptive predictive coding scheme, and as a consequence modify or generate new subjective feeling states.

The theory that I propose and the idea of interoceptive predictive coding target different phenomena, but they can be seen as complementary. I propose that we need to incorporate interoceptive information in the standard predictive-coding models of perceptual and cognitive inference. The interoceptive predictive coding view is a specific account of emotional feelings and conscious presence that only considers interoceptive inputs. The two mechanisms can interact bidirectionally, so that subjective feeling states and emotions influence perceptual and cognitive inference, which in turn produce new subjective feeling states and emotions. Understanding these dynamics in depth might shed light on self-regulatory brain mechanisms and the mind–body problem (see also Carhart-Harris & Friston, 2010).

What's in a bogeyman: The stuff dreams are made of

Up to now, I have focused on how sensory and interoceptive states constrain inference. However, my proposal entails an embodied view of concepts and representations, too. I argue that *interoceptive information is part and parcel of the representation of entities*, such as "wind," "thief," and many others. These concepts are formed and maintained in the brain through predictive coding hierarchies or more general generative architectures (Dayan, Hinton, Neal, & Zemel, 1995; Hinton, 2007a) that link low-level inputs to higher-level representations.

In this view, concepts are grounded in embodied states and link to a multifarious set of sensory and interceptive elements, all of which are considered in the embodied predictive coding inference. For example, your concept of a thief could include an affective (in this case, negative, frightening) component that derives from, say, a past theft experience. The concept thus links to interoceptive predictions that the embodied predictive coding inference must consider in the same way as all of the other available (e.g., sensory, conceptual) information. This idea can provide a computational basis for an embodied cognitive accounts of concepts, including nonperceptual and emotional concepts (Barsalou, 2008; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011).

Interoceptive information can include descending predictions or *corollary discharges* of motor control signals, too. In this way, the representation of objects and events can also include information about the *affordances* and the likely actions that we perform on them, which can be used as evidence in the inferential scheme in the same way that sensory stimuli are. Action-based approaches to cognition argue that objects can be recognized in terms of the expected sensory consequences of possible actions produced by forward models; for example, a sponge can be understood in terms of the (anticipated) softness when squeezing it or imagining squeezing it (Grush, 2004; Pezzulo, 2008, 2011; Roy, 2005). Here the perspective is slightly different, in that corollary discharges of actions, rather than only sensory predictions, are used as evidence; in other terms, the inference looks like, "because I am squeezing it, it should be a sponge." In this setting, an affordance is the attribute of a hidden cause in the environment that induces (through predictive coding) predictions in the exteroceptive and proprioceptive domains. In other words, one cannot divorce the perceptual attributes from how one would physically interact with the inferred object in an embodied context.

The arguments that I have made in this article leave open the possibility that some objects or events may be understood and experienced purely (or primarily) using interoceptive information. The bogeyman seems to be such an entity. According to Wikipedia (http://en.wikipedia.org/wiki/Bogeyman, retrieved on July 1st, 2013):

> A *bogeyman* (also spelled *bogieman*, or *boogeyman*) is an amorphous imaginary being used by adults to frighten children into compliant behaviour. The monster has no specific appearance, and conceptions about it can vary drastically from household to household within the same community; in many cases, he has no set appearance in the mind of an adult or child, but is simply

a non-specific embodiment of terror. Parents may tell their children that if they misbehave, the bogeyman will get them.

To the best of my knowledge, nobody has ever seen a bogeyman, so this entity could be understood mostly in terms of visceral and interoceptive signals that its putative presence (should) generate. In other words, you quite literally recognize a bogeyman with your body, and with your fear in particular. As a consequence, the bogeyman idea is a form of *self-fulfilling prophecy*, because a terrified child can take his or her terror as evidence that the bogeyman exists (and is probably close), and the terror itself can increase due to the circular causality mentioned earlier. This idea resonates well with recent proposals in embodied cognitive theories arguing that some abstract concepts (including, e.g., emotional concepts) could be grounded primarily on interoceptive information (Barsalou, 2008).

Note, however, that even in the bogeyman case, interoception is not the only available information. Part of the grounding and the understanding of nonperceptual concepts such as the bogeyman are in terms of "tales" heard from parents or friends; this explains why the attributed appearance of the bogeyman varies between communities (see Halloy, 2012 for a related discussion in the context of spirit possession phenomena). In principle, the predictive coding scheme can integrate diverse kinds of information, from sensory and interoceptive to tales and narratives, plausibly considering them at different hierarchical levels and weighting them depending on the reliability of the source (including social sources such as parents and friends). However, up to now the predictive coding framework has been used mostly to understand perceptual events that have clear sensory components. Understanding how different and heterogeneous elements combine in predictive coding architectures, and their relative importances, can shed light on the architecture of conceptual knowledge (Barsalou, 2008; Pezzulo, 2012; Pezzulo and Castelfranchi, 2009; Pezzulo et al., 2011, 2013).

Conclusions

I have proposed an embodied predictive coding model of perceptual and cognitive inference in which interoceptive dynamics are treated as evidence, similar to sensory dynamics. The inference is embodied, in that it is deeply influenced by the body and its motivational and emotional dynamics; in turn, the body and the motivational and emotional states also change in response to the inference. A prediction of this model is that interoceptive signals linked to body states (e.g., heart rate) and visceral and autonomic states (e.g., hunger or thirst)

should affect, and in turn be affected by, perceptual and cognitive inferences. Affective influences on perception have been reported (see, e.g., Anderson & Phelps, 2001), but further evidence will be necessary to assess the bidirectional interactions between mind and body proposed here.

The hypotheses discussed here can be tested using methods that track the temporal dynamics of inferences in both the brain and the body. Predictive coding schemes formulated in continuous time suggest that we can understand evoked neuronal responses in terms of suppression of prediction error that takes a finite amount of time. For example, the oddball event-related potential (ERP) peaks and oddball-dependent differences in ERPs (such as the mismatch negativity) have been described in terms of predictive coding by several authors (Wacongne, Changeux, & Dehaene, 2012). In principle, this logic can be extended to embodied predictive coding by considering jointly brain and body dynamics—for example, by simultaneously measuring event-related potentials, electromyographic, and autonomic signals during inference (while also manipulating and interfering with the brain and body states). By simultaneously recording time-course data in both the brain and the body, this method could help shed light on their bidirectional links and unveil the embodied aspects of perceptual and cognitive inference.

Malfunctioning of the embodied predictive coding mechanism can have dramatic effects. For example, patients affected by *Capgras syndrome* are unable to recognize their friends or family members and believe that impostors or aliens have replaced them. This syndrome is usually associated with impaired autonomic response and with the failure to sense normal affective states (e.g., pleasure or joy) associated with a friend's or a parent's face. It has been proposed that the syndrome depends on broken neuronal associations between the temporal cortices, where faces are recognized, and the limbic system, where the associated emotions are processed (Hirstein & Ramachandran, 1997). This idea can be easily cast within a predictive coding scheme. To infer who is the person in front of me, sensory predictions of the perceptual appearance (e.g., of the mother) and interoceptive predictions of affective states (e.g., pleasure and joy) are treated on the same grounds. The precision, and thus the weight, of interoceptive signals is so strong that it is hard to verbally convince a patient of the abnormality of his inference.

Indeed, people imagining a thief (or the bogeyman) in the night, or even Capgras patients, are not performing any irrational inference. Rather, their inference maximizes the probability of the correct hypothesis, given the evidence and the relative weights of the information sources. Because the embodied inference that I propose includes interoceptive information, rationality depends on an inextricable link between mind and body. Thus, the bad news is that the experience of "fear in the night" is not irrational; rather, you have valid reasons to fear thieves, killers, and even the bogeyman.

This view is at odds with *dual-process* theories that posit an emotional fast route (called "System 1") and a slower reasoning route (called "System 2"), with only the latter being rational (Kahneman, 2003; Stanovich & West, 2000). In the framework that I have proposed, sensory and interoceptive evidence can be considered simultaneously or one before the other, depending on how quickly the information sources contribute to the predictive coding inference. Still, sensory or interoceptive dynamics are not two separate systems: They are both expressions of a single process that is fully rational in how it considers the statistics of both the sensorium and interoceptive information. Understanding the interactions between sensory and interoceptive dynamics could be highly relevant for the emerging field of computational psychiatry (Montague, Dolan, Friston, & Dayan, 2012).

For the sake of simplicity, I skipped many of the complex details of predictive coding and generative inference, but the framework is rich enough to offer several interesting directions for future research on the mind–body problem. For example, sophisticated ways exist to incorporate the *precision* of information in a predictive coding inference (Friston, Adams, Perrinet, & Breakspear, 2012); precision-based optimization schemes could generate useful hypotheses as to how interoceptive information is integrated in perceptual inference. Furthermore, a more detailed inference should consider the utility of the hypotheses, rather than only their probability—for example, the expected gains or losses, if the inference is incorrect (e.g., the risk that I am not recognizing a thief or bogeyman). Risk-sensitive inference (Grau-Moya, Hez, Pezzulo, & Braun, 2013; Grau-Moya, Ortega, & Braun, 2012) formalizes these ideas and could help explain how subjects with different personality traits (e.g., risk-seeking or risk-avoidant) consider best or worst cases in the inference rather than an unbiased estimate.

Recent work in the predictive coding framework has emphasized that attention dynamics and active perception modulate perceptual inference by permitting us to probe the world actively and by running "experiments" to disambiguate the current hypotheses (Friston et al., 2012). An interesting direction for future research would be to study whether and how the active probing can be extended from sensory to interoceptive domains.

Finally, as we discussed, the active-inference theory extends predictive coding from perceptual to action domains (Friston et al., 2009). In this framework, action can be used to suppress sensory prediction errors, and goal achievement corresponds to fulfilling the predictions encoded in the high-level prior beliefs (e.g., states that are highly valued for an organism). An interesting direction for future research would be to look at homeostatic regulation within the active-inference framework. This framework permits studying the causation of proprioceptive and interoceptive responses through (motor and autonomic) reflexes and how action can

be used to suppress the interoceptive signals linked to undesired states such as fear. At minimum, this would provide some relief from the bogeyman, showing that simply turning the light on will dismiss it.

References

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411,* 305–309. doi:10.1038/35077083

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645. doi:10.1146/annurev.psych.59.103006.093639

Carhart-Harris, R. L., & Friston, K. J. (2010). The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas. *Brain, 133,* 1265–1283. doi:10.1093/brain/awq010

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36,* 181–204. doi:10.1017/S0140525X12000477

Damasio, A. (2000). *The feeling of what happens: Body and emotion in the making of consciousness*. New York, NY: Harcourt Brace.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation, 7,* 889–904.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8,* 162–169. doi:10.1016/j.tics.2004.02.002

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience, 4,* 215. doi:10.3389/fnhum.2010.00215

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B, 360,* 815–836. doi:10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11,* 127–138. doi:10.1038/nrn2787

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology, 3,* 151. doi:10.3389/fpsyg.2012.00151

Friston, K., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS ONE, 4,* e6421. doi:10.1371/journal.pone.0006421

Garrett, A. S., & Maddock, R. J. (2001). Time course of the subjective emotional response to aversive pictures: Relevance to fMRI studies. *Psychiatry Research, 108,* 39–48.

Grau-Moya, J., Hez, E., Pezzulo, G., & Braun, D. A. (2013). The effect of model uncertainty on cooperation in sensorimotor interactions. *Journal of the Royal Society Interface, 10,* 10130554. doi:10.1098/rsif.2013.0554

Grau-Moya, J., Ortega, P. A., & Braun, D. A. (2012). Risk-sensitivity in Bayesian sensorimotor integration. *PLoS Computational Biology, 8,* e1002698. doi:10.1371/journal.pcbi.1002698

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences, 27,* 377–396.

Halloy, A. (2012). Gods in the Flesh: Learning Emotions in the. Xangô Possession Cult (Brazil). *Ethnos: Journal of Anthropology, 77,* 177-202. doi:10.1080/00141844.2011.586465

Helmholtz, H. von. (1962). Concerning the perceptions in general. In J. P. C. Southall (Ed. and Trans.), *Helmholtz's Treatise on physiological optics* (Vol. 3). New York, NY: Dover. (Original work published 1866).

Hinton, G. E. (2007a). Learning multiple layers of representation. *Trends in Cognitive Sciences, 11,* 428–434. doi:10.1016/j.tics.2007.09.004

Hinton, G. E. (2007b). To recognize shapes, first learn to generate images. *Progress in Brain Research, 165,* 535–547. doi:10.1016/S0079-6123(06)65034-6

Hirstein, W., & Ramachandran, V. S. (1997). Capgras syndrome: A novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society B, 264,* 437–444. doi:10.1098/rspb.1997.0062

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition, 108,* 687–701. doi:10.1016/j.cognition.2008.05.010

James, W. (1890). *The principles of psychology*. New York, NY: Henry Holt.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697–720. doi:10.1037/0003-066X.58.9.697

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology, 4,* e1000209. doi:10.1371/journal.pcbi.1000209

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. Cambridge, MA: MIT Press.

Machens, C. K., Gollisch, T., Kolesnikova, O., & Herz, A. V. M. (2005). Testing the efficiency of sensory coding with optimal stimulus ensembles. *Neuron, 47,* 447–456. doi:10.1016/j.neuron.2005.06.015

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences, 16,* 72–80. doi:10.1016/j.tics.2011.11.018

Montague, P. R., & King-Casas, B. (2007). Efficient statistics, common currencies and the problem of reward-harvesting. *Trends in Cognitive Sciences, 11,* 514–519. doi:10.1016/j.tics.2007.10.002

Pezzulo, G. (2008). Coordinating with the future: The anticipatory nature of representation. *Minds and Machines, 18,* 179–225. doi:10.1007/s11023-008-9095-5

Pezzulo, G. (2011). Grounding procedural and declarative knowledge in sensorimotor anticipation. *Mind and Language, 26,* 78–114.

Pezzulo, G. (2012). An active inference view of cognitive control. *Frontiers in Theoretical and Philosophical Psychology, 3,* 478. doi:10.3389/fpsyg.2012.00478

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2011). The mechanics of embodiment: A dialog on embodiment and computational modeling. *Frontiers in Cognition, 2*(5), 1–21. doi:10.3389/fpsyg.2011.00005

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2013). Computational grounded cognition: A new alliance between grounded cognition and computational modeling. *Frontiers in Psychology, 3,* 612. doi:10.3389/fpsyg.2012.00612

Pezzulo, G., & Castelfranchi, C. (2009). Thinking as the control of imagination: a conceptual framework for goal-directed systems. *Psychological Research, 73,* 559–577.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2,* 79–87. doi:10.1038/4580

Roy, D. (2005). Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence, 167,* 170–205.

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology, 2,* 395. doi:10.3389/fpsyg.2011.00395

Spivey, M. (2007). *The continuity of mind*. Oxford, UK: Oxford University Press.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645–665. disc. 665–726.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331,* 1279–1285. doi:10.1126/science.1192788

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience, 32,* 3665–3678. doi:10.1523/JNEUROSCI.5003-11.2012

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia, 49,* 1105–1127. doi:10.1016/j.neuropsychologia.2010.12.032