Opinion

Getting ahead: forward models and their place in cognitive architecture

Martin J. Pickering[1*] **and Andy Clark**[2*]

[1] Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH9 9JZ, UK
[2] School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, EH8 9AD, UK

The use of forward models (mechanisms that predict the future state of a system) is well established in cognitive and computational neuroscience. We compare and contrast two recent, but interestingly divergent, accounts of the place of forward models in the human cognitive architecture. On the Auxiliary Forward Model (AFM) account, forward models are special-purpose prediction mechanisms implemented by additional circuitry distinct from core mechanisms of perception and action. On the Integral Forward Model (IFM) account, forward models lie at the heart of all forms of perception and action. We compare these neighbouring but importantly different visions and consider their implications for the cognitive sciences. We end by asking what kinds of empirical research might offer evidence favouring one or the other of these approaches.

Two roles for forward models

There is a great deal of evidence that people predict both themselves and other people [1,2]. But how do they do it? Recent proposals suggest that people use forward models [3–5] to make predictions (see Glossary). However, there are two different types of account of how they might be used. The first account (Figure 1) assumes a dedicated prediction mechanism implemented by additional circuitry distinct from the core mechanisms of perception and action. Those core mechanisms involve one or more distinct inverse models, which compute motor commands from desired effects. We call this complex the AFM account. The second account (Figure 2) is more integrated and posits a forward (generative) model as the core machinery of perception and action. We call this the IFM account. In the IFM, motor commands are replaced by the descending web of sensory predictions issued by the forward model. This removes any fundamental distinction between motor and sensory processing and sidesteps the need for a distinct inverse model or for the learning and use of multiple (paired) forward and inverse models. The two accounts thus share a great deal, but the differences between them are important, making them ripe (we argue) for directly contrastive research and experiment.

AFMs

When I plan an action, for instance moving my arm to a target, I construct an action command, use that command to perform the action, and experience the sensory (including proprioceptive) consequences of that action. If I repeatedly perform the action, I can learn from my mistakes (e.g., changing the plan slightly if my arm just misses the target). Over time, I can predict that if I instigate an action command, I will subsequently experience a particular result. In the same way, if I decide to name an object in

Glossary

Active inference: the combined mechanism by which perceptual and motor systems conspire to reduce prediction error using the twin strategies of altering predictions to fit the world and altering the world (including the body) to fit the predictions.
Corollary discharge: often (incorrectly) used synonymously with 'efference copy', this names the output of the forward model (the predictor mechanism), which may be used to influence further processing.
Efference copy: a copy of the current motor command that may be given as input to a forward model.
Forward model: a mechanism that predicts the future state of a system. In standard control theory, this would be an internal loop whose input is a copy (an efference copy) of a control signal such as a motor command and whose output is a prediction about the next sensory state. In AFMs, an 'inverse model' computes the motor command, which is then copied to the forward model, which issues predictions. In IFMs, proprioceptive predictions issued by the forward model act directly as motor commands and there is no role for an efference copy (Figures 1 and 2).
Generative model: a description that allows a system to self-generate data that are similar to the observed data. Usually, that means a model that captures the statistical structure of some set of observed inputs by tracking (in effect, schematically recapitulating) the causal matrix responsible for that structure. The dynamics of the units encoding such a model are used to predict inputs to the system. A generative model thus generates consequences from causes in the same way that a forward model maps from causes to consequences. Forward models are thus examples of generative models.
Inverse–forward scheme: a scheme that posits two distinct models – an inverse model (or optimal control model) that converts intentions into motor commands and a forward model that converts motor commands into sensory consequences (which are compared with actual outcomes for online error correction and learning).
Inverse model: a mechanism that takes the intended position of the body as input and estimates the motor commands that would transform the current position into the desired one.
Joint action: an action that involves appropriately timed and coordinated contributions from two (or more) agents.
Predictive coding: a data-compression strategy in which only the discrepancies between predicted and expected values (residual errors, or 'prediction errors') are used to drive further processing (Box 2).
Proprioception: the 'inner' sense that informs us about the relative locations of our bodily parts and the forces and efforts that are being applied.
Simulation: interpreting another's actions by reproducing the processes that one would use to perform that action.

Corresponding authors: Pickering, M.J. (martin.pickering@ed.ac.uk); Clark, A. (andy.clark@ed.ac.uk).
Keywords: forward model; active inference; prediction; simulation; predictive coding; joint action; language.
*The authors contributed equally to this work.

1364-6613/
 http://dx.doi.org/10.1016/j.tics.2014.05.006

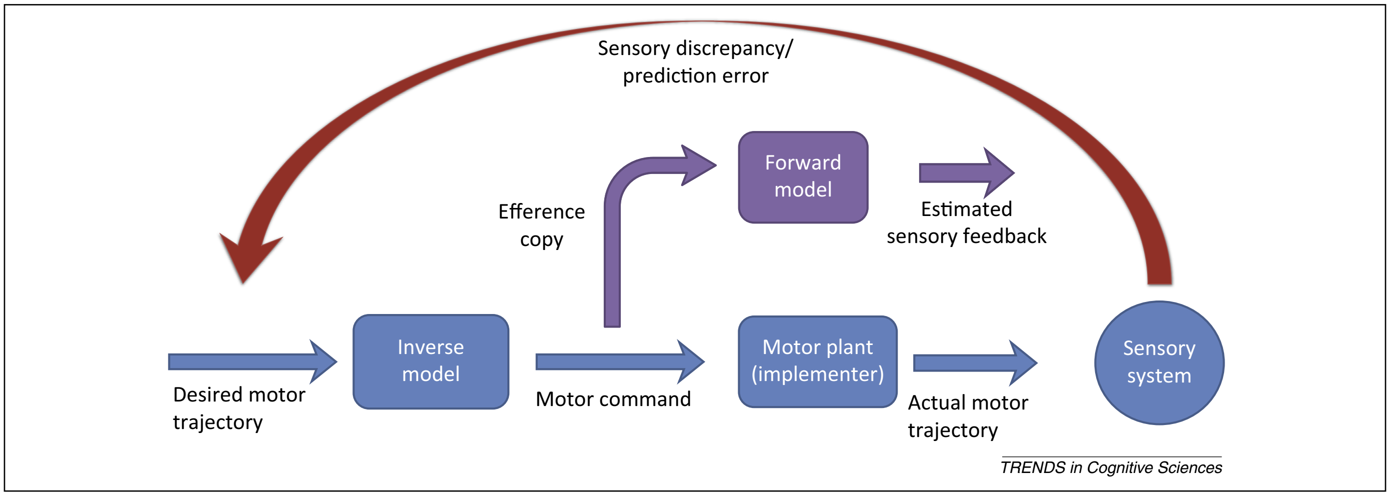

Figure 1. Auxiliary Forward Model (AFM) architecture. In this architecture, the output of the inverse model is a motor command, copied to the forward model, which is used to estimate sensory feedback.

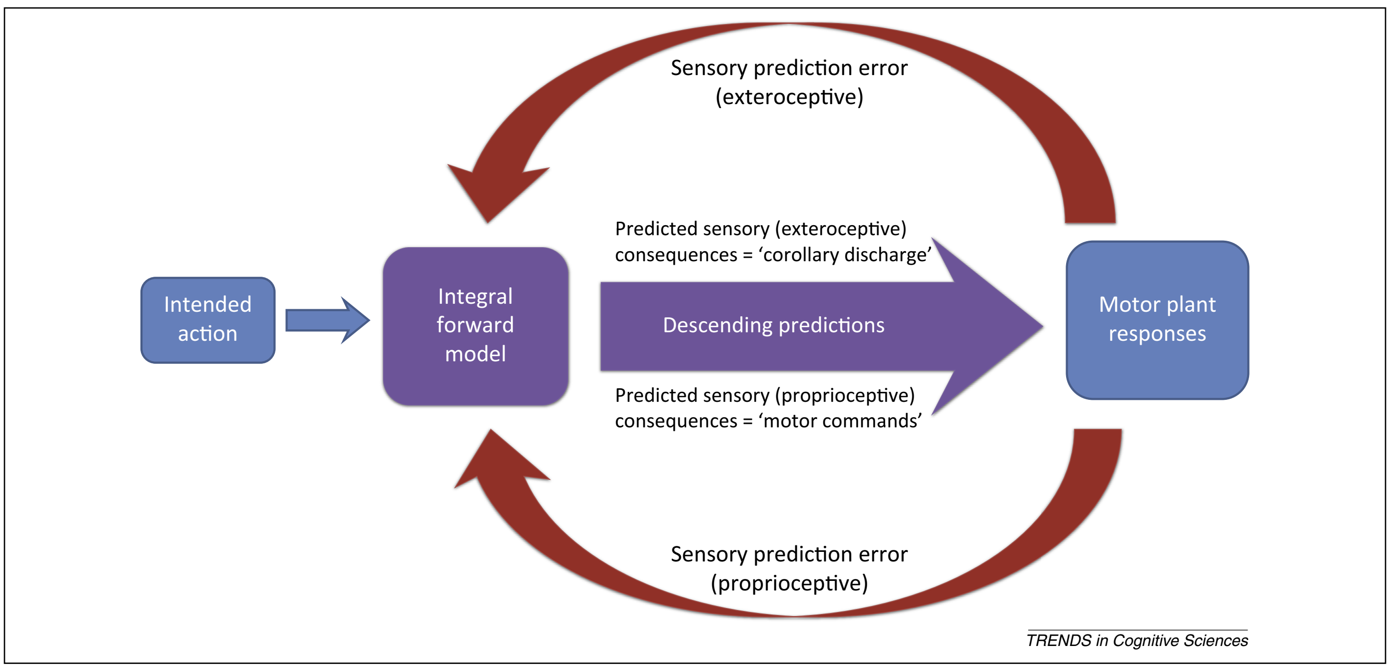

Figure 2. Integral Forward Model (IFM) architecture. In this architecture, the predictions from the forward model act as action commands and there is no need for an efference copy as such.

a picture (e.g., in a psycholinguistic experiment), I can predict what I will hear myself saying.

According to the AFM account (Figure 1), such predictions involve dedicated machinery that may involve different representations from those involved in generating the action itself [6–8]. In arm movement, the action command causes the action to be implemented using muscles, nerve fibres, and so on (these can be equated with the 'motor plant' in engineering). However, the action command also sends a so-called efference copy that is used to drive predictions that do not make use of the implementer. Instead, the efference copy serves as input to a forward model of the action that generates the projected sensory consequences of those commands as output. This forward model could simply involve a look-up table, but is more likely to involve calculations (e.g., approximations to the laws of mechanics), and the output is in general computed before the action is performed. As a simple analogy, I turn my radiator up from 'off' to half way. Well before the radiator heats up, I predict (based on repeated experience with my central heating) that it will take 5 min to heat by 10°C (using very simple equations; e.g., increase of 2°C/min for each 30° turn). I can act on the prediction right away (e.g., take my coat off) or compare the prediction with the results and learn from any discrepancy via my inverse model (e.g., and therefore turn the knob further).

Such accounts thus posit two distinct models – an inverse model (or optimal control model) that converts intentions into motor commands and a forward model that converts motor commands into sensory consequences (which are compared with actual outcomes for online error correction and learning). Importantly, the forward model is distinct from the plant (the central heating system). Similarly, my forward model of arm movement is distinct from

Box 1. Predicting language by simulation

After hearing a boy say 'I want to go and fly my...', you might predict that he will shortly utter 'kite' – a word that refers to a flyable object, is a noun, and begins with the phoneme /k/. According to prediction by simulation, you covertly imitate the boy's utterance, make contextual allowances for the differences between yourself and him, and use this to derive the production (i.e., motor) command that the boy would use to produce 'kite'. You then use the efference copy of this production command to derive forward models of the meaning, grammar, and sound of the utterance and can then act on these predictions; for example, comparing them with the boy's utterance when it occurs [28]. There is good evidence for covert imitation [29] and for predictions at different linguistic levels [30–32], some of which occur too quickly to be due to the implementer (i.e., the plant) [33]. Other researchers had Italian participants repeatedly hear a pseudoword (e.g., birro) and used transcranial magnetic stimulation (TMS) to reveal immediate appropriate articulatory activation (associated with rr) if they heard the first part of the same word (bi coarticulated with rro) compared with when they heard the first part of a different word (bi coarticulated with ffo). This suggests that the covert imitation facilitated speech recognition [34].

Box 2. Predictive coding, predictive processing, and active inference

Predictive coding was first developed as a data-compression strategy for commercial signal processing. Thus, consider a basic task such as image transmission: in most images, the value of one pixel regularly predicts the value of its nearest neighbours, with differences marking important features such as the boundaries between objects. That means that the code for a rich image can be compressed (for a properly informed receiver) by encoding only the 'unexpected' variation – the cases where the actual value departs from the predicted one. What needs to be transmitted is just the 'news' – the difference (also known as the prediction error) between the actual current signal and the predicted one. Descendants of this kind of compression technique are currently used in various forms of audio compression and (most notably) in motion-compressed encoding for video, where one assumes ('predicts') that the image remains the same from one frame to the next, encoding only the deviations due to motion, occlusion, and so on.

An emerging family of models of perception [10–15] deploy that core strategy of efficient encoding and transmission in the special context of a multilevel probabilistic generative model. If these 'predictive processing' [16] accounts are on track, perception is a process in which we (or rather, various parts of our brain) try to guess what is out there, using the incoming signal more as a means of tuning and nuancing the guessing rather than as a rich (and bandwidth-costly) encoding of the state of the world. Percepts here take shape only when downward predictions match the incoming sensory signal, at multiple levels of processing. Perception, if such models are correct, is a matter of the brain using stored knowledge to predict, in a progressively more refined manner, the patterns of multilayer neuronal response elicited by the current sensory stimulation.

Active inference (or 'action-oriented predictive processing') extends this story to encompass the generation of motor response. In these extensions [17,18] predictions of the proprioceptive trajectories that would ensue if a certain action were to occur generate prediction errors that are then eliminated by movement. Descending predictions thus function as motor commands. They do this by generating cascading prediction errors (because the predicted trajectory is not yet actual) that are eventually quashed by simple, low-level reflexes. Such a strategy avoids the need to solve difficult (arguably intractable) optimality equations during online processing while fluidly compensating for known signalling delays, perturbations, and sensory noise [26].

its muscular implementation. Additionally, my forward model of picture naming is distinct from the mechanisms of language production, although both 'routes' are cognitive and (largely) instantiated in the brain [1].

I could predict your behaviour simply based on my experiences as observer, just as I can predict events (e.g., how an object will fall). However, AFMs and IFMs (see below) both suggest that I predict your behaviour using the same forward models that I use for predicting my own behaviour – a process that Pickering and Garrod [1] call prediction by simulation. In the AFM version, I see your arm moving and covertly imitate your movements, using the inverse model and accommodating to differences between my own body and yours (i.e., the context) to determine the action command that I would use to move my own arm. I then take the efference copy of that command to predict your upcoming arm movement (e.g., to punch me). I can similarly predict a predictable word in your utterance. As when predicting my own behaviour, I make use of a dedicated mechanism distinct from the perceptual or language comprehension system (by 'borrowing' a mechanism developed for limb movement or language production). Importantly, this mechanism can be multilevel, so that it can be used to deal with complex hierarchical actions [9] or the meaning, grammar, and sound involved in language (Box 1).

IFMs

The alternative IFM approach originates in work on the role of prediction in perception [10–16]. In these accounts, perception itself involves the use of a forward (generative) model whose role is to construct the incoming sensory signal 'from the top down'. Mismatches between the predictions issued by the forward model and the sensory flow result in 'prediction error' signals that refine and alter the predictions, until the system settles into a coherent multilevel state. 'Predictive coding' models of perception (Box 2) deploy that strategy in the special context of multilevel systems encoding probabilistic forward models. In the multilevel (hierarchical) setting, each higher-level neuronal population provides predictions and contextual guidance to the level below.

Such accounts have recently been extended (under the umbrella of 'active inference' [17,18]) to include the control of action. This is accomplished (Box 2) by predicting the flow of sensation that would occur were some target action performed. The resulting cascade of prediction error is then quashed – ultimately at the level of spinal reflexes [19] – by moving the bodily plant to bring the action about. Action thus results from our own predictions concerning the flow of sensation – a version of the 'ideomotor' theory of James [20] and Lotze [21] according to which the idea of moving, when unimpeded by other factors, is what brings the moving about.

Predicting the behaviour of other agents is then possible by combining the two strategies (one for perception and one for action) just described. In active inference we learn to associate our own high-level plans and intentions with their sensory consequences and actions flow from the unpacking of those high-level states into predicted sensory patterns. If I am observing another agent, and that agent is, in relevant respects, like myself, the same forward

Box 3. Simulating others using IFMs

Imagine yourself observing another agent reaching, on a hot day, for a glass of chilled cola. To perceive the scene is (the IFM story suggests) to meet the incoming flow of sensory stimulations with an apt set of top-down predictions. However, those predictions may now be generated using the same forward model that would (were the weighting of predicted proprioceptive elements high) result in the agent herself, in a similar situation, reaching for the chilled cola. Such a forward model might include, for example, a high-level intention to quench one's thirst while experiencing a certain pleasant, bubbly sensation linked to a swathe of lower-level expectations concerning some probable trajectory of reaching and to a resulting flow of gross visible actions. We thus simultaneously infer both the most likely shape of the unfolding action and the most likely underlying intentions of the other agent. What ensues is thus (as in the AFM case) a complex mental simulation, but this time (unlike in the AFM approaches) that simulation reaches backwards, uncovering also the intentions of the observed agent. For fuller treatments of this scenario, in the context of a novel account of the mirror neuron system, see [22,23].

model used to predict the consequences of my own actions (the forward model linking my own intentional states to their visual and proprioceptive consequences) becomes available as a resource to predict the (visual) consequences of another's actions – and thereby infer their intentions [22].

To enable this to occur, however, the system needs somehow to differentiate between predicting our own and others' behaviour. This is achieved [23] by manipulating the weighting (the inverse variance or 'precision', reflecting the estimated reliability of the signal) of select aspects of the sensory (especially proprioceptive) prediction error signal. When engaged in self-generated action, the precision weighting of prediction errors on the predicted movement is set high, so that those proprioceptive predictions (reflecting the desired but latent trajectory of motion) are trusted and quashed by bodily motion. High precision weighting for the predicted proprioceptive trajectory allows the system to ignore sensory prediction errors reporting the fact that that the trajectory has not yet been enacted. This provides a fundamental reason (see [24]) for the often-observed attenuation of sensory information during self-produced movement.

When proprioceptive prediction error is highly weighted yet suitably resolved by action, we move, and we may feel a sense of agency or ownership regarding those actions. When observing the actions of another agent, by contrast, the weighting of proprioceptive prediction errors (associated with that specific action) is set low. Under those conditions, our own forward model becomes available to predict (and understand) the actions of others without engaging our own motor plant (see Box 3 for an informal example). In this scheme, variations in the precision weighting of prediction error thus set the context (self versus other) for perceptual inference about high-level intentions.

Comparisons between the accounts

AFMs and IFMs both implicate forward models in the production of fluent motor action. Both invoke prediction error-minimising schemes that are either identical with or formally related to familiar schemes such as Kalman filtering (see [25]; for a review and some formal comparison, see [26]). The key difference between the accounts lies in the need (or lack of it) to explicitly compute an inverse model. According to the AFM account, the forward model is distinct from the inverse model, because it involves apparatus that computes the motor commands used to drive online action. Such a model is thus free to depart considerably in form from whatever governs the true kinematics of the agent. Furthermore (in AFMs) the outputs of the forward model (i.e., the corollary discharge) do not cause movements – they are just used to finesse and predict outcomes and in learning. According to the IFM account, however, the forward model drives our own online action by generating a flow of descending predictions that constitute the set points for our reflexes (e.g., if my reflexes were a thermostat, I would simply need to predict that the temperature had increased by 10°C).

This might seem to imply that the IFM account has less freedom of form (for example, less scope for simplifications) because the model used to predict outcomes is the same model that generates our own actions. However, it is now reused (in ways modulated by precision weighting and sensitive to contextual information, including any salient information distinguishing the two agents) when predicting the actions of others. The scope for contextualisation and nuancing means, however, that we need deploy only select aspects of our own forward (generative) model, so there remains plenty of room for simplifications and mistakes (just as there is in AFMs). Moreover, to whatever extent we have found it useful to use various shortcuts (simplified or idealised models) to predict the sensory consequences of our own behaviour, we may use those same shortcuts to predict the unfolding behaviours of others.

The core difference between IFMs and AFMs is thus that, in IFMs, there is no distinct or dedicated mechanism used to predict the outcomes of our own (or others') actions. Instead, this is achieved using the same mechanism that (by predicting sensory consequence) drives our own actions, subtly contextualised both for self-made acts and for the understanding of others. What differs is thus the location of the forward model as part of a larger architecture.

A well-known example of the AFM approach is the Hierarchical Modular Selection and Identification for Control (HMOSAIC) account. The HMOSAIC account exploits a hierarchical, prediction-driven multilevel model as a means of supporting action imitation. In [27] (p. 599), Wolpert *et al.* suggested that the HMOSAIC account might likewise address issues concerning action understanding and the 'extraction of intentions'. The HMOSAIC account, however, relies on a stack of predictor–controller (forward–inverse model) pairs, whereas in IFMs there is no separate inverse model or controller. An important upshot is that IFM accounts dispense entirely with the need for efference copies. In their place, there is simply the variegated, context-sensitive web of descending predictions that underwrite perception and drive action. Differences between myself and another agent are now treated as just additional context that is able to nuance the use of a common forward (generative) model. Predicting how the visual scene will unfold from where I stand and imagining myself in your shoes (hence viewing the scene from elsewhere) are thus accomplished using the same core machinery, contextually modulated. IFMs and AFMs thus agree that a single resource (the forward model) is used to deal with self- and other prediction, but they disagree over whether this resource (the shared forward model) is distinct from the machinery that drives our own motor behaviours.

Because each of these strategies predicts self and other using a shared forward model, AFMs and IFMs are both hostage to the other agents being 'similar enough' to ourselves. At some point, this balance tips and it becomes better to treat the other as simply exotic: another part of the world to be modelled and hence an object apt for what might be thought of as 'generic prediction'. Here, too, IFMs and AFMs share a common structure, each allowing for cases where we rely on similarity-based simulation and others where we must fall back on the forms of prediction that are used for non-agents (which Pickering and Garrod [1] call 'prediction by association').

In summary, IFMs and AFMs agree that we often predict other, similar agents by applying a forward model geared to predicting sensory consequences. According to the AFM account, however, that forward model inheres in a distinct piece of circuitry that is geared to anticipating our own movements (including our own speech) but is different from the machinery that produces our movements. IFMs, by contrast, invoke a single, complex forward model geared both to predicting and bringing about the consequences of intended acts. This is, in effect, a single generative model of embodied exchange with the world. That model is variously contextually nuanced to imagine our own behaviours, to bring them about, and to predict, imagine, and understand the behaviours of others. AFMs, meanwhile, posit at least two kinds of resource here: one for predicting ourselves and others and one (not prediction based) for bringing our own actions about.

Testing grounds

The AFM account, we have seen, invokes two distinct models: an inverse model that converts intentions into motor commands and a separate forward model that converts motor commands into sensory consequences. IFMs, by contrast, automatically invert a single forward or generative model that converts intentions into sensory consequences. At the most abstract level, AFMs thus depict movements as driven by descending motor commands and simulations as handled by a further (efference copy-driven) resource, whereas IFMs depict both movement and simulation as resulting directly from descending predictions issued by the forward model. What kinds of evidence (see also Box 4) might help tease these possibilities apart?

A promising strategy is to look for double dissociations. IFMs predict that self-prediction and the prediction of other agents will be simultaneously impaired whenever action production is (nonsuperficially) impaired. In AFMs, however, impaired action production need not immediately impair simulation (either of oneself or others). Thus in AFMs the distinction between forward and inverse models makes room for double dissociations that IFMs seems to exclude. If there are distinct forward and inverse models, patients with lesions to the forward model should still be able to execute skilled movements, even if they are unable to perform online error correction or learn new movements.

Box 4. Outstanding questions

- Can evidence concerning double dissociations help decide between IFMs and AFMs?
- Can IFMs successfully accommodate cases of joint action of the kind described in the text?
- What are the computational costs and benefits associated with IFMs and AFMs?
- Are hybrid (IFM/AFM) solutions possible?
- The AFM formulation suggests that motor commands need not be engaged during action observation. By contrast, the IFM approach predicts that the same motor command units will be activated during action and its observation. IFMs thus seem to predict that motor units should display 'mirror neuron' characteristics (for some evidence that they do, see [35]), whereas the AFM formulation does not. Is this correct or can AFMs also account for these results?
- Some theories propose that forward and inverse models are represented in the cerebellum [36]. Can the data that are used to motivate such accounts be interpreted in terms of IFM as well as AFM theories? At issue here is the thorny question of what (exactly) it is that the cerebellum learns and what kinds of prediction it specialises in [37–41].
- How can we tell whether a descending neural signal is a sensory prediction (as assumed by IFMs) or a traditional motor command (as assumed by AFMs)?

By contrast, IFMs suggest that this scenario is impossible and that lesions to the forward model will always preclude skilled movements. Conversely, AFMs suggest that lesions to the inverse model should prevent movement (motor commands) but need not thereby affect action observation. In IFMs, by contrast, loss of, or damage to, the integrated forward model would affect both the enacting of intentions and the recognition of those intentions in others.

Another promising conceptual testing ground for the two approaches might be provided by studies of conversation and (more generally) joint action. Imagine that agent A is trying to hold a conversation with agent B. IFMs claim that agent A relies on predictions issued by her own forward (generative) model to produce her speech, while also using that same model to simulate and predict agent B (who is busy simulating and predicting agent A). Both of the key activities of agent A (simulating agent B and producing speech) depend, if the IFM account is correct, on a single forward model, elements of which are repurposed (by alterations in precision weighting) to predict and understand agent B's actions.

In cases of joint action, however, this requires a very delicate use of that single resource. The same subelements of a single integrated forward model may now be required both to drive an action and to predict and understand that kind of action when it is being performed by another agent. To achieve this, it might seem that proprioceptive precision needs to be set high (to drive the action) and low (to simulate it) with respect to the same model elements at the same time. According to AFMs, however, the resource that handles the simulations is distinct from (although based on and interacting with) the resource that produces the agent's own actions. This seems to provide additional degrees of freedom that might be important for joint action.

We do not think it is impossible for an IFM account to accommodate such cases (they are, essentially, the special class of cases discussed at length, under the AFM umbrella, in [1]). One possibility may be to think of cases of joint action as setting up special contexts that recruit rapidly alternating, or very delicately interwoven, sets of precision weightings. Addressing and modelling such complex scenarios may help adjudicate between these two powerful yet surprisingly divergent prediction-based schemas.

Trading complexity

Despite their many similarities, the IFM and AFM accounts represent fundamentally different views of the shape and functioning of the human cognitive architecture. In common is the core emphasis on the need to predict our own upcoming sensory states. Such predictions can power learning, help finesse time delays, and enable a suite of potent capacities for motor imagination and simulation-based reasoning. In common too is the resulting emphasis on the learning and use of a forward (generative) model able to anticipate the sensory consequences of our own actions.

However, IFMs differ from AFMs in several important ways. In IFMs, motor commands are replaced by predictions about proprioceptive consequences (that implicitly minimise various energetic costs). The need for a distinct inverse model/optimal control calculation disappears and along with it the need for an efference copy of the motor command. Instead, IFMs posit a more complex (distributed) forward model mapping prior beliefs about desired trajectories to sensory consequences. The 'heavy lifting' that (in AFMs) required the use of an efference copy, inverse models, and optimal controllers now shifts to the acquisition and use of this more complex predictive (generative) model. Whether this is a worthwhile (or biologically realistic) trade-off remains to be seen.

References

1 Pickering, M.J. and Garrod, S. (2013) An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–392
2 Wilson, M. and Knoblich, G. (2005) The case for motor involvement in perceiving conspecifics. *Pscyhol. Bull.* 131, 460–473
3 Jordan, M.I. and Rumelhart, D.E. (1992) Forward models: supervised learning with a distal teacher. *Cogn. Sci.* 16, 307–354
4 Desmurget, M. and Grafton, S. (2000) Forward modeling allows feedback control for fast reaching movements. *Trends Cogn. Sci.* 4, 423–431
5 Grush, R. (2004) The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–442
6 Kawato, M. (1999) Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* 9, 718–727
7 Wolpert, D.M. and Kawato, M. (1998) Multiple paired forward and inverse models for motor control. *Neural Netw.* 11, 1317–1329
8 Franklin, D.W. and Wolpert, D.M. (2011) Computational mechanisms of sensorimotor control. *Neuron* 72, 425–442
9 Haruno, M. *et al.* (2003) Hierarchical mosaic for movement generation. *Int. Congr. Ser.* 1250, 575–590
10 Gregory, R.L. (1980) Perceptions as hypotheses. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 290, 181–197
11 Hinton, G.E. *et al.* (1995) The wake–sleep algorithm for unsupervised neural networks. *Science* 268, 1158–1160
12 Rao, R. and Ballard, D. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87
13 Lee, T.S. and Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1434–1448
14 Friston, K. (2005) A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 360, 815–836
15 Bastos, A.M. *et al.* (2012) Canonical microcircuits for predictive coding. *Neuron* 76, 695–711
16 Clark, A. (2013) Whatever Next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204
17 Friston, K. (2009) The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301
18 Friston, K.J. *et al.* (2011) Action and behavior: a free-energy formulation. *Biol. Cybern.* 102, 227–260
19 Adams, R.A. *et al.* (2013) Predictions not commands: active inference in the motor system. *Brain Struct. Funct.* 218, 611–643
20 James, W. (1890) *The Principles of Psychology*, Harvard University Press
21 Lotze, H. (1852) *Medicinische Psychologie oder Physiologie der Seele*, Weidmannsche
22 Kilner, J.M. *et al.* (2007) Predictive coding: an account of the mirror neuron system. *Cogn. Process.* 8, 159–166
23 Friston, K. *et al.* (2011) Action understanding and active inference. *Biol. Cybern.* 104, 137–160
24 Brown, H. *et al.* (2013) Active inference, sensory attenuation and illusions. *Cogn. Process.* 14, 411–427
25 Wolpert, D.M. *et al.* (1995) An internal model for sensorimotor integration. *Science* 269, 1880–1882
26 Friston, K. (2011) What is optimal about motor control? *Neuron* 72, 488–498
27 Wolpert, D.M. *et al.* (2003) A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 358, 593–602
28 Pickering, M.J. and Garrod, S. (2014) Self-, other-, and joint monitoring using forward models. *Front. Hum. Neurosci.* 8, 132
29 Fowler, C.A. *et al.* (2003) Rapid access to speech gestures in perception: evidence from choice and simple response time tasks. *J. Mem. Lang.* 49, 296–314
30 Altmann, G.T.M. and Kamide, Y. (1999) Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264
31 DeLong, K.A. *et al.* (2005) Probabilistic word pre-activation during comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121
32 Staub, A. and Clifton, C., Jr (2006) Syntactic prediction in language comprehension: evidence from either...or. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 425–436
33 Dikker, S. *et al.* (2009) Sensitivity to syntax in visual cortex. *Cognition* 110, 293–321
34 D'Ausilio, A. *et al.* (2011) Tongue corticospinal modulation during attended verbal stimuli: priming and coarticulation effects. *Neuropsychologia* 49, 3670–3676
35 Vigneswaran, G.L. *et al.* (2013) M1 corticospinal mirror neurons and their role in movement suppression during action observation. *Curr. Biol.* 23, 236–243
36 Ito, M. (2008) Control of mental activities by internal models in the cerebellum. *Nat. Rev. Neurosci.* 9, 304–313
37 Bastian, A. (2006) Learning to predict the future: the cerebellum adapts feedforward movement control. *Curr. Opin. Neurobiol.* 16, 645–649
38 Roth, M. *et al.* (2013) The cerebellum optimizes perceptual predictions about external sensory events. *Curr. Biol.* 23, 930–935
39 Koziol, L.F. *et al.* (2014) Consensus paper: the cerebellum's role in movement and cognition. *Cerebellum* 13, 151–177
40 Lesage, E. *et al.* (2012) Cerebellar rTMS disrupts predictive language processing. *Curr. Biol.* 22, R794–R795
41 Moberget, T. *et al.* (2014) Generalized role for the cerebellum in encoding internal models: evidence from semantic processing. *J. Neurosci.* 34, 2871–2878