Cognition 254 (2025) 105969

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Communicated priors tune the perception of control

George Blackburne[a,b], Chris D. Frith[c,d], Daniel Yon[a,e,f,*]

[a] School of Psychological Sciences, Birkbeck, University of London, UK
[b] Department of Experimental Psychology, University College London, UK
[c] Institute of Philosophy, University of London, UK
[d] Wellcome Centre for Human Neuroimaging, University College London, UK.
[e] Institut d'Études Avancées de Paris, France.
[f] Department of Psychology, Goldsmiths, University of London, UK.

ARTICLE INFO

Keywords:
Agency
Control
Expectation
Prediction
Communication

ABSTRACT

Action allows us to shape the world around us. But to act effectively we need to accurately sense what we can and cannot control. Classic theories across cognitive science suppose that this 'sense of agency' is constructed from the sensorimotor signals we experience as we interact with our surroundings. But these sensorimotor signals are inherently ambiguous, and can provide us with a distorted picture of what we can and cannot influence. Here we investigate one way that agents like us might overcome the inherent ambiguity of these signals: by combining noisy sensorimotor evidence with prior beliefs about control acquired through explicit communication with others. Using novel tools to measure and model control decisions, we find that explicit beliefs about the controllability of the environment alter both the sensitivity and bias of agentic choices; meaning that we are both better at detecting *and* more biased to feel control when we are told to expect it. These seemingly paradoxical effects on agentic choices can be captured by a computational model where expecting to be in control exaggerates the sensitivity or 'gain' of the mechanisms we use to detect our influence over our surroundings – making us increasingly sensitised to both true and illusory signs of agency. In combination, these results reveal a cognitive and computational mechanism that allows public communication about what we can and cannot influence to reshape our private sense of control.

1. Introduction

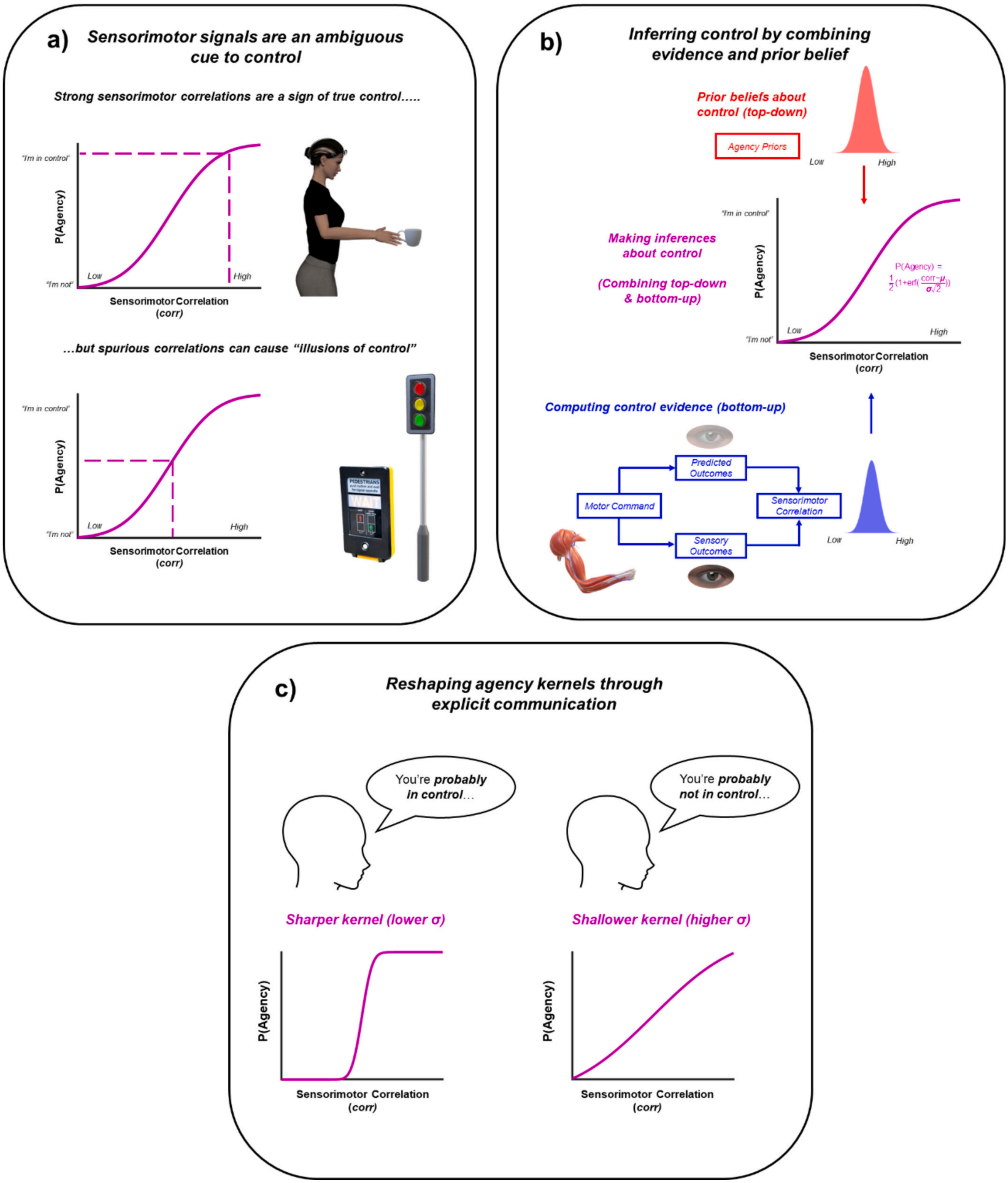
Human beings are agents. Through our actions we shape the world around us, manipulating our environments to suit our needs and desires. But to interact with our surroundings effectively, agents like us need to discover the causal structure of the environments we inhabit – learning which parts of the external world we can shape through action, and which parts lay outside our sphere of control.

Neuroscientists and psychologists interested in how we construct this 'sense of control' have tended to emphasise the role of direct experience (Haggard, 2017; Haggard & Chambon, 2012; Press, Thomas, & Yon, 2023). For instance, the comparator model of agency supposes that we formulate feelings of control by tracking directly experienced sensorimotor signals –inferring that we are in control of our environments when we experience strong contingencies between the actions we perform and the outcomes that ensue (Frith, 1987; Frith, Blakemore, & Wolpert, 2000; **see** Fig. 1). A rich vein of evidence supports the idea that we are exquisitely sensitive to matches and mismatches between action and outcome as we judge what we can and cannot control, and even subtle spatiotemporal disturbances can dramatically disrupt feelings of agency (Blakemore, Frith, & Wolpert, 1999; Penton, Wang, Catmur, & Bird, 2022; Perrykkad, Lawson, Jamadar, & Hohwy, 2021).

However, relying on sensorimotor signals alone to construct feelings of control presents brains like ours with a problem. The ambiguous, noise-ridden signals emanating from our sensorimotor circuits can provide a misleading picture of whether we are truly in control or not. For example, felt correlations between action and outcome are not always a signal of true agency (Alloy & Abramson, 1979), and we can readily experience 'illusions of control' over objectively uncontrollable objects when their behaviour spuriously aligns with our own (Yon, Bunce, & Press, 2020). For instance, we might press a 'placebo button' at a pedestrian crossing everyday on our morning commute, experiencing a sense that our actions stop the traffic, while unbeknownst to us the lights are programmed by a timer and the button is entirely inert (Luo, 2004).

* Corresponding author at: School of Psychological Sciences, Birkbeck, University of London, UK.
E-mail address: d.yon@bbk.ac.uk (D. Yon).

https://doi.org/10.1016/j.cognition.2024.105969
Received 8 April 2024; Received in revised form 23 September 2024; Accepted 25 September 2024
Available online 4 October 2024



(caption on next page)

Fig. 1. Estimating control with sensorimotor signals and prior beliefs. a) Classic theories suggest that we track sensorimotor signals to infer what we can and cannot control. For instance, when we are controlling our bodies (e.g., when we reach for a coffee cup) we will usually experience tight correlations between action and outcome (*top*). However, these signals can also be misleading. For instance, we can experience spurious correlations between our actions and outcomes (e.g., when we press 'placebo buttons') and these correlations can lead us to experience illusory feelings of control (*bottom*). **b)** One way to improve inferences about agency is to combine sensorimotor experience with prior beliefs about control. Here we propose a model where agents track the correlations between actions and outcomes to generate 'bottom up' evidence about what they can and cannot control (*blue*) – as in classic comparator accounts. This sensorimotor evidence is mapped onto feelings of control by an *agency kernel* (*purple*) – with stronger felt correlations mapping onto stronger feelings of control. However, our model assumes that the shape of this kernel is also influenced 'top-down' by beliefs that agents hold about the probability of control (*red*). **c)** This model offers a way of conceptualising how explicit communication about control might alter inferences of agency. In particular, if we are explicitly told to expect a high level of control (*left*) this prior belief could 'sharpen' an agent's decision kernel – such that modest-to-strong correlations are mapped onto strong inferences of agency. In contrast, if we are explicitly told to expect a low level of control *(right)* this prior belief could 'dampen' an agent's decision kernel – meaning that modest correlations fail to 'ignite' feelings of agency. This kind of biasing would be adaptive, since if we expect *not* to be in control, modest correlations between action and outcome are likely to be spurious (cf. placebo buttons, Luo, 2004). NB: This kind of biasing is indeed found in the present work (see *Computational Modelling)*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Since direct sensorimotor signals can be misleading, models of our own agency crafted from these signals have the potential to mislead us too (Fig. 1).

Bayesian models of perception are preoccupied with a very similar problem: how can we construct an accurate perceptual image of the world around us given the noise and ambiguity that corrupts the sensory signals we sample? (de Lange, Heilbron, & Kok, 2018; Summerfield & de Lange, 2014). Bayesian models of perception propose that the inherent ambiguity of our sensory signals can be overcome by combining noisy sensory evidence with prior beliefs about what the world is likely to contain (Yuille & Kersten, 2006). Biasing our perception in line with our expectations will tend to render a more veridical picture of the outside world, because what we expect to occur (by definition) usually does.

These prior beliefs could emerge slowly, as we integrate our direct experiences through trial-and-error to learn about what we can and can't control. But direct experience is not the only window humans have onto our environments. Other domains of cognitive science have increasingly come to appreciate that while it is often very challenging to discover the causal structure of our environments from direct experience alone, the ambiguity inherent in noisy experiences can be finessed by explicit communication. For example, work in reinforcement learning has found that human agents virtually never discover the structure of some reward environments when left to their own devices, but can rapidly become 'model-based' learners when given explicit instructions about how the learning environment is arranged (Castro-Rodrigues et al., 2022).

Explicit communication with others could also allow us to form more accurate models of our own agency. By sharing beliefs about action and influence, we may discover what we can and cannot control in ways that would be challenging based on direct experience alone. For instance, it might be hard to discover through direct experience that our traffic crossing is wired up to a placebo button – but if somebody tells us the button has been disconnected for years, we can immediately stop pushing it.

It seems intuitive that our sociocultural world is a rich source of information about what we can control and what we are responsible for. Since antiquity philosophers have speculated that explicitly teaching children about the consequences of their actions is central to endowing them with a sense of agency (Bobzien, 2006), and human societies codify expectations about responsibility, credit and blame in legal texts and in social norms. But it remains mysterious how beliefs about control acquired through this kind of explicit communication become integrated into the sense of control that we feel.

The influence of instructions has been formulated in more Bayesian terms by other cognitive scientists, suggesting that explicit communication can directly inject prior beliefs into agents' heads without the need for direct learning (Lindström, Golkar, Jangard, Tobler, & Olsson, 2019). But how does explicit communication allow us to directly inject prior beliefs from one head into another? And how could these communicated priors shape feelings of control?

One way this predictive biasing can be achieved in perceptual systems is through 'sharpening': reshaping patterns of sensory gain so that our sensory systems are particularly sensitised to the signals that we expect to occur (de Lange et al., 2018; Press & Yon, 2019). In neural circuits, this kind of sharpening creates higher fidelity representations of sensory events when they conform with prior expectations (Kok, Jehee, & de Lange, 2012; Yon et al., 2023; Yon, Gilbert, de Lange, & Press, 2018), and in perceptual terms, exaggerating sensory gain in this way renders agents better able to detect the perceptual events they expect to occur. At the same time, observers become more likely to 'hallucinate' the events they expect even when they are truly absent – as manipulating the gain on sensory circuits in this way also makes it easier to mistake spurious noise for genuine signal (Wyart, Nobre, & Summerfield, 2012).

In this paper, we investigated whether explicit prior beliefs lead to a similar kind of predictive biasing in the sense of control. We present a new model which casts the detection of control as analogous to a perceptual decision, where evidence sampled from sensorimotor circuits is used to make inferences about what we can and cannot control – just as classic comparator models of agency suggest (Frith, 1987; Frith et al., 2000). Crucially though, our theory supposes that prior beliefs can percolate into these inferences, controlling how experienced sensorimotor signals are mapped onto inferences about controllability (see Fig. 1). In this way, our model allows agents to filter their perception of noisy sensorimotor signals through prior beliefs about what they can and cannot control – becoming more sensitive to control-like evidence when we believe we are likely to be causal agents, and becoming less sensitive to this same evidence when we believe we are not.

Here we present a series of experiments testing this hypothesis, using recently developed motion-tracking and modelling tools to measure how explicit prior knowledge alters the sensitivity and bias of agentic decisions (Yon et al., 2020). These experiments reveal that expecting a high level of control exaggerates both the sensitivity and bias of agency judgements – such that we become better able to distinguish control from its absence, but also more likely to experience illusions of controlling things we cannot. Crucially, a computational model implementing our new theory can explain the emergence of these seemingly incompatible effects – by assuming that explicit beliefs about our causal power reshape our sensitivity to the noisy sensorimotor evidence we use to judge what we can and cannot control. In so doing, our results reveal a cognitive and computational mechanism that allows public communication with other people to reshape private feelings of agency.

2. Methods

2.1. Participants

We conducted two experiments. Experiment 2 was a replication of Experiment 1, and procedural details were identical. Both were preregistered on AsPredicted.org (Exp 1: https://aspredicted.org/9jq7s.pdf, Exp 2: https://aspredicted.org/d7wj5.pdf). Twenty-four separate participants took part in each experiment. A further ten participants were tested (five each in both Experiments 1 and 2) but excluded from the final sample, as they failed to complete at least 200 valid trials of the main task (see below). This sample size and this exclusion criterion were pre-registered for both studies. Experimental methods and task-based statistical analyses were also pre-registered, but computational modelling analyses (see below) were not.

All participants were healthy adults, fluent English speakers, resident in the United Kingdom, with normal or corrected vision and no current neurological or psychiatric illness. Participants had a mean age of 32.6 ± 8.6 years, including 31 of whom identified as female, 16 as male, and 1 as outside the gender binary. Participants were recruited via Prolific, completed the experimental task online, and were paid a small honorarium for taking part. All procedures were approved by local ethics committees.

2.2. Experimental task

Participants completed a 'control detection task' closely modelled on techniques introduced by Yon et al. (2020) – implemented in PsychoPy (Fig. 2). On each trial, participants executed a circular mouse movement and observed a small dot move around a 'bagel' on screen (2000 ms). On *control* trials, the movement of the dot cursor was yoked to the participant's actions, while on *no control* trials the cursor followed a preprogrammed trajectory, repeating the path of a previous movement. On both trial types, a small amount of random jitter was added to the drawn trajectory. At the end of each trial, participants had to report whether they controlled the trajectory of the moving dot ('yes' or 'no') and their confidence in this decision ('confident' or 'guess') on a four-point scale. There was no enforced time limit on decisions, and the response screen remained until a choice was registered.

Over different blocks of the experiment, participants received explicit instructions about the levels of control to expect. In *expect high* blocks, participants were explicitly informed (by on-screen instructions) that this was a 'high control block' and told "*on most trials, you will control the dot's path*" In contrast, in *expect low* blocks, participants were instructed that the upcoming was a "low control block" and were told "*on most trials, you will not control the dot's path*". These instructions were valid, and in *expect high* blocks 70 % of trials were *control* trials and 30 % were *no control* trials (vice versa for *expect low* blocks). Comparing participant decisions between these different blocks thus allows us to examine how probabilistic expectations about the controllability of the environment influence judgements of control.

The main experiment comprised 240 trials. This was divided into four expectation blocks – two *expect high,* two *expect low* – of 60 trials each. Block expectation alternated within participants (e.g., *high-low-high-low*) and block order was counterbalanced across participants. Participants also completed short practice sessions before the main blocks where they practiced moving the dot around the ring (10 trials), and practiced making control judgements (10 trials), before beginning the main experiment. Breaks were offered every 20 trials.

2.3. Inclusion criteria and statistical analysis

The decision screen at the end of each trial only appeared when movements were successful (i.e., the dot cursor moved counter clockwise around the ring, and stayed within the bounds of the bagel). If participants made movement errors, the decision screen was replaced with warning feedback ("Stay in the ring", "Move round the ring counter-clockwise") and no decision was recorded. These trials were not included in subsequent analyses, and participants who contributed fewer than 200 valid trials across the experiment were excluded entirely, as small trial counts limit the sensitivity of subsequent analyses. This exclusion criterion was pre-registered for both experiments.

We used participants' choices to compute signal detection theoretic measures of sensitivity (d') and bias (c) (Green & Swets, 1974). Both d' and c are calculated from hit rates (HR) and false alarm rates (FAR) where $d' = z(HR) - z(FR)$ and $c = -0.5(z(HR) + z(FR))$. Thus, higher values of d' indicate a superior ability to objectively distinguish *control*

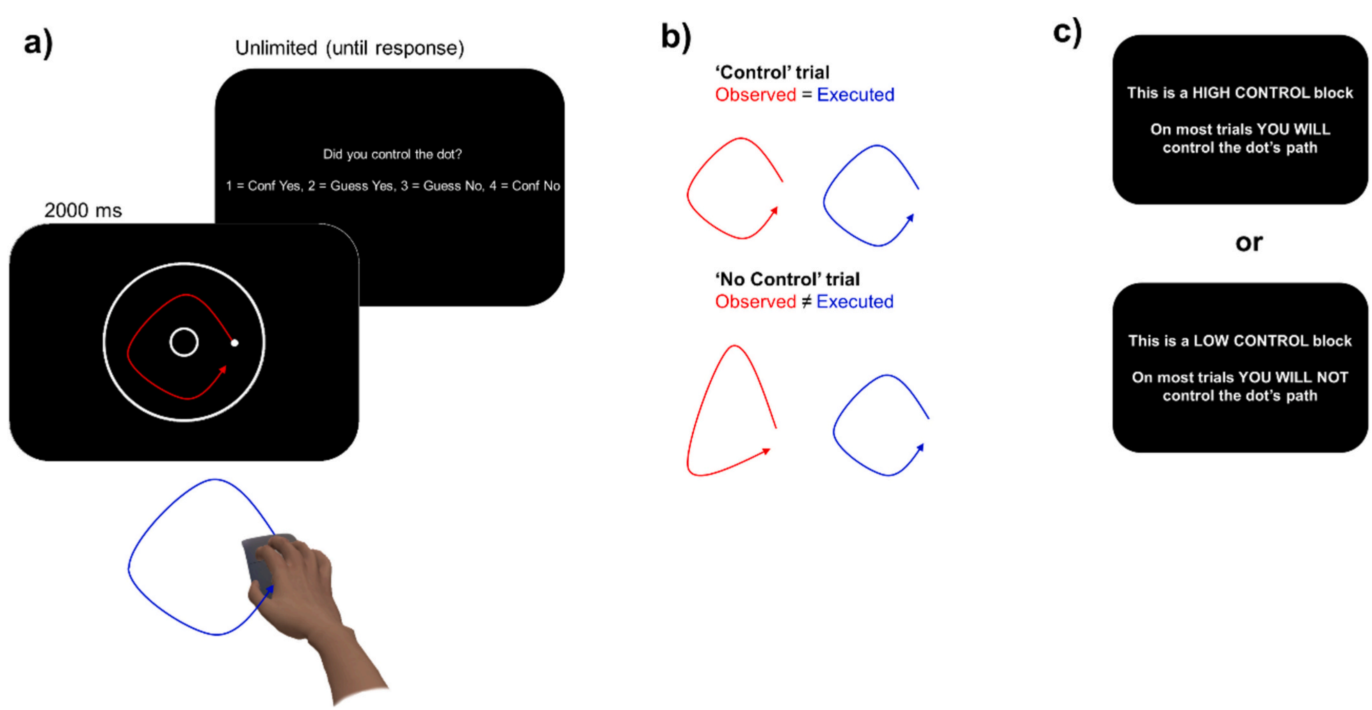


Fig. 2. Control detection task. a) Participants completed a task where they made a counter-clockwise hand movement and observed similar movements of an onscreen cursor. They were asked to judge whether they controlled the path of the dot or not, and report their confidence (*confident* or *guess*). **b)** Sometimes these onscreen movements were entirely yoked to the participant's movements (*control* trials) and other times the cursor followed a trajectory from a previous trial (*no control* trials). **C)** We manipulated participant's expectations about the probability of control in upcoming blocks with explicit cues, allowing us to determine how prior beliefs about controllability influence the sensitivity and bias of agency judgements.

from *no control* trials. Lower values of *c* indicate more liberal responding – such that agents are more likely to report 'control' over the dot, whether they truly influenced it or not.

We also used participants' confidence ratings to compute measures of metacognitive sensitivity (meta-*d'*) and bias (confidence level; Fleming & Lau, 2014; Maniscalco & Lau, 2012). Higher values of meta-*d'* indicate more accurate introspection, such that agents tend to feel more confident in their control judgements where their decisions are

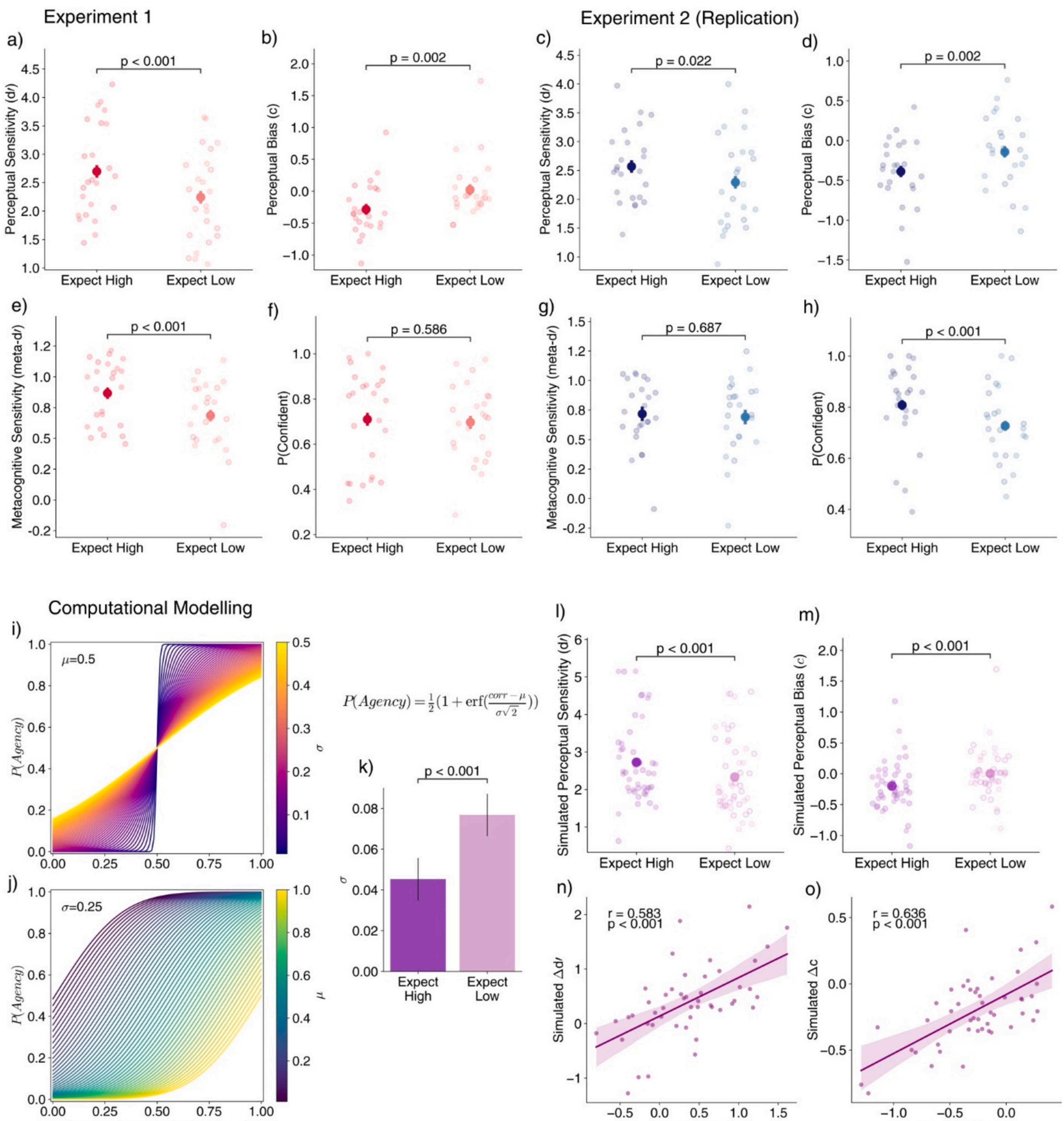


Fig. 3. Empirical results and computational modelling. Signal detection theoretic analyses revealed that expecting high (rather than low) levels of control, exaggerated both sensitivity to control (d' – panels *a* and *c*) and bias to feel control (c – panels *b* and *d*) in both experiments. Larger dots indicate means with bars displaying 95 % within-subjects confidence intervals, reflecting the magnitude of differences between conditions(Loftus & Masson, 1994). Effects of these expectations measures of metacognition were present in each experiment, but inconsistent across studies (panels *e-h*). We modelled feelings of control as arising from 'agency kernels' which transform sampled sensorimotor evidence into inferences of agency. Our data was best fit by a model where expectations about control change the slope (σ – sigma, panel i) but not bias (μ – mu, panel j) of these kernels. Analysing the best-fitting sigma parameters found that kernels had a more exaggerated slope when agents expected higher levels of control (panel *k*) and simulating data from the model recreated the empirical effects observed – i.e., when the model performs the task it is also more sensitive to control (panel *l*), and biased to hallucinate control (panel *m*), when higher levels of control are expected.

correct, and less confident when they err. In contrast, confidence level simply reflects metacognitive bias – and is calculated as the proportion of trials where participants report high (rather than low) confidence in their choices. Meta-d values were calculated using single-subject Bayesian estimation with the HMetad package (Fleming, 2017).

To evaluate how expectations alter these objective and subjective dimensions of the sense of agency, we compared these measures between *expect high control* and *expect low control* trials. Participants were considered statistical outliers if their individual effect scores (i.e., difference between *expect high* and *expect low* trials) was >2.5 SDs from the sample mean. For the purposes of inferential tests, we winsorised these effect scores to be equivalent of +/− 2.5 SDs. This planned adjustment was pre-registered for both experiments, and was applied to one participant in the confidence level comparison in Experiment 1, and to two participants who were outliers in the comparison of sigma values in the computational modelling analyses – though in no case did this change any statistical patterns observed.

3. Results

Across both experiments, expectations altered the decisions agents made about what they could and could not control. Agents showed greater perceptual sensitivity (d') in *Expect High Control* blocks than *Expect Low Control* blocks – Exp 1: $t_{23} = 3.995$, $p < .001$, $dz = 0.816$, Exp 2: $t_{23} = 2.448$, $p = .022$, $dz = 0.500$ (see Fig. 3a & 3c) – indicating a better ability to objectively distinguish true control from it's absence. But at the same time, both experiments revealed that agents were also biased (c) to report control more often when high levels of control were expected - Exp 1: $t_{23} = 3.525$, $p = .002$, $dz = 0.720$, Exp 2: $t_{23} = 3.393$, $p = .002$, $dz = 0.693$ (see Fig. 3b & 3d) – consistent with the idea that expecting to be in control causes agents to more readily hallucinate control over objectively uncontrollable events.

We also conducted analyses of agentic metacognition, and how introspection about agency might be influenced by these expectations too – but patterns were inconsistent across studies. In Experiment 1, agents showed superior metacognitive insight (meta-d') into their agency judgements when control was expected -: $t_{23} = 3.801$, $p < .001$, $dz = 0.776$ (see Fig. 3e) – but this effect was not replicated in Experiment 2 – $p = .687$ (see Fig. 3g). Likewise, in Experiment 2 agents reported generally higher confidence in their control judgements when control was expected - $t_{23} = 4.094$, $p < .001$, $dz = 0.836$ (See Fig. 3h) - but this effect was not found in Experiment 1 – $p = .662$ (see Fig. 3f).

Traditionally in signal detection theory, shifts in sensitivity and bias have been thought to be governed by separate processes. For instance, changes in sensitivity might arise because of improvements in perceptual precision, while shifts in bias might reflect strategic changes in decision making without any changes in what observers perceive (Swets, Tanner, & Birdsall, 1961). But while it may seem counterintuitive, a range of studies have found that top-down shifts in sensory gain (e.g., through attention) can simultaneously improve detection sensitivity and generate more liberal detection biases (Cheadle, Egner, Wyart, Wu, & Summerfield, 2015; Wyart et al., 2012). Indeed, recently we have shown that 'illusions' of control in agency tasks could arise if otherwise unbiased agents are especially sensitive to spurious correlations between action and outcome – mistaking this 'noise' for 'signal' (Yon et al., 2020). Thus – in principle – it is possible that a single mechanism could explain why agents become both more sensitive to control *and* more biased to report agency when control is expected.

3.1. Computational modelling

We used computational modelling to investigate this possibility. Our model assumes that agents begin to judge whether they are in control of their surroundings by tracking the sensorimotor correspondence between actions they perform and outcomes that ensue – just as in comparator models of agency (Frith, 1987; Frith et al., 2000). However, our model then assumes that the sensorimotor evidence generated by this process is passed through an *agency kernel,* which translates sensorimotor signals into feelings of control. The shape of this kernel determines which kinds of sampled sensorimotor signals lead to inferences of control, and how strong sensorimotor evidence must be before feelings of agency are 'ignited'.

Thus, our model generates feelings of control in two steps. First, the model assumes that agents have access to moment-by-moment information about the actions they are performing in the task, and the trajectory of the moving dot that appears on-screen. This allows agents, on each trial, to compute a *sensorimotor correlation* between action and outcome – which we take as the two-dimensional correlation between the executed and observed motion trajectories (Yon et al., 2020). Higher values of this *sensorimotor correlation* denote a strong coupling between actions and outcomes – where 1 is perfect contingency and 0 is complete independence. Naturally, correlations tend to be higher when agents are truly in control than when they are truly not – but substantial trial-by-trial variability means that agents experience a wide range of action-outcome correlations throughout the task.

In the next step, the model passes this *sensorimotor correlation* value through an *agency kernel.* This kernel is modelled as a cumulative gaussian function mapping sampled correlations onto inferences of control. The shape of this function is controlled by two parameters *sigma* - σ and *mu* - μ. Sigma controls the slope or sensitivity of the function – with lower values of sigma leading to sharper, exaggerated kernels and higher values of sigma leading to a shallower, noisier mapping between sampled correlations and decisions (Fig. 3i). In contrast, mu controls the bias of the function: as mu decreases, it becomes more likely that agents feel a sense of agency over weaker and weaker correlations (i.e., becoming more liberal in detecting control), while conversely as mu increases agents become less likely to feel a sense of control even when experienced correlations are relatively high (i.e., detection becomes more conservative; see Fig. 3j).

We explored whether explicit prior beliefs alter the way agents translate experienced sensorimotor signals into feelings of control. In our model, this reshaping could happen in one of two ways. Prior expectations could alter the slope the agency kernel (σ), such that agents are more sensitive to the distinction between strong and weak correlations when high levels of control are expected. Alternatively, expectations could induce a general bias that shifts the decision kernel (μ), such that agents are generally more biased to report control when it is expected, with the overall sensitivity or 'slope' of the kernel remaining unchanged.

Intuitively, one might expect that *both* kinds of reshaping would be needed to account for our empirical findings – where expectations change both perceptual sensitivity (d') and bias (c). From the point of view of traditional signal detection theory, changes in the actual or expected probability of a signal (in our case, 'control') should primarily influence measures of bias (like c) – as the optimal criterion is affected by the relative probabilities of 'signal' and 'noise' trials, and this can lead to symmetric changes in both hit rates and false alarm rates (Swets et al., 1961). In contrast, measures of sensitivity (like d') are traditionally thought of as bias-free, and largely independent of these changes in criterion (Swets et al., 1961).

However, it may be possible to create both enhanced sensitivity and exaggerated illusions of control by exaggerating only the slope or sensitivity of the decision kernel (σ) – rather than introducing tonic biases or 'criterion shifts' in decision making. This is because a sharpened kernel is both more likely to generate feelings of agency when observers experience a moderate correlation between action and outcome – which increases an agent's ability to detect instances of true control ('hits') – and also more likely to hallucinate an illusory sense of control when correlations are spurious ('false alarms').

To compare these possibilities, we modelled kernels to participant agency decisions, allowing the kernels to vary in three different ways. In the *slope* model, we fit separate slope parameters to the kernel for each expectation condition – allowing, for example, kernels to have steeper slopes on *Expect High* than *Expect Low* trials – but estimated a single value of μ – meaning that the bias of kernel did not change as a function of expectations. The second *bias* model was the opposite to *slope* model, meaning that we fit separate bias parameters *Expect High* and *Expect Low* trials but estimated only a single value of σ – meaning that the slope of the kernel did not change as a function of expectations. In the final *slope* + *bias* model, we allowed both σ and μ of the kernel to varying according to expectations. In both cases, best fitting parameter values of the gaussian kernels were found using a Python implementation of the Levenberg-Marquardt algorithm for non-linear least squares fitting.

First, we examined whether each of these three models could reproduce the empirical results obtained in Experiments 1 and 2, since a failure to capture key results effectively 'falsifies' a model as a candidate explanation (Palminteri, Wyart, & Koechlin, 2017). This revealed that while the *slope* and *slope* + *bias* models could reproduce shifts in d' and c according to expectations, the *bias* model could not reproduce shifts in d' (see *Supplementary Material)* – ruling out this model as a potential explanation.

We adjudicated between the remaining *slope* and *slope* + *bias* models by comparing model fits, computing R^2 values that reflect the error between participants' empirical decisions and decisions simulated from their best fitting kernel. This analysis revealed that decisions simulated from the *slope* + *bias* model had slightly better fit ($R^2 = 0.81$) than the *slope* model ($R^2 = 0.80$) - though the benefit of including these additional μ parameters was marginal. Indeed, analysing μ values in the *slope* + *bias* model found that values estimated for $\mu_{\text{Expect High}}$ and $\mu_{\text{Expect Low}}$ did not significantly differ ($p = .186$). This suggests that, even when we allow the μ parameter (bias) of the agency kernels to vary with expectations about control, the parameters do not consistently differ. Given this, we concluded that the simpler *slope* model provided the better account of our empirical data. And indeed, analysing parameter values of the *slope* model revealed that sigma values were significantly lower on *Expect High* than *Expect Low* trials – $t_{47} = 3.807$, $p < .001$, $dz = 0.55$ – indicating that decision kernels were indeed sharpened when agents expect to be control. (NB: The same difference between $\sigma_{\text{Expect High}}$ and $\sigma_{\text{Expect Low}}$ parameters was also found in the *slope* + *bias* model – $p = .003$).

To further investigate whether the *slope* model could reproduce our empirical results, we simulated decisions from these agency kernels and analysed these in the same way as the real data. In effect, these simulations involve having the model perform the same task as the participants – exposing the model to the sensorimotor correlation a participant experienced on a given trial, and passing it through the parametrised kernel to create a simulated decision about whether the cursor on that trial was controllable or not. By repeating this process for every trial in the experiment, we can create simulated hit rates and false alarm rates that can be analysed to yield signal detection theoretic measures like *d'* and *c* in the same way as the real empirical data.

As shown in **Fig. 3l** and **m,** data simulated from the model in this way recapitulates our empirical results: the model is more sensitive to objective facts about control (higher *d'*) when control is expected –- $t_{47} = 4.048$, $p < .001$, $dz = 0.584$ – but also biased to experience more 'illusions of control' when control is expected too - $t_{47} = 5.170$, $p < .001$, $dz = 0.746$. As well as reproducing these effects at the group level, correlational analyses also show the model does a good job of capturing variability in effects between participants - with a strong correlation between empirical effects of expectation on of d' and c for each participant and the effect simulated for them by the model – $r_{48} = 0.583$, $p < .001$ and $r_{48} = 0.636$, $p < .001$, respectively (see Fig. 3n & o). In combination, these results lend support to our model where explicit priors alter the sensitivity and bias of agency judgements by reshaping the slope or gain of the function that transforms sampled sensorimotor evidence into feelings of control.

4. Discussion

To effectively interact with the world around us we need an accurate sense of control, but the sensorimotor signals we use to judge our agency over our surroundings are riven with ambiguity and noise. One way we might overcome this ambiguity is explicitly communicating with others – forming prior beliefs that help to optimise our inferences about what we can and cannot influence.

Here, across two experiments, we find that explicit instructions about what we can and cannot control alter both the sensitivity and bias of agentic judgements. When we are told to expect high levels of control over the environment we become better able to distinguish control from its absence, but also biased to hallucinate a sense of control over uncontrollable outcomes. Computational modelling revealed that this seemingly counterintuitive combination of effects could be captured by a model where priors only change a single parameter of the decision process – the slope or gain of the function mapping sensorimotor signals onto inferences of control.

These results reveal how explicit communication with others can help to refine our sense of control. Exaggerating our sensitivity to sensorimotor signs of control when we expect a high degree of agency is adaptive, as it enables us to more readily detect signs of our influence over events that we probably do control. Likewise, dampening our sensitivity to control-like evidence when control is more improbable is also adaptive, as if we have strong prior beliefs that we are not influencing our surroundings, experienced correlations are likely to be spurious coincidences rather than signs of genuine agency (Yon et al., 2020).

Our combination of signal detection methods and computational modelling provide evidence that explicit beliefs about control alter the *sensitivity* of agentic judgements. Some past work has explored how implicit measures of agentic experience – like intentional binding illusions or sensory attenuation – change when agents believe they are (not) authors of an outcome (Desantis, Roussel, & Waszak, 2011; Desantis, Weiss, Schütz-Bosbach, & Waszak, 2012; Dogge, Schaap, Custers, Wegner, & Aarts, 2012). But by design, such measures of agentic experience are indirect, making it difficult to probe how prior beliefs are integrated into the computations our brains use to furnish the sense of control. In contrast, our psychophysical methods and modelling approach allow us to diagnose how prior beliefs alter experiences of control – by altering our sensitivity to control-like evidence depending on whether we believe we are agents or not.

It may seem initially counterintuitive that this kind of 'gain control' mechanism can simultaneously improve the accuracy of control judgements (d') and lead to exaggerated biases (c) in the control that we feel. Classic accounts like signal detection theory have tended to assume that separate underlying mechanisms are responsible for a perceiver's sensitivity and their bias (Swets et al., 1961). However, our findings join an extant literature which shows that explicit instructions – such as instructions to attend – can simultaneously improve perceptual sensitivity and engender perceptual biases (Hawkins et al., 1990). One way that perceptual scientists have explained these joint effects is to assume that top-down gain control exaggerates a perceiver's sensitivity to signals and signal-like noise – rendering us more sensitive to true signals but also more prone to overinterpret spurious noise (Wyart et al., 2012). Here we have found an analogous mechanism can account for the ways that explicit communication about controllability sharpens or dampens our sensitivity to genuine and spurious signals of agency over our surroundings – altering both sensitivity and bias in our sense of control.

Here we have used a signal detection approach to study feelings of control (Yon et al., 2020), building on the classic idea from perception research that these tools allow us to separately quantify an individual's sensitivity and bias. However, as noted above, a fertile approach in sense of agency research has also contrasted methods that involve explicit judgements of control (like the present study) with implicit markers of agentic feelings – like the intentional binding illusion (Haggard, Clark, &

Kalogeras, 2002). Theorists have likewise fruitfully distinguished between non-conceptual *feelings* of control and conceptual *judgements* of control (Synofzik, Vosgerau, & Newen, 2008), which may be particularly important given that judgements and feelings of control can track similar quantities (Dickinson, Shanks, & Evenden, 1984; Moore, Lagnado, Deal, & Haggard, 2009) but also sometimes decouple (Dewey & Knoblich, 2014). As we have not used any implicit measures, it is possible in principle that the influences of control priors we describe here primarily reflect changes in *beliefs* about control rather than changes in the control that agents *experience*. However, our results do reveal that priors about controllability don't simply induce generic biases in control judgement, but deeper changes in our sensitivity to sensorimotor evidence. Such effects are more redolent of changes seen in low-level perception than high-level judgement (Wyart et al., 2012), and our theory and model predict that prior beliefs about control should shape feelings and judgements in equivalent ways. Testing this model prediction is a promising avenue for future work.

In any case, these findings reveal a cognitive and computational mechanism that can explain how our background beliefs about agency, responsibility and control alter agentic feelings. For example, when somebody else coerces us into generating a particular outcome (e.g., pressing a button to deliver an electric shock), feelings of agency over and neural responses to the outcome become attenuated (Caspar, Christensen, Cleeremans, & Haggard, 2016), even though objective sensorimotor contingencies are unchanged by coercion (e.g., it remains our button press that causes the shock to occur). These kinds of distortions in the sense of agency are somewhat mysterious if feelings of control simply depend on direct sensorimotor experience, but can be readily accommodated by our model, where prior beliefs alter our sensitivity to control-like evidence. For example, if sociocultural learning means that we *believe* that we have less influence when we are only following orders, such priors may 'dampen' our agency kernels, rendering us less sensitive to evidence that we remain truly in control.

Here we have drawn inspiration from Bayesian models of perception, which assume that our subjective experience of the sensory world unfolds as we combine incoming evidence with prior beliefs. A key component of Bayesian models is the notion of 'precision' (Yon & Frith, 2021). These accounts suggest that, as we make inferences about the world around us, the weight we afford to different sources of information is proportional to their reliability or uncertainty. For instance, in vision perceivers rely more on prior expectations when progressively more noise is added to incoming signals i.e., as the sensory evidence becomes relatively less 'precise' (Olkkonen, McCarthy, & Allred, 2014). If our sense of control is constructed in a similarly Bayesian way (Moore & Fletcher, 2012) – depending on similar estimates of uncertainty (Constant, Salomon, & Filevich, 2022) - we ought to find that the influence of prior beliefs on feelings of agency is exaggerated when sensorimotor signals are most ambiguous – leaning more on prior knowledge when the evidence provided by our senses is most uncertain. For instance, humans (Bloom, Venard, Harden, & Seetharaman, 2007) and other animals (Skinner, 1948) can experience 'superstitious associations' between actions and outcomes in the absence of any true causal relation - particularly when there are many possible actions agents might take and the timing between action and outcome becomes more delayed or more variable. These situations can be usefully thought of as cases where our sensorimotor experiences are objectively more uncertain or 'imprecise'.

However, from a Bayesian point of view, this objective ambiguity and variability in our sensorimotor experience is not the entire story. Recent Bayesian models suggest that the balance between expectations and evidence doesn't only depend on the objective uncertainty in the different sources of information being monitored, but on what agents *believe* about their precision (Friston, 2018; Yon & Frith, 2021). For instance, in perceptual decision making, observers come to have exaggerated sense of confidence in their percepts when they *expect* incoming sensory signals to be more reliable, even if objective perceptual precision remains unchanged (Olawole-Scott & Yon, 2023). If inferences about our own agency unfold in a similar way, we would expect that the balance agents strike between incoming evidence and prior beliefs will depend on their *beliefs* about the reliability of sensorimotor signals and prior expectations.

Thinking about how these weights are calibrated could allow us to develop and test new hypotheses about how disturbances in the sense of control arise and persist. For instance, in psychotic illnesses like schizophrenia patients can experience delusions that their bodies are controlled by an alien force, or that they possess grandiose powers to influence features of their surroundings which they – objectively speaking – cannot (Isham et al., 2021; Mellor, 1970). In controlled experiments, patients with psychosis show reduced sensitivity to sensorimotor evidence about what they can and cannot influence (Blakemore, Smith, Steel, Johnstone, & Frith, 2000), and poorer insight into the reliability of the sensorimotor signals they sample (Krugwasser, Stern, Faivre, Harel, & Salomon, 2022). A key clinical feature of these delusions is that these strange beliefs about action are resistant to revision when challenged by other people. In the context of our model, we could speculate that such anomalous experiences arise because patients do not give appropriate weight to explicitly communicated beliefs – meaning that noisy sensorimotor evidence is not appropriately 'filtered' through the expectations provided by other people. This way of thinking may thus explain not only how social communication shapes feelings of control in general, but also how some of us become detached from social influence (Baptista, Jacquet, Sidarus, Cohen, & Chambon, 2021).

Our findings suggest that it is possible to implant explicit models about agency and control directly into the minds of others through explicit communication. Though we have focused on how these models shape the subjective experience of control, inheriting models about our agency from others could have other consequences for cognition about control (Yon & Corlett, 2022). For example, entertaining causal models about our own agency makes it possible to engage in "hypothesis testing" behaviour that probes our level of control over the environment (Wen et al., 2020) and allows us to become sensitive to counterfactual information about what we can and cannot influence (Kulakova, Khalighinejad, & Haggard, 2017). Future work may probe whether explicit communication about what we can control also shapes these features of agentic behaviour too.

Our capacity for goal-directed action hinges on our ability to track what we can control. But the sensorimotor signals served up by direct experience are inherently ambiguous, and can provide us with an unrealistic picture of what we can and cannot influence. Here we have revealed a cognitive and computational mechanism that allows agents like us to overcome this ambiguity, by allowing prior beliefs about control to retune the mechanisms we use to construct agentic experiences. This mechanism in turn reveals one way that the ideas we inherit from others about action and control can shape our personal sense of agency.

CRediT authorship contribution statement

George Blackburne: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Chris D. Frith:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Daniel Yon:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

None

Data availability

Data is available at https://osf.io/365p7/.

Acknowledgements

This work was supported by the Wellcome Trust. DY completed work on this paper while a Visiting Fellow at the Paris Institute for Advanced Study, funded by FIAS and the Marie Sklodowska-Curie Actions COFUND Programme of the European Commission.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2024.105969.

References

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General, 108*(4), 441–485. https://doi.org/10.1037/0096-3445.108.4.441

Baptista, A., Jacquet, P. O., Sidarus, N., Cohen, D., & Chambon, V. (2021). *Susceptibility of agency judgments to social influence*.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience, 11*(5), 551–559. https://doi.org/10.1162/089892999563607

Blakemore, S. J., Smith, J., Steel, R., Johnstone, C. E., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological Medicine, 30*(5), 1131–1139. https://doi.org/10.1017/s0033291799002676

Bloom, C. M., Venard, J., Harden, M., & Seetharaman, S. (2007). Non-contingent positive and negative reinforcement schedules of supersitious behaviors. *Behavioural Processes, 75*(1), 8–13. https://doi.org/10.1016/j.beproc.2007.02.010

Bobzien, S. (2006). *Moral responsibility and moral development in Epicurus' philosophy*. https://philarchive.org/rec/BOBCHA.

Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. (2016). Coercion changes the sense of Agency in the Human Brain. *Current Biology: CB, 26*(5), 585–592. https://doi.org/10.1016/j.cub.2015.12.067

Castro-Rodrigues, P., Akam, T., Snorasson, I., Camacho, M., Paixão, V., Maia, A., … Oliveira-Maia, A. J. (2022). Explicit knowledge of task structure is a primary determinant of human model-based action. *Nature Human Behaviour, 6*(8). https://doi.org/10.1038/s41562-022-01346-2. Article 8.

Cheadle, S., Egner, T., Wyart, V., Wu, C., & Summerfield, C. (2015). Feature expectation heightens visual sensitivity during fine orientation discrimination. *Journal of Vision, 15*(14), 14. https://doi.org/10.1167/15.14.14

Constant, M., Salomon, R., & Filevich, E. (2022). Judgments of agency are affected by sensory noise without recruiting metacognitive processing. *eLife, 11*, Article e72356. https://doi.org/10.7554/eLife.72356

Desantis, A., Roussel, C., & Waszak, F. (2011). On the influence of causal beliefs on the feeling of agency. *Consciousness and Cognition, 20*(4), 1211–1220. https://doi.org/10.1016/j.concog.2011.02.012

Desantis, A., Weiss, C., Schütz-Bosbach, S., & Waszak, F. (2012). Believing and perceiving: Authorship belief modulates sensory attenuation. *PLoS One, 7*(5), Article e37959. https://doi.org/10.1371/journal.pone.0037959

Dewey, J. A., & Knoblich, G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing? *PLoS One, 9*(10), Article e110118. https://doi.org/10.1371/journal.pone.0110118

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology Section A, 36*(1), 29–50. https://doi.org/10.1080/14640748408401502

Dogge, M., Schaap, M., Custers, R., Wegner, D. M., & Aarts, H. (2012). When moving without volition: Implied self-causation enhances binding strength between involuntary actions and effects. *Consciousness and Cognition, 21*(1), 501–506. https://doi.org/10.1016/j.concog.2011.10.014

Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness, 2017*(1), nix007. https://doi.org/10.1093/nc/nix007

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 443. https://doi.org/10.3389/fnhum.2014.00443

Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience, 21*(8). https://doi.org/10.1038/s41593-018-0200-7. Article 8.

Frith, C. D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine, 17*(3), 631–648. https://doi.org/10.1017/s0033291700025873

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 355*(1404), 1771–1788. https://doi.org/10.1098/rstb.2000.0734

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics (pp. xiii, 479)*. Robert E. Krieger.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience, 18*(4), Article 4. https://doi.org/10.1038/nrn.2017.14

Haggard, P., & Chambon, V. (2012). Sense of agency. *Current Biology, 22*(10), R390–R392.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*(4), 382–385. https://doi.org/10.1038/nn827

Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology. Human Perception and Performance, 16*(4), 802–811. https://doi.org/10.1037//0096-1523.16.4.802

Isham, L., Griffith, L., Boylan, A.-M., Hicks, A., Wilson, N., Byrne, R., … Freeman, D. (2021). Understanding, treating, and renaming grandiose delusions: A qualitative study. *Psychology and Psychotherapy: Theory, Research and Practice, 94*(1), 119–140. https://doi.org/10.1111/papt.12260

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Krugwasser, A. R., Stern, Y., Faivre, N., Harel, E. V., & Salomon, R. (2022). Impaired sense of agency and associated confidence in psychosis. *Schizophrenia (Heidelberg, Germany), 8*(1), 32. https://doi.org/10.1038/s41537-022-00212-4

Kulakova, E., Khalighinejad, N., & Haggard, P. (2017). I could have done otherwise: Availability of counterfactual comparisons informs the sense of agency. *Consciousness and Cognition, 49*, 237–244. https://doi.org/10.1016/j.concog.2017.01.013

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences, 22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Lindström, B., Golkar, A., Jangard, S., Tobler, P. N., & Olsson, A. (2019). Social threat learning transfers to decision making in humans. *Proceedings of the National Academy of Sciences, 116*(10), 4732–4737. https://doi.org/10.1073/pnas.1810180116

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*(4), 476–490. https://doi.org/10.3758/BF03210951

Luo, M.. For exercise in New York Futility, Push Button. The New York Times. https://www.nytimes.com/2004/02/27/nyregion/for-exercise-in-new-yorkfutility-push-button.html.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition, 21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Mellor, C. S. (1970). First rank symptoms of schizophrenia. I. The frequnncy in schizophrenics on admission to hospital. II. Differences between individual first rank symptoms. *The British Journal of Psychiatry, 117*(536), 15–23.

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition, 21*(1), 59–68. https://doi.org/10.1016/j.concog.2011.08.010

Moore, J. W., Lagnado, D., Deal, D. C., & Haggard, P. (2009). Feelings of control: Contingency determines experience of action. *Cognition, 110*(2), 279–283. https://doi.org/10.1016/j.cognition.2008.11.006

Olawole-Scott, H., & Yon, D. (2023). Expectations about precision bias metacognition and awareness. *Journal of Experimental Psychology. General, 152*(8), 2177–2189. https://doi.org/10.1037/xge0001371

Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision, 14*(11). https://doi.org/10.1167/14.11.5

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences, 21*(6), 425–433. https://doi.org/10.1016/j.tics.2017.03.011

Penton, T., Wang, X., Catmur, C., & Bird, G. (2022). Investigating the sense of agency and its relation to subclinical traits using a novel task. *Experimental Brain Research, 240*(5), 1399–1410. https://doi.org/10.1007/s00221-022-06339-1

Perrykkad, K., Lawson, R. P., Jamadar, S., & Hohwy, J. (2021). The effect of uncertainty on prediction error in the action perception loop. *Cognition, 210*, Article 104598. https://doi.org/10.1016/j.cognition.2021.104598

Press, C., Thomas, E. R., & Yon, D. (2023). Cancelling cancellation? Sensorimotor control, agency, and prediction. *Neuroscience and Biobehavioral Reviews, 145*, Article 105012. https://doi.org/10.1016/j.neubiorev.2022.105012

Press, C., & Yon, D. (2019). Perceptual prediction: Rapidly making sense of a Noisy world. *Current Biology, 29*(15), R751–R753. https://doi.org/10.1016/j.cub.2019.06.054

Skinner, B. F. (1948). 'Superstition' in the pigeon. *Journal of Experimental Psychology, 38*(2), 168–172. https://doi.org/10.1037/h0055873

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience, 15*(11), Article 11. https://doi.org/10.1038/nrn3838

Swets, J., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68*, 301–340.

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition, 17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Wen, W., Shibata, H., Ohata, R., Yamashita, A., Asama, H., & Imamizu, H. (2020). The active sensing of control difference. *iScience, 23*(5), Article 101112. https://doi.org/10.1016/j.isci.2020.101112

Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences, 109*(9), 3593–3598. https://doi.org/10.1073/pnas.1120118109

Yon, D., Bunce, C., & Press, C. (2020). Illusions of control without delusions of grandeur. *Cognition, 205*, Article 104429. https://doi.org/10.1016/j.cognition.2020.104429
Yon, D., & Corlett, P. R. (2022). Causal surgery under a Markov blanket. *The Behavioral and Brain Sciences, 45*, Article e218. https://doi.org/10.1017/S0140525X22000218
Yon, D., & Frith, C. D. (2021). Precision and the Bayesian brain. *Current Biology, 31*(17), R1026–R1032. https://doi.org/10.1016/j.cub.2021.07.044
Yon, D., Gilbert, S. J., de Lange, F. P., & Press, C. (2018). Action sharpens sensory representations of expected outcomes. *Nature Communications, 9*(1). https://doi.org/10.1038/s41467-018-06752-7. Article 1.
Yon, D., Thomas, E. R., Gilbert, S. J., de Lange, F. P., Kok, P., & Press, C. (2023). Stubborn predictions in primary visual cortex. *Journal of Cognitive Neuroscience, 35*(7), 1133–1143. https://doi.org/10.1162/jocn_a_01997
Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences, 10*(7), 301–308. https://doi.org/10.1016/j.tics.2006.05.002