# Metacognition bridges experiences and beliefs in Sense of Agency

John P. Veillette[1†]
Letitia Ho[1]
Howard C. Nusbaum[1]

[1]Department of Psychology, University of Chicago

[†]Correspondence should be addressed to John P. Veillette; E-mail: johnv@uchicago.edu

## Author Contributions (CRediT Taxonomy)

**H.C.N.:** Conceptualization, Funding acquisition, Resources, Supervision, and Writing - review & editing. **J.P.V.:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing - original draft, and Writing - review & editing. **L.H.:** Conceptualization, Data curation, Investigation, Software, Validation, and Writing - review & editing.

## Funding

J.P.V. was supported by NSF GRFP DGE 1746045 and a Neubauer Family Distinguished Doctoral Fellowship. This study was supported by NSF NCS 2024923 awarded to H.C.N.

## Conflict of Interest

The authors declare no conflict of interest.

## Abstract

Cognitive scientists differentiate the "minimal self" – subjective experiences of agency and ownership in our sensorimotor interactions with the world – and the "narrative self" that encompasses those beliefs about the self that are sustained over time. How exactly moment-to-moment experiences are integrated into narrative beliefs, however, remains an open question. We administered a battery of sensorimotor tasks and surveys to index subjects' ($n$ = 195) propensity to classify stimuli as self-caused and their metacognitive monitoring of such agency judgements, and we compared these behavioral metrics to trait-level beliefs about their own agency. Subjects who were less sensitive to sensory control cues in the sensorimotor tasks also reported lower trait-level agency beliefs. Importantly, however, this relationship all but disappears when controlling for metacognitive accuracy. These results suggest narrative beliefs about self-agency are not just the sum of individual experiences of self-causation but rather the product of a metacognitively-driven integration process.


## Statement of Relevance

Philosophers have discussed the relationship between low-level phenomenological experiences and high-level beliefs pertaining to the sense of self for thousands of years, and this relationship is far from a mere intellectual curiosity – our understanding of the self both informs theory in the psychology of action and has practical applications. For instance, patients suffering from numerous psychiatric disorders report disturbances in agency beliefs, and scientists often assume such symptoms are caused by dysfunctions of the low-level cognitive mechanisms that produce moment-to-moment agency judgements. Our findings, however, suggest that the process by which these individual judgments are integrated into higher level beliefs better explains individual differences in people's self-assessment of their own agency (at least within a healthy population). This study brings empirical evidence to bear on the fundamental question of how subjective experiences of selfhood are related across levels of abstraction.

## Introduction

Sense of agency (SoA) is the feeling or belief that one controls one's own actions and, through those actions, can influence events in the world. We experience a feeling of "I did that" as we intentionly take action. This phenomenological SoA, together with body ownership, has been argued by philosophers, psychologists, and neuroscientists alike to be the most basic building block of a minimal conscious self-awareness (Gallagher, 2000; Tsakiris et al., 2006). However, beyond the "minimal self," SoA is also discussed as a high-level belief about one's level of control incorporated into the "narrative self" that is sustained over time (Dennett, 1993; Gallagher, 2000). It is an important aspect of this idea that self-reported beliefs about one's own agency appear to be relatively stable over time, constituting a trait-level phenomenon (Tapal et al., 2017). It is assumed that the more elaborated narrative self is, in some way, constructed from our moment-to-moment experience of minimal selfhood, but the operating characteristics of this putative integration have been left vague. Are narrative beliefs about the self merely the sum of individual experiences, or is there more to the construction of self-narratives? Since there are already validated tools for measuring SoA as both a "low-level", pre-reflexive, sensorimotor phenomenon and as a high-level belief, SoA provides a framework to investigate the relationship between experiences and beliefs in the construction of conscious self-awareness.

At a sensorimotor level, humans and other organisms are constantly engaged in a process of distinguishing between self-caused and externally-caused sensory stimuli. The most basic expression of this distinction is the neural suppression of predictable sensations resulting from voluntary movement (C. D. Frith, 1987). That is, sensations are perceived as self-caused when they are predictable from intended actions, but whether one experiences agency over a particular sensation further depends upon other cues such as prospective movement intentions (C. Frith, 2012; Haggard, 2017) and short-term beliefs about whether causality is possible in the current context (Desantis et al., 2011), to be differentiated from the long-term/trait-level beliefs in question in the present study. The degree to which the sensory consequence of an action must match the predicted consequence before agency is felt may vary from subject to subject; one's sensitivity to cues to their own control has been measured using control discrimination tasks, in which subjects identify which of two stimuli they are able to influence with their movements (Wang et al., 2020; Wen & Haggard, 2018, 2020). In addition to *whether* one feels agency over an action or outcome (i.e. an agency judgement), one can also discuss *how* one experiences agency (i.e. a phenomenological experience). A basic finding is that, when one experiences agency over an action-outcome pairing (say, a button press with a resulting tone), the perceived time of the action is shifted toward the time of the outcome and vice versa, putatively *intentionally binding* the two into a single event according to the common theoretical interpretation (Haggard et al., 2002). While often used as an implicit measure of agency, the magnitude of the intentional binding effect is also used as an index of the depth to which SoA affects conscious experience (Lush et al., 2017). While, depending on a number of contextual factors, judgements of agency may be influenced by metacognitive processes (Chambon et al., 2014), a combination of empirical findings and computational modelling has been used to convincingly argue that agency judgements need not depend on metacognitive processing (Constant et al., 2022).

At the level of the narrative self, explicit trait-level beliefs about one's SoA can be measured psychometrically using the Sense of Agency Scale (SoAS) (Tapal et al., 2017). Two

factors, termed "positive" SoA (SoPA) and "negative" SoA (SoNA) can be derived from subject responses to the scale items. The stability of this factor structure has been confirmed and replicated, and these factors can be differentiated from other related constructs such as self-efficacy beliefs and free will beliefs, and there is high test-retest reliability that can be seen even when separated by months (Hurault et al., 2020; Tapal et al., 2017). Moreover, the measured factors predict obsessive-compulsive symptoms and differ between patients with psychosis and healthy controls, which has suggested that these factors meaningfully, though necessarily crudely, quantify clinically important differences in subjective experience (Kruse et al., 2022; Tapal et al., 2017).

The present study, then, aims to assess the degree to which high-level beliefs about agency are informed by moment-to-moment agency judgements in our sensorimotor interactions. We use validated online tasks to measure, for each subject, sensitivity to sensory evidence of control during agency judgements and the accuracy of metacognitive monitoring of those judgements (Wang et al., 2020) as well as the magnitude of the intentional binding effect as an index of how deeply inferences of agency affect consciousness awareness (Galang et al., 2021). We then estimate the extent to which individual differences in these indices of moment-to-moment SoA predict beliefs about SoA measured by the SoAS. Our results, accordingly, inform our understanding of how SoA at the sensorimotor level relates to SoA at the level of narrative beliefs.

# Methods

## Subject recruitment and ethics

200 subjects were recruited online from across the United States using Prolific (prolific.co), the behavioral task was hosted on Pavlovia (pavlovia.org), and all experiment code was written in JavaScript using the jsPsych library (de Leeuw, 2015). Subjects in Prolific's recruitment pool were only allowed to participate if 95% of their previous submissions on the site had been approved and they were using the Windows operating system. 5 subjects' data were lost due to technical error, resulting in $n = 195$. Subjects' data were excluded from analysis of particular tasks or scales (not removed from analyses of other tasks) if they failed to pass exclusion criteria/attention checks for that task, such as failing an unacceptably high proportion of trivially easy "catch" trials, or if they had partially missing data. Please see task descriptions below for exact exclusion criteria and counts of subjects removed from each task.

All subjects gave written, informed consent before participating. All of the methods performed in the study were in accordance with relevant safety and ethics guidelines for human subject research and were approved by the Social and Behavioral Sciences Institutional Review Board at the University of Chicago (IRB21-1458). This study was not a clinical trial.

## Sense of Agency Scale

Subjects completed the Sense of Agency Scale (SoAS) as introduced and first validated by Tapal and colleagues (Tapal et al., 2017). The factor loadings reported by Tapal et al. were multiplied by the individual Likert-scale responses to the scale items to obtain numerical values for sense of positive agency (SoPA) and sense of negative agency (SoNA), factors which explain separable components of variance in item responses (Hurault et al., 2020; Tapal et al., 2017). The

test-retest correlations reported by Tapal et al., measured two months apart, are $r = 0.78$ for SoPA and $r = 0.74$ for SoNA; these correlations may be useful to consider as a point estimate of the potentially explainable variance in scores when considering the effect size estimates we report here.

Subjects' data were excluded from this task if they only replied with 1's and 7's on the Likert scale, indicating lack of honest effort in responding. 20 subjects were excluded on this basis, resulting in $n = 175$.

**Intentional binding task**

We use a conventional Libet clock paradigm (REF) for measuring the magnitude of intentional binding. In summary in this task, the subject sees a moving clock hand on each trial, during which they are asked to press a key on their keyboard (at their leisure but earlier than 8 seconds into the trial and but after the first rotation of the clock hand has occurred, or the trial restarts). Following the key press, a tone is presented. At the end of each trial, subjects are asked to move the clock hand back to either (1) the time they pressed the *key* or (2) the time they heard the *tone*. (A) In the *operant* condition, the tone followed the keypress with a constant 0.25 second delay. (B) In the *baseline* condition, the tone occurred at a random time, uniformly distributed throughout the trial. The magnitude of intentional binding for each subject was computed separately for the key and for the tone by subtracting the average overestimation of the event onset (in milliseconds relative to the true event onset) in the baseline condition from that in the operant condition. Thus, these measurements reflect the degree to which perception of the keypress and tone events are shifted toward each other in time when the former is perceived as causing the latter, or intentional binding. Each 2x2 condition (i.e. baseline-tone, operant-tone, etc.) had 40 trials preceded by 5 practice trials, and we used a preexisting jsPsych implementation of the Libet clock paradigm which had already been validated for online use (Galang et al., 2021).

For analyses of individual differences (see *Data Analysis* below), the sign of the individual subject tone binding effect was flipped, such that more positive values mean stronger binding – just as for key binding.

Subjects' data were excluded from this task if their overestimation in any one condition was farther than 5 standard deviations from the mean overestimation in order to ensure included subjects were not responding randomly. Only 5 subjects were excluded on this basis, indicating decent task compliance overall. The resulting sample size was $n = 190$.

**Motion tracking task**

We used Wang and colleagues' jsPsych implementation of the sensorimotor task they introduced and validated (Wang et al., 2020). As with other control discrimintation tasks in the literature (Wen & Haggard, 2018, 2020), the task of the subject is to determine which of two moving dots they are able to influence the trajectory of by moving their mouse, while the actual degree of control is low enough to make accurate discrimintation challenging. In this way, the task measures perceptual sensitivity to control cues.

Specifically, two moving dots, following independent, pseudorandom trajectories, were presented within separate circles on the screen. The subject could move their cursor to influence the trajectory of one of the two dots (the "target" dot) but were not told which dot they were influencing. The percentage of the target dot's trajectory that the subject could influence ("percent control") was manipulated across trials. Subjects had 4 seconds to view/influence the dot stimuli, followed by a 0.5 second blank screen before they were asked to identify which dot they thought they were influencing. Subsequently, they were asked to rate their confidence in their answer.

As described by Wang and colleagues, the task begins with 5 practice trials starting at 25% control (very easy). After the practice trials, the experiment proceeded in two interwoven adaptive staircase procedures by which percent control was adjusted 13 times over 100 trials per staircase, resulting in a total of 200 trials. Please refer to Wang and colleagues' (2020) paper for a full description of the staircase procedure. 15% of those 200 trials were randomly inserted "catch" trials, in which percent control was always 25%. The goal of the staircase was to hone in on the percent control in which the subject could identify the target dot with 75% accuracy. To this end, the average percent control along the last five staircase reversals was taken as the "percent control threshold," which served as our metric for each subject's sensitivity to visual control cues during sensorimotor agency judgements. As pointed out by Wang et al. (2020), the distribution of percent control threshold measurements is highly skewed, so these values were log transformed so as to be closer to normally distributed ("log control threshold").

Moreover, to quantify metacognitive accuracy, we computed the area under the receiver operating characteristic curve for classifying correct vs. incorrect trials (over all trials) from subjects' confidence ratings (type II AUROC). This common measure of metacognitive performance reflects how calibrated subjects' uncertainty judgements are; that is, are they actually wrong more often when they are more uncertain?

Following Wang et al. (2020), subjects' data were excluded from this task if they failed over 40% of the easy catch trials, indicating they were responding effectively randomly. Moreover, we excluded subjects whose AUROC scores were significantly *below* chance (0.5) with a significance threshold of $p < 0.05$ via a Mann-Whitney test; that is, reported confidence was inversely correlated with success. We interpreted such cases as a misunderstanding of task instructions, such as believing 1 was "most confident" rather than "least confident," whereas subjects with AUROCs moderately but not significantly below chance were assumed to represent legitimate variance in performance. One additional subject was removed because their data contained unexplained missing values. Based on these criteria, 13 subjects were excluded, resulting in a sample size of $n = 182$.

**Other self-report measures**

We asked subjects to complete two other brief surveys for the purpose of obtaining pilot effect size estimates for future studies. Thus, these scales were never analyzed and results are not reported here, but the raw data are available in our open dataset and may be of use to other researchers. These scales were the Tellegan Absorption Scale (Tellegen & Atkinson, 1981) and the Embodied Sense of Self Scale (Asai et al., 2016).

## Data analysis

All analyses were done using Python. Distributions of measurements from the two sensorimotor tasks were visualized using the *DABEST* and *ptitprince* packages (see Figure 1). Confidence intervals were derived by bootstrap for the Cohen's *d* effect size of intentional binding effects for purposes of replication.

Before fitting statistical models, all sensorimotor measures and SoAS scores were *Z*-normalized to put them on a common scale. We then took a Bayesian approach to estimating posterior distributions of effect sizes of interest using the *PyMC* package for posterior sampling. In all Bayesian analyses reported here, we used a Normal(0, 1) prior for population means and regression coefficients, an Exponential(1) prior for noise terms, and an LKJ(eta = 2) prior for correlations. These conservative priors all serve to shrink posterior effect size estimates toward zero, placing only a small prior probability on large effect sizes. Posteriors were approximated by drawing 10,000 posterior samples across 2 sampling chains using *PyMC*'s no-U-turn sampler.

We estimate Bayesian posteriors for the correlations between each sensorimotor/behavioral measure and SoPA and SoNA separately. These posteriors are visualized in Figure 2 and summary statistics are reported in Table 1. Given a strong observed correlation between log control threshold and AUROC measures observed the motion tracking task (see Figure 3), we performed a mediation analysis to assess the degree to which this correlation was confounding the two metrics' pairwise correlations with SoNA (see Figure 2). Posterior distributions for total, direct, and indirect effects were estimated with a linear mediation model and shown in Figure 4.

## Open practices statement

All code for both the experiment (https://doi.org/10.5281/zenodo.8173285) and the analysis (https://doi.org/10.5281/zenodo.8173283) is permanently archived on Zenodo. Deidentified raw data is available on the Open Science Framework (https://osf.io/753c2/) and is organized roughly according to the Brain Imaging Data Structure specifications for behavioral data to facilitate easy navigation. The analysis plan was not preregistered, but estimation-oriented statistics were used to provide a transparent representation of the statistical uncertainty about the magnitude of effects (Cumming, 2014).

# Results

We observe distributions of behavioral effects consistent with the prior work from which our sensorimotor tasks (Libet and motion tracking) were taken (Galang et al., 2021; Wang et al., 2020). We replicate the previously reported intentional binding effect for both key ($d$ = 0.39, 95% CI [0.23, 0.54]) and tone ($d$ = -0.72, 95% CI [-0.87, -0.58]). In comparison, the meta-analytic effect sizes for the intentional binding effect are $d$ = 0.45 and $d$ = -0.73 for action (key) and outcome (tone) binding, respectively (Tanaka et al., 2019). In other words, we obtained effect size estimates consistent with "gold standard" in-lab measurements. Additionally, the distributions we observe in the motion tracking task are qualitatively quite similar to those obtained by Wang and colleagues in their original validation of the task (Wang et al., 2020). Observed distributions of all these behavioral measurements are visualized in Figure 1.

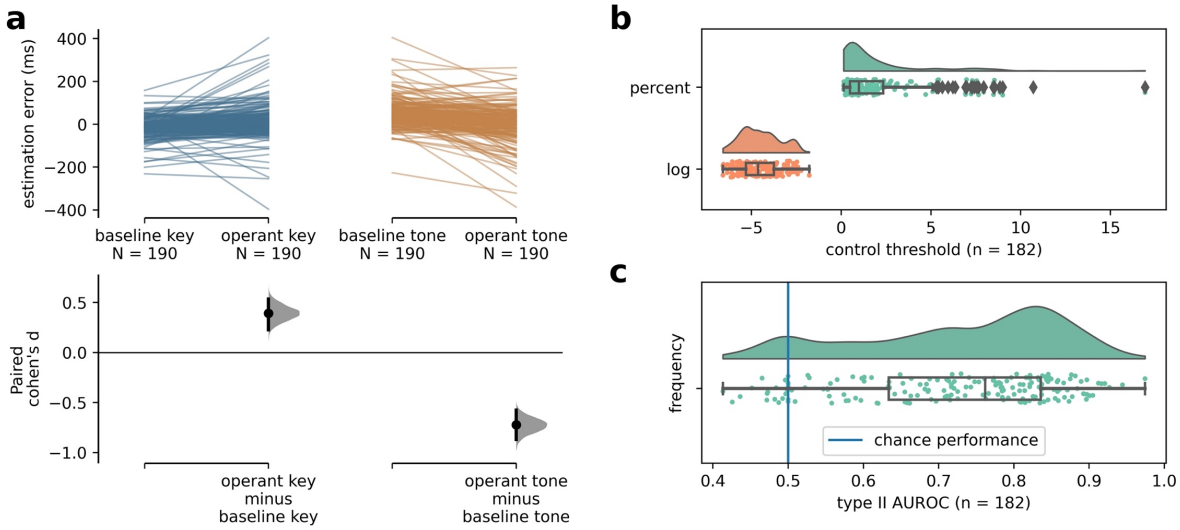


**Figure 1: Distributions of sensorimotor, behavioral measures.** (a) Each subjects' mean estimates of the timing of keypress and tone events relative to the true event times in the Libet task in each condition are shown on top, with bootstrapped distributions and 95% confidence intervals of the Cohen's *d* effect size for group-level intentional binding effects on bottom. (b and c) Raincloud plots of control threshold (in both percent and log scale), measuring control cue sensitivity, and type II AUROC, measuring metacognitive performance. Box component of raincloud plots shows the median and quantiles, while whiskers show the extent of the distribution excluding extreme points (for visualization only). (b) Extreme points (those that fall more than 1.5 the interquartile range from the closest quartile are marked with diamonds; those points are no longer extreme once log scaled.

Bayesian posterior distributions for Pearson correlations between binding effects, control thresholds (i.e. sensitivity), and metacognitive accuracy, on one hand, and trait-level SoPA and SoNA on the other are shown in Figure 2, with summary statistics in Table 1. We find evidence of a positive correlation between log control threshold (i.e. the inverse of sensitivity to control cues) and SoNA. That is, those who are less sensitive to control cues report having less agency. Moreover, we find evidence of a negative correlation between metacognitive ability (type II AUROC) and SoNA. That is, those with better metacognitive ability report feeling more agency (or less negative agency) overall. We do not find sufficient evidence to draw a conclusion as to whether intentional binding magnitudes predict beliefs about agency, but we report 95% "highest density intervals" (HDIs, i.e. Bayesian credible intervals) in Table 1 which place upper bounds on how large such an effect could plausibly be based on our data.

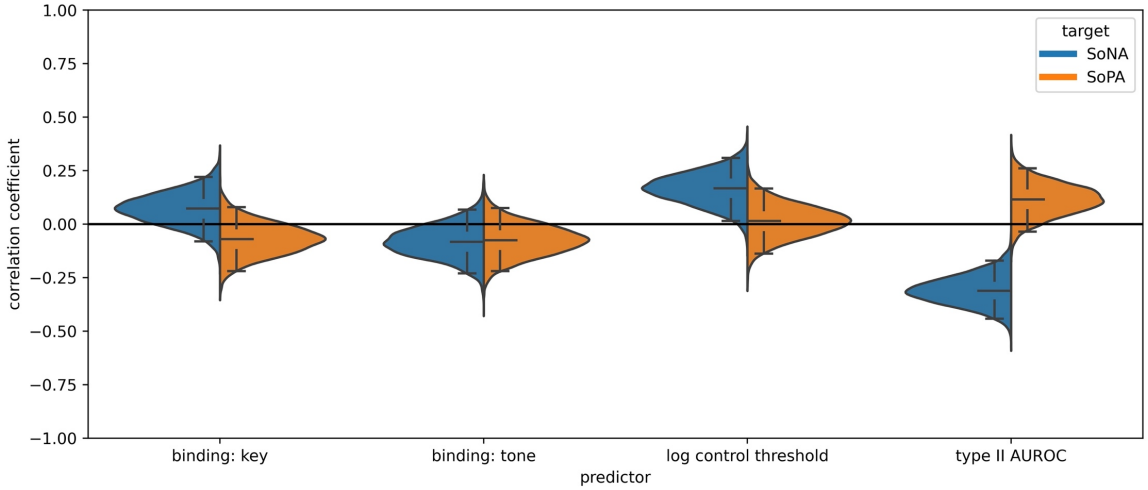


**Figure 2: Posterior distributions of correlations between sensorimotor, behavioral measures and agency beliefs.** Behavioral measures are as in Figure 1. Agency beliefs, measured by the Sense of Agency Scale, are subdivided into sense of positive agency (SoPA) and negative agency (SoNA). Whiskers overlaid atop the violin plots extend to the 2.5% and 97.5% quantiles of the posteriori distributions, representing 95% credible intervals.

| Predictor | Target | Mean | Lower HDI | Upper HDI | Prob. Neg. | Prob. Pos. | R-hat |
|---|---|---|---|---|---|---|---|
| **binding: key** | **SoNA** | -0.071 | -0.217 | 0.082 | 0.823 | 0.177 | 1.000 |
| **binding: key** | **SoPA** | 0.072 | -0.080 | 0.221 | 0.172 | 0.828 | 1.000 |
| **binding: tone** | **SoNA** | -0.075 | -0.219 | 0.076 | 0.838 | 0.163 | 1.000 |
| **binding: tone** | **SoPA** | -0.082 | -0.230 | 0.068 | 0.857 | 0.143 | 1.000 |
| **log control thres.** | **SoNA** | 0.014 | -0.138 | 0.165 | 0.426 | 0.574 | 1.000 |
| **log control thres.** | **SoPA** | 0.166 | 0.015 | 0.309 | 0.017 | **0.983** | 1.000 |
| **type II AUROC** | **SoNA** | 0.115 | -0.034 | 0.261 | 0.069 | 0.931 | 1.000 |
| **type II AUROC** | **SoPA** | -0.310 | -0.448 | -0.179 | **1.000** | 0.000 | 1.000 |

**Table 1: Posterior summary statistics for correlations behavioral measures and agency beliefs.** Summary statistics include posterior mean (expected value), lower edge of 95% highest density interval (HDI), upper edge of 95% HDI, posterior probability effect size is negative, probability effect size is positive, and R-hat: a measure of the convergence of the posterior sampling procedure that is optimal at R-hat = 1.

While we found that control sensitivity (log control threshold) and metacognitive performance (type II AUROC) both correlate with SoNA separately, visualizing the joint distributions of the three measurements as in Figure 3 reveals a clear correlation between log control threshold and AUROC of $r$ = -0.57 (95% HDI [-0.67, -0.46]) between the two measures. This finding motivated a mediation analysis to determine whether this correlation between behavioral predictors confounded our estimate of their correlation with SoNA.

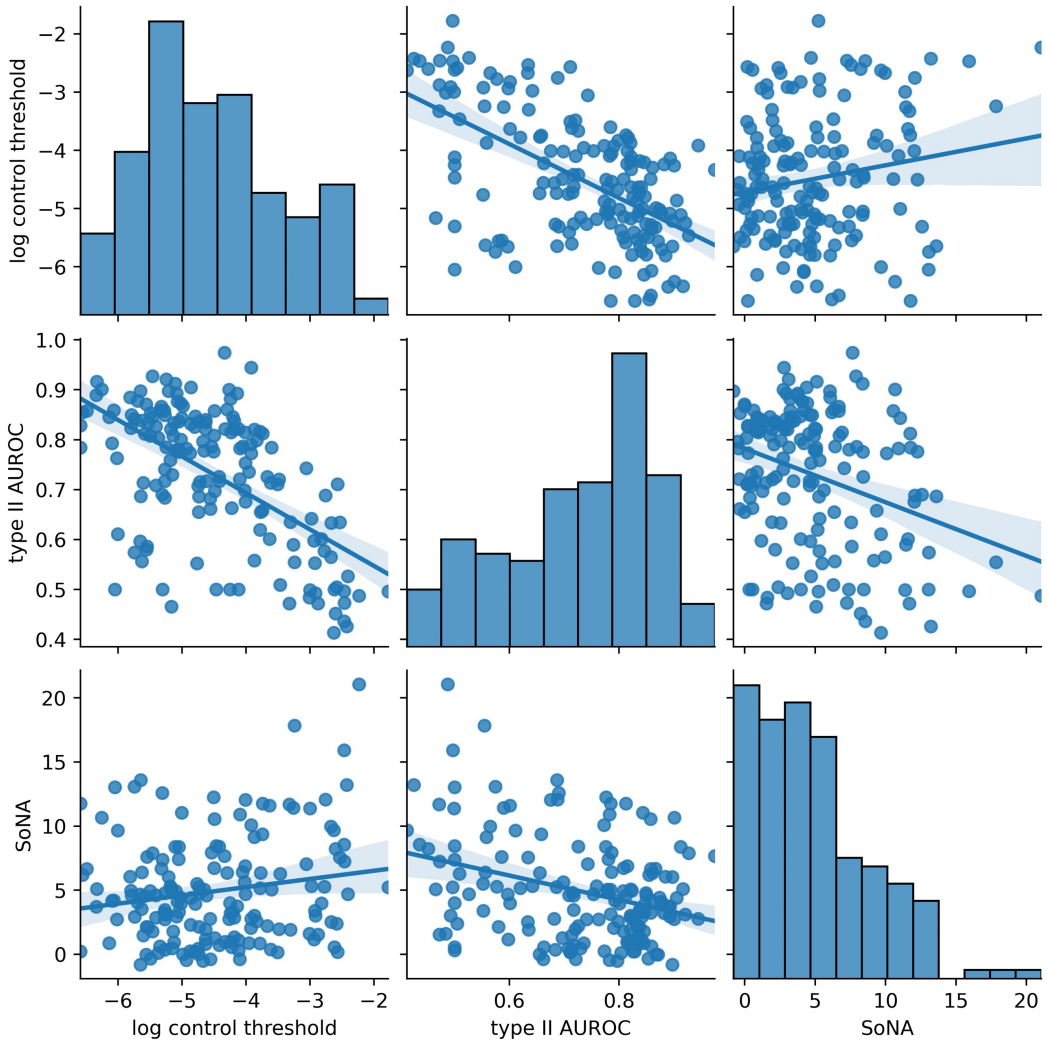


**Figure 3: Joint distributions of control cue sensitivity, metacognitive performance, and sense of negative agency.** Histograms for each variable are shown on the diagonal, raw data with best-fit linear regression lines and 95% confidence bands are shown off the diagonal. Data are shown in their original scale for visualization only.

When we investigate further whether the correlation between control cue sensitivity and metacognitive ability confounds our estimate of the relationship between control cue sensitivity and SoNA, we find that it indeed does. In a linear mediation analysis, we find that the total effect of log control threshold on SoNA (beta = 0.17, 95% HDI [0.02, 0.33]) can be decomposed into a weak, if present at all, direct effect (beta = -0.02, 95% HDI [-0.21, 0.15]) and a clear indirect effect mediated by type II AUROC (beta = 0.19, 95% HDI [0.08, 0.31]). In other words, we find that the effect of control cue sensitivity on agency beliefs is driven, in large part, by a shared correlation with metacognitive performance. Posterior distributions for this mediation analysis are visualized in Figure 4. While we statistically model control sensitivity as affecting agency beliefs, please note our analysis does not rule out the possibility that causality may flow in the reverse direction as well (Rigoni et al., 2011). Our mediation analysis simply constrains the possibility space of causal structures relating control sensitivity and agency beliefs (with whatever directionality) to those that are mediated by metacognition.

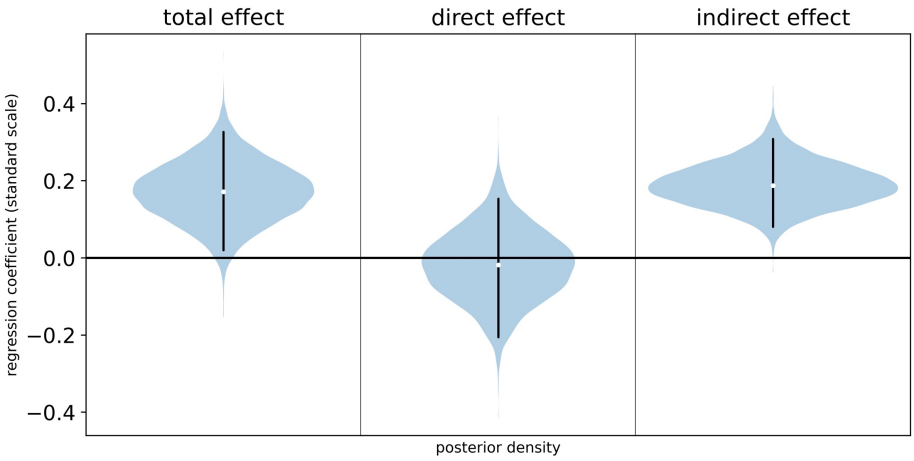


**Figure 4: Effect of control cue sensitivity on agency beliefs, mediated by metacognitive performance.** Posterior distributions for regression coefficients from a linear mediation analysis with 95% HDIs overlaid are shown. The total effect is accounted for, in large part, by an indirect effect mediated by metacognitive performance. All variables (log control threshold, type II AUROC, and SoNA) were standardized before mediation analysis, so the regression coefficient estimates are on roughly the same scale as the correlations visualized in Figure 2.

## Discussion

Agency judgements (or self-vs-other judgements in general) do not only occur at the sensorimotor level, nor is there a sudden jump from experiences of individual agency judgments to narrative beliefs. Sense of agency (SoA) has been studied at many levels of abstraction (e.g. mental, social, etc.), and recent controversies have cast doubt on the notion that a common cognitive or neural substrate can account for agency judgements across all these scales (Galang et al., 2021; Wang et al., 2020). Indeed, the neural predictors of agency judgements appear to meaningfully differ even between different types of sensorimotor judgements, such as those concerning muscle movements (Veillette et al., 2023) and downstream outcomes (Timm et al., 2016). If asked, however, we suspect most people would say that the "I" to which they attribute actions and consequences does not differ across these domains. To the extent that SoA is generated by different mechanisms at different scales of biological, cognitive, and behavioral organization, what is it that links these levels of abstraction such that self-caused phenomenon in different domains are all experienced as belonging to the same, unified self in consciousness?

Our results begin to address this fundamental question. While it would be intuitive to theorize that those who experience agency more frequently in their moment-to-moment agency judgements will report higher SoA when asked about their narrative-level beliefs – as a matter of statistical learning – what we find instead is more nuanced. While the intuitive correlation between (in)sensitivity to control cues (i.e. log control threshold) and sense of negative agency (SoNA) does appear to exist, this effect was mediated primarily by metacognitive accuracy (i.e. type II AUROC, see *Methods*) about such agency judgements – that is, the accuracy with which one monitors uncertainty about agency judgements. In other words, we do not find evidence of a direct relationship between sensitivity to control cues and trait-level agency beliefs. While sensitivity to control and metacognitive accuracy were correlated in the present study, this need not be the case in all settings; agency judgements can be made without recruiting metacognitive resources

(Constant et al., 2022). In such cases, however, metacognition may still play a role in determining how individual experiences of agency are integrated into a larger self-concept. In this vein, our findings suggest that metacognition may provide a link between (at least some of the) different levels of abstraction at which one experiences selfhood. Moreover, our data challenge recent arguments that higher levels beliefs about agency are entirely socially constructed, clearly linking such beliefs to behavioral indices of sensorimotor experience – albeit through a metacognitive mediator (Jenkins, 2001).

While we saw trends toward the magnitude of intentional binding predicting SoA beliefs, we do not have enough evidence to conclude one way or another whether such a relationship exists. Interestingly, previous work has suggested that the magnitude of intentional binding influences free will beliefs (Aarts & van den Bos, 2011) and, conversely, that free will beliefs affect motor preparatory neural activity (Rigoni et al., 2011). Free will beliefs, which concern the existence of mental causation in general, do differ from SoA beliefs, which pertain only to one's own ability to exert control over the world. One possibility is that free will beliefs are more influenced by *how* one experiences volitional action, as reflected in intentional binding, and SoA beliefs are more influenced by *whether* one experiences actions as volitional. This distinction could be a fruitful subject for future study. Moreover, we did not find compelling evidence that any sensorimotor metric predicted positive SoA (SoPA), only SoNA. This finding (or lack thereof) makes sense in light of existing theory, as interruptions of normal sensorimotor control become salient intrusions in consciousness, but the routine flow from action to outcome naturally falls into the background (Synofzik et al., 2008). Nonetheless, it remains an open question from where the meaningful variance in SoPA originates.

It is important to note some limitations on the inferences we can draw from the present data. Obviously, we did not measure all possible behavioral indices of agency experience at either the sensorimotor or the narrative level; indeed, no single study can. Consequently, we cannot rule out a direct effect of control sensitivity or some other index that would affect the frequency of positive agency judgements on SoA beliefs. Not all such relationships, if they exist, are necessarily mediated by metacognition. Moreover, our correlation estimates fall far below the test-retest correlation of the Sense of Agency scale (Tapal et al., 2017), suggesting that there is still much meaningful variance in agency beliefs left to be explained – in all likelihood by factors that are not to be found at the sensorimotor level. Further, the extent to which the observed correlations are explained by a causal effect of sensorimotor experience on beliefs, rather than of beliefs on sensorimotor experience, remains unclear. However, our results clearly show that a (surprisingly) substantial portion of the individual differences in self-agency beliefs are concretely related to one's sensorimotor experience of volitional action, and that the observed relationship is mediated by metacognition. Overall, our findings point toward a model of conscious selfhood in which moment-to-moment experiences are aggregated into beliefs by a process of metacognitive integration, which may serve to connect the facets of self which are experienced across scales of biological, mental, and behavioral organization.

## References


Aarts, H., & van den Bos, K. (2011). On the Foundations of Beliefs in Free Will: Intentional Binding and Unconscious Priming in Self-Agency. *Psychological Science*, *22*(4), 532–537. https://doi.org/10.1177/0956797611399294

Asai, T., Kanayama, N., Imaizumi, S., Koyama, S., & Kaganoi, S. (2016). Development of Embodied Sense of Self Scale (ESSS): Exploring Everyday Experiences Induced by Anomalous Self-Representation. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01005

Chambon, V., Filevich, E., & Haggard, P. (2014). What is the Human Sense of Agency, and is it Metacognitive? In S. M. Fleming & C. D. Frith (Eds.), *The Cognitive Neuroscience of Metacognition* (pp. 321–342). Springer. https://doi.org/10.1007/978-3-642-45190-4_14

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

Dennett, D. (1993). *Consciousness explained*. Penguin uk.

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

Frith, C. (2012). Explaining delusions of control: The comparator model 20 years on. *Consciousness and Cognition*, *21*(1), 52–54. https://doi.org/10.1016/j.concog.2011.06.010

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

Galang, C. M., Malik, R., Kinley, I., & Obhi, S. S. (2021). Studying sense of agency online: Can intentional binding be observed in uncontrolled online settings? *Consciousness and Cognition*, *95*, 103217. https://doi.org/10.1016/j.concog.2021.103217

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

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

Hurault, J.-C., Broc, G., Crône, L., Tedesco, A., & Brunel, L. (2020). Measuring the Sense of Agency: A French Adaptation and Validation of the Sense of Agency Scale (F-SoAS). *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.584145

Jenkins, A. H. (2001). Individuality in Cultural Context: The Case for Psychological Agency. *Theory & Psychology*, *11*(3), 347–362. https://doi.org/10.1177/0959354301113004

Kruse, E., Lesh, T., Board, S., Carter, C., & Joiner, W. (2022). P588. Exploring the Neural Correlates of Sense of Agency Deficits in Psychosis: A DTI Study. *Biological Psychiatry*, *91*(9), S327–S328. https://doi.org/10.1016/j.biopsych.2022.02.825

Lush, P., Caspar, E. A., Cleeremans, A., Haggard, P., Magalhães De Saldanha da Gama, P. A., & Dienes, Z. (2017). The Power of Suggestion: Posthypnotically Induced Changes in the Temporal Binding of Intentional Action Outcomes. *Psychological Science*, *28*(5), 661–669. https://doi.org/10.1177/0956797616687015

Rigoni, D., Kühn, S., Sartori, G., & Brass, M. (2011). Inducing Disbelief in Free Will Alters Brain Correlates of Preconscious Motor Preparation: The Brain Minds Whether We Believe in Free Will or Not. *Psychological Science*, *22*(5), 613–618. https://doi.org/10.1177/0956797611405680

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

Tanaka, T., Matsumoto, T., Hayashi, S., Takagi, S., & Kawabata, H. (2019). What Makes Action and Outcome Temporally Close to Each Other: A Systematic Review and Meta-Analysis of Temporal Binding. *Timing & Time Perception*, *7*(3), 189–218. https://doi.org/10.1163/22134468-20191150

Tapal, A., Oren, E., Dar, R., & Eitam, B. (2017). The Sense of Agency Scale: A Measure of Consciously Perceived Control over One's Mind, Body, and the Immediate Environment. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01552

Tellegen, A., & Atkinson, G. (1981). Tellegen Absorption Scale. *Journal of Abnormal Psychology*. https://doi.org/10.1037/t14465-000

Timm, J., Schönwiesner, M., Schröger, E., & SanMiguel, I. (2016). Sensory suppression of brain responses to self-generated sounds is observed with and without the perception of agency. *Cortex*, *80*, 5–20. https://doi.org/10.1016/j.cortex.2016.03.018

Tsakiris, M., Prabhu, G., & Haggard, P. (2006). Having a body versus moving your body: How agency structures body-ownership. *Consciousness and Cognition*, *15*(2), 423–432. https://doi.org/10.1016/j.concog.2005.09.004

Veillette, J. P., Lopes, P., & Nusbaum, H. C. (2023). *Temporal Dynamics of Brain Activity Predicting Illusory Agency over Involuntary Movements* (p. 2023.05.06.539706). bioRxiv. https://doi.org/10.1101/2023.05.06.539706

Wang, S., Rajananda, S., Lau, H., & Knotts, J. D. (2020). New measures of agency from an adaptive sensorimotor task. *PLOS ONE*, *15*(12), e0244113. https://doi.org/10.1371/journal.pone.0244113

Wen, W., & Haggard, P. (2018). Control Changes the Way We Look at the World. *Journal of Cognitive Neuroscience*, *30*(4), 603–619. https://doi.org/10.1162/jocn_a_01226

Wen, W., & Haggard, P. (2020). Prediction error and regularity detection underlie two dissociable mechanisms for computing the sense of agency. *Cognition*, *195*, 104074. https://doi.org/10.1016/j.cognition.2019.104074