https://doi.org/10.1093/cercor/bhad547 Advance access publication date 13 February 2024 Original Article

# Neural correlates of an illusionary sense of agency caused by virtual reality

Yiyang Cai<sup>1,2</sup>, Huichao Yang<sup>3</sup>, Xiaosha Wang<sup>3</sup>, Ziyi Xiong (b<sup>3</sup>, Simone Kühn<sup>4,5</sup>, Yanchao Bi<sup>3,6</sup>, Kunlin Wei (b<sup>1,2,\*</sup>

<sup>1</sup>School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China,

<sup>2</sup>Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China,

<sup>3</sup>State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China,

<sup>4</sup>Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, 20251 Hamburg, Germany,

<sup>5</sup>Lise Meitner Group for Environmental Neuroscience, Max Planck Institute for Human Development, 14195 Berlin, Germany,

<sup>6</sup>Chinese Institute for Brain Research, Beijing 102206, China

\*Corresponding author: School of Psychological and Cognitive Sciences, Peking University, 5 Yiheyuan Road, Beijing 100871, China. Email: wei.kunlin@pku.edu.cn

Sense of agency (SoA) is the sensation that self-actions lead to ensuing perceptual consequences. The prospective mechanism emphasizes that SoA arises from motor prediction and its comparison with actual action outcomes, while the reconstructive mechanism stresses that SoA emerges from retrospective causal processing about the action outcomes. Consistent with the prospective mechanism, motor planning regions were identified by neuroimaging studies using the temporal binding (TB) effect, a behavioral measure often linked to implicit SoA. Yet, TB also occurs during passive observation of another's action, lending support to the reconstructive mechanism, but its neural correlates remain unexplored. Here, we employed virtual reality (VR) to modulate such observation-based SoA and examined it with functional magnetic resonance imaging (fMRI). After manipulating an avatar hand in VR, participants passively observed an avatar's "action" and showed a significant increase in TB. The binding effect was associated suggest that the experience of controlling an avatar may potentiate inferential processing within the right inferior parietal cortex and give rise to the illusionary SoA without voluntary action.

Key words: binding; embodiment; self-consciousness; sense of agency; virtual reality.

### Introduction

Sense of agency (SoA) is the sensation that self-initiated actions influence the external environment. We implicitly experience the feeling of the connection between our action and the resulting consequence and attend to its disruption only when the actual action feedback conflicts with our expected consequences. As an integral part of self-consciousness, SoA enables one to feel fluent control over one's surroundings (Haggard 2017), distinct from others (Kahl and Kopp 2018), and responsible for one's own actions (Haggard and Tsakiris 2009). A clear understanding of the computations underlying SoA is still lacking but two major mechanisms, i.e. the prospective and the reconstructive mechanism, are currently attested and might both contribute to the manifestation of SoA (Moore and Obhi 2012). The prospective mechanism emphasizes that SoA is based on a predictive process in the motor system and a comparative process for comparing the predicted and actual action feedback (Frith et al. 2000; Gallagher 2000; Wolpert and Ghahramani 2000; Haggard 2005). The predictive process uses the efference copy of the current motor commands to generate expectations of action consequences. A mismatch between the prediction and actual sensory feedback can disrupt the otherwise fluent SoA. On the other hand, the reconstructive mechanism emphasizes that SoA arises from retrospective explanations of sensory feedback after movement (Wegner and Wheatley 1999; Wegner 2003; Buehner and Humphreys 2009). This inferential sensemaking process

evaluates the action feedback and its contingency with prior intentions and goals, reconstructing the causal links between them. While both mechanisms depend on the processing of sensory feedback, they differ in predictive aspects of motor control: the prospective mechanism necessitates the forward model of motor control, i.e. the sensory prediction of action consequence, while the reconstructive mechanism does not rely on the forward model but necessitates post-movement inferential processing of action feedback.

Previous neuroimaging studies typically modulate the magnitude of SoA by either manipulating the authorship of the action, e.g. externally moving people's effector to generate passive "actions" (Balslev et al. 2006; Tsakiris et al. 2010; Kühn et al. 2013; Straube et al. 2017; van Kemenade et al. 2017, 2019; Uhlmann et al. 2020; Zapparoli et al. 2020), or perturbing the sensory feedback of the movement or its outcome by implementing temporal and spatial discrepancies (Farrer and Frith 2002; Farrer et al. 2003, 2008; Leube et al. 2003a, 2003b; Matsuzawa et al. 2005; Balslev et al. 2006; David et al. 2007; Schnell et al. 2007; Spengler et al. 2009; Yomogida et al. 2010; Nahab et al. 2011; Chambon et al. 2013; Kühn et al. 2013; de Bezenac et al. 2016; Sasaki et al. 2018; Kikuchi et al. 2019; van Kemenade et al. 2019; Di Plinio et al. 2020; Ohata et al. 2020; Uhlmann et al. 2020; Zapparoli et al. 2020). Such contrasts between voluntary action and perturbed action revealed neural correlates of SoA in extensive cortical areas such as frontal, parietal, temporal, and insula cortices and

subcortical regions such as the cerebellum and striatum (Haggard 2017; Seghezzi et al. 2019; Charalampaki et al. 2022).

Temporal binding (TB), adopted by many as an indicator of implicit SoA (Haggard et al. 2002; Moore and Obhi 2012; Haggard 2017; Tanaka et al. 2019), refers to the fact that people's timing judgment of an action (e.g. a key press) and its delayed outcome (e.g. a beep sound or flash) are biased toward each other whenever the movement is voluntary as compared to involuntarily made (e.g. the finger pushed by others or triggered by transcranial magnetic stimulation (TMS; Haggard et al. 2002). Studies on the neural substrate underlying TB have highlighted the activity in a brain network including the supplementary motor area (SMA; Kühn et al. 2013), pre-SMA, and dorsal parietal cortex (Seghezzi and Zapparoli 2020; Zapparoli et al. 2020). In fact, modulating the activity over the pre-SMA by repetitive TMS selectively at the motor planning phase affects the binding effect (Zapparoli et al. 2020), with similar findings by the use of transcranial direct current stimulation and theta-burst TMS (Moore et al. 2010; Cavazzana et al. 2015). Given that the SMA and pre-SMA are crucial for preparing and initiating spontaneous actions (Fried et al. 1991; Cunnington et al. 2003), these findings have been used as neural support for the prospective mechanism of SoA.

Behavioral studies, however, highlighted that motor planning and execution are not necessary for TB, given that it can be elicited without voluntary action (Buehner and Humphreys 2009; Buehner 2012; Poonian and Cunnington 2013; Dewey and Knoblich 2014; Poonian et al. 2015; Borhani et al. 2017; Kong et al. 2017; Vastano et al. 2018; Suzuki et al. 2019). The TB effect can be generated by merely observing another human's or even a machine's causal action (i.e. a key press), while observing a non-causal event (i.e. a visual flash) could not (Buehner 2012). These observationelicited binding effects thus support the reconstructive mechanism, which conceptualizes SoA as a consequence of post-hoc inference after movements (Wegner and Wheatley 1999; Wegner 2003; Buehner and Humphreys 2009; Desantis et al. 2011). Given the behavioral evidence, many researchers propose that both predictive and retrospective processes contribute to the manifestation of SoA (e.g. Moore and Obhi 2012). However, the neural evidence supporting observation-elicited implicit SoA and thus the reconstructive mechanism is currently lacking.

Here we used virtual reality (VR) to modulate people's SoA, which is measured by a modified TB task without requiring them to execute movements, and examined whether its neural correlates were specifically tied to the inferential processing of action feedback rather than to motor planning and execution. Our recent behavioral study showed that after controlling an avatar in a first-person perspective in VR for a brief period, people increased TB when passively observing an avatar's "action" (Kong et al. 2017). This "embodiment" effect was thus caused by the prior experience of controlling the avatar since the TB was unchanged for people who experienced the identical VR environment but without controlling the avatar. Hence, such a VR setting would allow us to modulate the implicit TB and reveal its related neural changes when no voluntary action is engaged. We hypothesized that if the change in binding involved motor processes, we should find its neural correlates in sensorimotor regions, especially those planning areas (e.g. pre-SMA and SMA proper) implicated in motor intention and planning (Sperduti et al. 2011; Seghezzi et al. 2019). Alternatively, if post-movement inferential processes contributed heavily to the binding effect, we should observe its neural correlates in the regions outside the frontal motor areas. The targeted areas included posterior parietal areas that had been attributed to causal inference and action awareness (Wende et al. 2013;

Renes et al. 2015; Haggard 2017). In particular, inferior parietal regions deserved special attention since direct stimulation of these regions induced subjective experiences of intending to move or even increased (false) reports of movements that were not objectively measured (Desmurget et al. 2009).

Another venue of our study was that our VR manipulation enabled us to examine the neural basis of SoA over a virtual body. VR experience could change people's self-consciousness (Slater et al. 2009; Banakou and Slater 2014), SoA included (Banakou and Slater 2014; Kokkinara et al. 2016; Padrao et al. 2016; Nierula et al. 2021). However, previous neuroimaging studies focused on the sense of bodily ownership (Bach et al. 2012; Bekrater-Bodmann et al. 2014; Pamplona et al. 2022) and self-localization (Ionta et al. 2011; Lenggenhager et al. 2011). The neural substrate underlying SoA over a virtual body is still understudied (but see Nahab et al. 2011; Padrao et al. 2016; Limanowski et al. 2017, 2018), especially for the TB effect. Furthermore, previous neural studies on SoA over virtual body typically contrast conditions with different levels of spatiotemporal mismatch between virtual and actual actions. Our paradigm, instead, enables us to examine the neural correlates of embodying an avatar by contrasting before and after a VR experience.

# Materials and methods Participants

Our study recruited 48 college students as paid volunteers. Half of the participants were randomly assigned to the experimental group and half to the control group. Both groups were exposed to a VR environment, but only the experimental group viewed an avatar hand in VR. Three participants from the control group were excluded from data analysis, one for excessive head motion (>2 mm maximum translation or 2° rotation), and two for technical failure (details in Procedures), leaving 24 participants in the experimental group (age:  $M \pm SD = 23.57 \pm 2.39$  yr, 12 females) and 21 participants in the control group (age:  $M \pm SD = 22.18 \pm 2.76$  yr, 13 females). Power analysis was conducted based on the reported effect size in our previous study with a similar design (Cohen's f = 0.5 for the interaction effect in a two-way mixed analysis of variance [ANOVA]; Kong et al. 2017), and indicated that a sample size of n = 14 per group would lead to a power of 0.9 with an  $\alpha$  level of 0.05 (G\*Power 3.1; Faul et al. 2007). Thus, despite the data loss of three participants, we had enough participants for detecting possible group effects. The two groups were matched on age ( $t_{43} = 1.81$ , P = 0.078) and gender ( $\chi_1^2 = 0.643$ , P = 0.423). All participants were right-handed, had normal or corrected-to-normal vision, and reported no neurological diagnoses. The experiment was conducted according to the principles of the Declaration of Helsinki and was approved by the Ethical Committee of the School of Psychological and Cognitive Sciences at Peking University.

#### Procedures

#### Experimental procedure overview

Each participant went through three consecutive phases of the experiment, i.e. a pre-test, VR exposure, and a post-test. The pretest and post-test were carried out in the magnetic resonance imaging (MRI) scanner (each lasting ~35 min), in which participants performed the modified TB task (see "Temporal binding task" below) and a hand laterality judgment task (Ferri et al. 2012) in a sequel; The laterality task was designed to study questions unrelated to the purpose of the current investigation and was not reported here. After the pre-test, the participants walked into the waiting room next to the scanning room to receive VR exposure for ~30 min. During VR exposure, participants wore a head-mounted display (HMD, HTC Vive Pro) and a motion-tracking glove (Noitom Hi5 VR Glove) to perform four gamified motor tasks (see "VR exposure" below). The experimental group could view an avatar hand, whose motion spatially and temporally matched with that of their actual right hand; by controlling the avatar hand for these goal-directed movements, VR exposure would enable participants to embody the avatar hand. The control group was never given a chance to see the avatar hand and performed the same motor tasks. After finishing the VR exposure, the participants removed the HMD and the motion-tracking glove, and walked with their eyes closed to the scanning room under the guidance of the experimenter. They were instructed not to open their eyes until they were properly positioned in the scanner to get ready for the post-test. This procedure was employed to minimize the visual experience of the real settings and to preserve the effect of VR exposure. After the post-test, participants re-entered the waiting room and were asked to evaluate their subjective sense of embodiment during the post-test by means of questionnaires.

#### Temporal binding task

The TB task was a replicate of the same task in our previous study, though it had been performed with the HMD earlier (Kong et al. 2017). In brief, participants judged the timing of an auditory stimulus with the aid of a Libet clock projected in the MRI scanner in keeping with the TB task previously performed outside the VR (Haggard et al. 2002).

The task involved temporal judgments of tones in two types of trials, baseline and operant trials. For each operant trial (top panel in Fig. 1a), the Libet clock started to rotate clockwise from a random location. After a random interval of 2,560-5,120 ms, the right avatar hand pressed the white button. A tone would be presented 250 ms later (100 ms in duration). Note that the participant was required to refrain from any movement during the stimulus presentation. The clock hand kept rotating after the tone for a random duration of 1,280–2,560 ms, then reset its position to 12 o'clock. Participants were required to report the location of the clock hand when the tone was perceived by pressing the left and right keys using the left middle and index finger to move the clock hand clockwise and counterclockwise, respectively. They then confirmed their estimated position by pressing the key with the right index finger. The report was self-paced. The next trial started upon the confirmation. The baseline trial (bottom panel in Fig. 1a) was identical to the operant trial, except that the tone was presented without the button press, i.e. the avatar's hand remained static all the time. The mean trial duration was 12.06 s (SD: 2.98 s). Each run contained 20 trials, lasting ~6–8 min. Both the pre-test and the post-test were comprised of two operant runs and two baseline runs, whose orders were randomized and counterbalanced between participants.

Participants familiarized themselves with the TB task before the formal experiment by conducting four operant trials and four baseline trials outside the scanner. Due to technical failure, two control participants underwent one operant run and three baseline runs in one of the two phases, and were excluded from data analysis.

All the visual stimuli were back-projected onto a translucent screen located inside the scanner (resolution:  $1,024 \times 768$ ; refresh rate: 60 Hz; viewing distance ~90 cm; Fig. 1a). The avatar hand was the one that the participants visually controlled during the VR exposure outside the scanner but fitted for rendering in the two-dimension display in the scanner. The clock had a 10-pixel-long hand, which rotated with a period of 2,560 ms per circle.

The clock face (radius = 110 pixels) was marked with conventional "1 min" intervals. Th30.4(f)-.9.1(0)-393-1.405TTci50e<sup>A</sup> tone3087.4(t).the

## a Temporal binding task in the Pre-test and Post-test phases

Operant trial



**Fig. 1.** Illustrations of behavioral tasks and the observed TB effects. (a) Graphical illustration of an operant trial (top panel) and a baseline trial (bottom panel) in the TB task. In both conditions, participants were required to report the location of the clock hand when the tone was perceived. Throughout the baseline trials, the avatar hands on the screen kept stationary. But in the operant trials, its right index finger "pressed" a white button 250 ms prior to the tone. Participants' real hands kept unmoved before making temporal judgments. (b) Scenes of motor tasks in the VR exposure phase. Participants performed four VR motor tasks: The gesture-imitation task: Bending right-hand fingers to match a target gesture, shown by the distant avatar hand; the bubble-poking task: Poking the bubble with the right index finger; the cube-picking task: Picking target cubes that are specified by color or shape in instructions; the pad-tapping task: Memorizing a multi-digit number before it disappears and then recalling it by tapping on a keypad. The experimental group viewed the avatar hand (shown here), but the control group did not. (c) TB effects were quantified by perceptual shifts evoked by the avatar's movement. The average binding effects in the pre-test and post-test are shown for the two groups separately. Error bars indicate standard error.

hand illusion (Botvinick and Cohen 1998; Slater et al. 2008; Braun et al. 2014; Kalckert and Ehrsson 2014; Ma and Hommel 2015). The items were designed to assess explicit SoA and sense of ownership (SoO) or to control for possible response biases by using the reversed control items (Table S1).

### **Behavioral analysis**

To examine whether the VR exposure could enhance participants' implicit SoA associated with an avatar hand, we compared the

TB effect using a two-way mixed-design ANOVA with Phase as the within-group factor (pre- versus post-test) and Group as the between-group factor (experimental versus control group). Each trial yielded a perceptual error of temporal judgment, quantified as the difference between the reported and the actual onset of the tone. The TB effect was operationally defined as the difference in perceptual error, i.e. a perceptual shift, between the operant condition and the baseline condition. A negative perceptual shift indicated the TB effect.

The present study focused on the implicit SoA, measured using the behavioral TB effects mentioned above, not the explicit measures of embodiment from questionnaires, as our previous behavioral study showed that explicit SoA did not change after brief VR exposure (

#### Table 1. Summary of the ROI-level results.

Location		Coordinates  (MNI, mm)			Phase × Group 				Correlation with binding changes			
	Н								ALL		EXP	
Area		x	у	Z	F <sub>1, 43</sub>	р	Cohen's f	BF <sub>incl</sub>	r <sub>43</sub>	р	r <sub>22</sub>	р
Literatu	e-based	l ROIs										
SMA	L	-7	-4	69	0.11	0.745	0.050	0.302 <sup>H0</sup>	0.05	0.737	-0.18	0.401
Insula	L	-41	2	1	0.13	0.716	0.056	0.314 <sup>H0</sup>	-0.09	0.566	-0.36	0.083
CAL	R	18	-90	-1	< 0.001	0.993	0.001	0.325 <sup>H0</sup>	0.08	0.607	0.07	0.724
CE	R	24	-53	-27	0.08	0.783	0.042	0.294 <sup>H0</sup>	0.03	0.827	-0.11	0.601
IPL	L	-46	-48	51	0.58	0.450	0.116	0.407	-0.20	0.195	-0.37	0.077***
STG	R	54	-49	22	0.39	0.538	0.095	0.349	-0.06	0.697	-0.29	0.168
AG	R	45	-60	43	7.67	0.008**	0.422	13.546	-0.25	0.092***	-0.36	0.087***
Task-ba	sed ROI											
IPL	R	50	-44	44	5.29	0.026*	0.351	3.708	-0.35	0.017*	-0.44	0.029*

Notes: H = Hemisphere, L = left, R = right; SMA = Supplementary motor area, Insula = Posterior insula, CAL = Calcarine scissure, IPL = Inferior parietal lobule, AG = Angular gyrus, STG = Superior temporal gyrus, CE = Cerebellum; <sup>H0</sup>: moderate evidence in support of the null hypothesis of no interaction effect (BF<sub>incl</sub> < 0.33). \*0.01 < P < 0.05. \*\* P < 0.10.

quantified by subtracting the whole-brain contrast images of the

pre-test (operant versus baseline) from the post-test (operant

versus baseline) for each participant. A two-sample t-rsu1.404208DC65.5(A.4318e))-476.6(f)-.5(or)-4761Tm211.2(h).5(7f2.82/F3thh).5(7f2.8.ma)dthe



Downloaded from https://academic.oup.com/cercor/article/34/2/bhad547/7607165 by guest on 07 May 2024

# **a** Bindin



(Khalighinejad et al. 2016), pre-SMA with intentionality (Yomogida et al. 2010; Zapparoli et al. 2018, 2020), and SMA proper with motor planning and initiation (Kühn et al. 2013; Passingham and Lau 2019). Consistent with the rationale of our task design, all these prefrontal and frontal areas, related to the generation of action before action feedback, returned a null effect in our data (Fig. 1a and Table 1), except the right AG identified by the meta-analysis based on voluntary action-related SoA (Seghezzi et al. 2019).

The TB change elicited by observing an embodied action is not only specifically associated with the right AG, but also associated with a cluster extended to the right IPL. Both are not directly tied to motor planning and movement initiation. This finding supports the reconstructive mechanism of SoA, which emphasizes that the TB is grounded by retrospective causality since the right AG and IPL participate in both SoA and causal processing in general. The right IPL has been one of the most frequently revealed neural correlates of SoA (Farrer et al. 2003, 2008; Schnell et al. 2007; Nahab et al. 2011; Chambon et al. 2013, 2015). Even anosognosia patients who often assert that they performed an action with their paralyzed, immobile limb typically have lesions in the right parietal lobule (Fotopoulou et al. 2008). More importantly, the right IPL is broadly involved in causal processing since the explicit judgment of both physical and social causality relates to neural activations in the right IPL, along with other areas (Wende et al. 2013; Renes et al. 2015). Even seeing a causal event, such as an object collision, elicits more activations in the right IPL than seeing a non-causal event, such as an object launching (Fugelsang et al. 2005).

The AG, similarly implicated by numerous SoA studies, engages in diverse cognitive tasks that require inferential sensemaking. For SoA, meta-analyses have shown that the TPJ, with the right AG included, is related to attributing SoA to others (external SoA, Sperduti et al. 2011) and to the reduction of self-agency (negative SoA, Zito et al. 2020). A recent review also finds AG as a common node for encoding motor intention and SoA (Seghezzi et al. 2019). Beyond agency tasks, the AG has been reliably shown to engage in a wide range of tasks, including reasoning, semantic processing, word reading and comprehension, memory retrieval, attention and spatial cognition, default mode network, and social cognition. A well-received unified theory about the AG's function, based on the commonality of these tasks, highlights its role in sensemaking, i.e. giving meaning to external sensory information or internal thoughts (Seghier 2013). For instance, the AG engages in the comprehension of speech and written languages (Xu et al. 2005; Obleser and Kotz 2010), especially in solving referential ambiguity (Nieuwland et al. 2007). It also engages in inferring human intention in the theory of mind tasks (Mason and Just 2011). Given its rich anatomical connectivity to widely distributed brain regions, the AG appears suitable for combining diverse information, linguistic and nonverbal (e.g. body movements), prior knowledge (experiences, context, and purpose), and new sensory information, to converge toward plausible accounts of the events. This sensemaking process can be implemented as an active optimization process that combines bottom-up information (i.e. sensory information) with top-down predictions (i.e. prior knowledge and purpose) to minimize surprise according to the free energy principle (Friston 2010). Pertinent to our findings here, the AG is a central region for the inferential sensemaking process in various tasks, among which the agency-related task is an important genre since SoA sets the boundary between self and the external environment (Seghier 2013, 2022).

The involvement of AG and IPL in our observation-based TB is in line with the reconstructive mechanism of SoA (Wegner and Wheatley 1999; Wegner 2003; Buehner and Humphreys 2009;

Desantis et al. 2011; Tramacere 2022). Our findings, of course, should not be taken as evidence against the importance of the prospective processing for SoA during voluntary action. There exists extensive behavioral and neural evidence that both the prospective motoric process in motor planning and the retrospective process in outcome evaluation contribute to SoA, though their relative importance depends on available cues and task goals (Moore and Obhi 2012; Synofzik et al. 2013). Even AG, the region we identified as crucial for retrospective processing of SoA, has been shown to monitor signals related to action selection in the dorsolateral prefrontal cortex when participants are required to explicitly report their SoA (Chambon et al. 2013, 2015). Our findings highlight that the brain can indeed invoke SoA-related processing retrospectively when no action is involved.

In line with the widely reported dissociation between explicit SoA and TB (Buehner 2012; Dewey and Knoblich 2014; Lynn et al. 2014; Saito et al. 2015; Wen et al. 2015; Kong et al. 2017; Kirsch et al. 2019; Ma et al. 2019, 2021; Schwarz et al. 2019; Suzuki et al. 2019), our findings can also be viewed as a challenge to the validity of treating TB as an implicit measurement of SoA. TB with voluntary actions indeed changes according to SoA manipulations, including the aforementioned experimental comparisons between active and passive movements and between congruent and incongruent action feedback. However, TB can also be elicited without action and supported by distinct neural substrates, as shown here. Thus, a parsimonious account of TB posits that it results from top-down causal belief about the timing of sequential events, with or without voluntary action (Hoerl et al. 2020). The belief is about the causal relationship between a movementrelated event, not necessarily an intentional action, and a subsequent outcome event. The causal belief is subject to influence from priming, instruction, statistical contingency, and prior belief, which all have been shown to affect the TB (Wegner 2003; Aarts et al. 2005; Moore et al. 2009; Ebert and Wegner 2010; Desantis et al. 2011; Ma et al. 2019). The causal account thus views TB as a general phenomenon in timing perception and casual belief, beyond a reflection of implicit SoA that has been argued to embed in the motor system. This view resembles the reconstructive mechanism of SoA with its emphasis on inferential processing for sensory events. In this light, the VR experience with the avatar might strengthen the internal model, thus facilitating the formation of causal belief between avatar "action" and subsequent sensory outcome. The AG and IPL underlie the VR binding effects, and might play a role in representing the causal belief. Though risking the curse of reverse inference from neural findings to cognitive processes, our findings support the causal account by showing that the neural substrate underlying our observed VR binding effect involves AG and IPL, important areas supporting the causal inference of sensory events.

Though a quantitative model of SoA is currently lacking, various aspects of TB have been accounted for by probabilistic inference models based on Bayesian cue combination (Moore and Fletcher 2012; Wolpe et al. 2013; Legaspi and Toyoizumi 2019; Lush et al. 2019). The temporal shift of the action and the action outcome are modeled as resulting from optimal estimates of their specific timing when relevant sensory cues and prior expectations are integrated according to causality between cues. Specifically, the shifts occur only when the "action" is inferred as causal for the subsequent effect (Legaspi and Toyoizumi 2019). In computational terms, the binding builds upon a prior belief of a causal relationship and the sensory evidence of related timing cues, independent of whether intentional action is involved. From the perspective of the Bayesian model, our increased binding of the outcome event can be viewed as reflecting an enhanced prior belief of the causal relationship between the avatar movement and the subsequent beep tone. Both our VR and control groups received identical sensory feedback in the TB task, and the only difference is that the VR group had prior experience visually controlling the avatar before the post-test. The embodiment of the avatar is thus akin to an enhanced prior belief that the avatar hand is responsible for the outcome (Desantis et al. 2011; Haering and Kiesel 2012), which leads to an increased timing shift according to the Bayesian model of TB (Legaspi and Toyoizumi 2019). In fact, a similar Bayesian model based on causal inference also explains the sense of bodily ownership, another component of self-consciousness, as investigated in the classical rubber hand illusion paradigm (Chancel et al. 2022a). More importantly, causal beliefs about relevant ownership cues, estimated from this paradigm, are implicated in the IPS, a region often associated with cue combination, as well as the AG (Chancel et al. 2022b). These modeling and neuroimaging work thus suggest that classical measures of self, i.e. the rubber hand illusion in the sense of bodily ownership and the TB in the SoA, might be governed by the same causal inference mechanism with the involvement of IPL and AG.

Previous studies on VR embodiment have largely focused on how multisensory integration affects people's self-consciousness (Slater et al. 2009; Banakou and Slater 2014). With a brief exposure to VR, people erroneously feel that they own a virtual body part or even a full virtual body (Petkova et al. 2011; Blanke et al. 2015), mislocate themselves (Ehrsson 2007), or change the perception of one's identity (Petkova et al. 2011; Banakou et al. 2013). The common technique is to present a vivid visual representation of an avatar and match it spatiotemporally with sensory cues from other modalities, including tactile, auditory, and proprioceptive cues (Slater et al. 2009; Banakou and Slater 2014). Neuroimaging studies have shown that the premotor cortex and TPJ are key areas for bodily ownership (Bekrater-Bodmann et al. 2014; Pamplona et al. 2022) and self-location (Ionta et al. 2011; Lenggenhager et al. 2011). However, the neural correlate of SoA over a virtual body is understudied. Existing studies typically manipulated spatiotemporal mismatch between avatar and actual action (Nahab et al. 2011; Limanowski et al. 2017) as in other embodiment studies, say, on bodily ownership. Interestingly, the neural correlates to these parametrical modulations of SoA (not necessarily about the degree of SoA) also include IPL, along with other regions like STS (Limanowski et al. 2017). Our study differed from these studies by showing that sensorimotor control experience with an avatar can lead to subsequent SoA changes over the avatar movement, whose neural correlates center at the right AG and IPL, key areas that are also associated with SoA arising from actions in real settings. Given this cluster covers high-order associative regions, we postulate that the VR embodiment effect is potentially generalizable to other tasks beyond the TB. For instance, SoA arising from voluntary actions contributes to perceptual attenuation of action-induced sensory stimuli (Blakemore et al. 1998; Shergill et al. 2005) or self-other distinction (Kahl and Kopp 2018). Whether these perceptual tasks are affected by similar avatarcontrol experiences in VR warrants further investigation.

Our findings raised possible problems for the era of VR or metaverse. First, despite the fact that our participants did not change their self-reported SoA rating with the brief VR experience (see Supplementary Materials), it is still possible that people's explicit judgment of SoA can be modulated by long-term VR use. Second, individuals with neurological and psychiatric disorders experience disrupted SoA and illusions in their daily lives (Frith et al. 2000; Edwards and Bhatia 2012), and even neurotypical individuals can occasionally experience faulty SoA, say, with sensory priming (Wegner and Wheatley 1999; Aarts et al. 2005). Whether certain populations' self-consciousness is negatively affected by the experience of controlling an avatar is an important open question from the perspective of psychopathology. Third, given the observed immediate behavioral and neural effect of an embodied avatar on SoA, we expect that unintended "actions" of the avatar, accidentally caused by technical glitches in the virtual worlds, might affect the avatar owner's SoA and even lead to psychological harm (Cheong 2022). These previously rare scenarios might lead to potential legal issues about how to account for the responsibility of compromising someone's SoA in the metaverse.

In conclusion, the TB elicited by passive observation of an embodied virtual body is subserved by the right AG and IPL, regions related to causal inference and inferential sensemaking but not directly related to motor control. In contrast, traditional motor planning areas (e.g. pre-SMA), widely observed in studies on the SoA arising from voluntary actions, are not implicated. These findings support the reconstructive mechanism of SoA that emphasizes retrospective processing of SoA-related cues and suggests that the experience of controlling an avatar might enhance the causal belief of avatar action and its action outcome, leading to increased TB. Our behavioral and fMRI results also questioned the validity of using TB as a measure of implicit SoA. Furthermore, given that people's causal belief over an avatar's action can be built up, both behaviorally and neurophysiologically, by a brief period of avatar control in immersive VR, how the embodiment of a virtual body affects our self-consciousness and other psychological constructs would pose as a novel problem when our populace spends increasing time in virtual or digital worlds.

### Acknowledgments

We thank Chunfang Yan and Dr Weiwei Men for the help with the VR fMRI experiment settings.

### Author contributions

Yiyang Cai (Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Huichao Yang (Formal Analysis, Funding acquisition, Methodology, Validation, Writing—review & editing), Xiaosha Wang (Formal Analysis, Methodology, Validation, Writing—review & editing), Ziyi Xiong (Formal Analysis, Methodology, Validation, Writing—review & editing), Simone Kühn (Conceptualization, Funding acquisition, Writing—review & editing), Yanchao Bi (Methodology, Supervision, Writing—original draft, Writing—review & editing), and Kunlin Wei (Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing—original draft, Writing—review & editing).

### Supplementary material

Supplementary material is available at Cerebral Cortex online.

# Funding

This work was supported by STI2030-Major Projects (2021ZD 0202600) and the National Natural Science Foundation of China

12 | Cerebral Cortex

Fried I, Katz A, McCarthy G, Sass KJ, Williamson P, Spencer SS, Spencer DD. Functional organization of human supplementary

- Obleser J, Kotz SA. Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb Cortex*. 2010:20(3):633–640.
- Ohata R, Asai T, Kadota H, Shigemasu H, Ogawa K, Imamizu H. Sense of agency beyond sensorimotor process: decoding self-other action attribution in the human brain. *Cereb Cortex*. 2020:30(7): 4076–4091.
- Padrao G, Gonzalez-Franco M, Sanchez-Vives MV, Slater M, Rodriguez-Fornells A. Violating body movement semantics: neural signatures of self-generated and external-generated errors. NeuroImage. 2016:124(Pt A):147–156.
- Pamplona GS, Salgado JA, Staempfli P, Seifritz E, Gassert R, Ionta S. Illusory body ownership affects the cortical response to vicarious somatosensation. *Cereb Cortex*. 2022:32(2):312–328.
- Passingham RE, Lau HC. Acting, seeing, and conscious awareness. Neuropsychologia. 2019:128:241–248.
- Petkova VI, Björnsdotter M, Gentile G, Jonsson T, Li T-Q, Ehrsson HH. From part- to whole-body ownership in the multisensory brain. *Curr Biol.* 2011:21(13):1118–1122.
- Poonian SK, Cunnington R. Intentional binding in self-made and observed actions. Exp Brain Res. 2013:229(3):419–427.
- Poonian SK, McFadyen J, Ogden J, Cunnington R. Implicit agency in observed actions: evidence for N1 suppression of tones caused by self-made and observed actions. J Cogn Neurosci. 2015:27(4): 752–764.
- Renes RA, Van Haren NEM, Aarts H, Vink M. An exploratory fMRI study into inferences of self-agency. Soc Cogn Affect Neurosci. 2015:10(5):708–712.

- Yomogida Y, Sugiura M, Sassa Y, Wakusawa K, Sekiguchi A, Fukushima A, Takeuchi H, Horie K, Sato S, Kawashima R. The neural basis of agency: an fMRI study. *NeuroImage*. 2010:50(1): 198–207.
- Zapparoli L, Seghezzi S, Scifo P, Zerbi A, Banfi G, Tettamanti M, Paulesu E. Dissecting the neurofunctional bases of intentional action. Proc Natl Acad Sci. 2018:115(28):7440–7445.
- Zapparoli L, Seghezzi S, Zirone E, Guidali G, Tettamanti M, Banfi G, Bolognini N, Paulesu E. How the effects of actions become our own. Sci Adv. 2020:6(27): eaay8301.
- Zito GA, Wiest R, Aybek S. Neural correlates of sense of agency in motor control: a neuroimaging meta-analysis. *PLoS One.* 2020:15(6):e0234321.