# scientific reports



# **OPEN** The ambiguous feeling between "mine" and "not-mine" measured by integrated information theory during rubber hand illusion

Takayuki Niizato<sup>1,4<sup>\ZI</sup></sup>, Yuta Nishiyama<sup>2,4<sup>\ZI</sup></sup>, Kotaro Sakamoto<sup>3</sup>, Takumi Kazama<sup>2</sup>, Tatsuya Okabayashi<sup>2</sup> & Taiki Yamaguchi<sup>2</sup>

Human body awareness is adaptive to context changes. The illusory sense of body ownership has been studied since the publication of the rubber hand illusion, where ambiguous body ownership feeling was first defined. Phenomenologically, the ambiguous body ownership is attributed to a conflict between feeling and judgement: it characterises a discrepancy between first- and thirdperson processes. Although Bayesian inference can explain this malleability of body image, it still fails to relate the subjective feeling to physiological data. This study attempts to explain subjective experience during rubber hand illusions by using integrated information theory (IIT). The integrated information  $\Phi$  in IIT measures the difference between the whole system and its subsystems. By analysing seven different time-series of physiological data representing a small body-brain system, we demonstrate that the integrity of the whole system during the illusion decreases, while the integrity of its subsystems increases. These general tendencies agree with many brain-image analyses and subjective reports; furthermore, we found that subjective ratings as ambiguous body ownership were associated with  $\Phi$ . Our result suggests that IIT can explain the general tendency of the sense of ownership illusions and individual differences in subjective experience during the illusions.

The malleability of human body image is a manifestation of the flexible perceptual system. This malleability has been demonstrated in several experiments where the body awareness is altered by multisensory stimuli. One famous example is the rubber hand illusion (RHI) introduced by Botvinick and Cohen<sup>1</sup>, where a subject observes a brush stroking a dummy hand while the subject's real hand, which is hidden behind a partition, is also stroked. Synchronised visuo-tactile stimuli lead to an illusory sense of ownership of the dummy hand. The body ownership of the dummy body parts can be extended to the whole body. For instance, Lenggenhager et al. showed that a person who observed a virtual body being touched while they were simultaneously also being touched experienced an out-of-body experience<sup>2</sup>. Interestingly, the stimulation of specific brain regions, specifically the temporo-parietal junction area, can evoke a similar experience<sup>3</sup>. This region is known to work in consonance with multiple body sensations<sup>4</sup>. Yet despite numerous studies conducted in this regard, a comprehensive understanding of the sense of ownership is still lacking. Recent phenomenological interpretations provide a concise scheme for understanding the origin of body ownership and agency<sup>5-8</sup>. Ataria summarised the type of feelingjudgement conflict to clarify the various states of body ownership9. Phenomenologically, feelings involve closed sensor-motor loops, whereas judgements are more abstract reasoning processes. In summary, feelings provide a first-person perspective, i.e., a non-conceptual, pre-reflective, bottom-up process; by contrast, judgement gives a third-person perspective, i.e., an observational, reflective, top-down process. According to this argument, during an RHI, subjects feel the dummy hand as their hand but never admit (or judge) it as their hand. This conflict between feeling (i.e., 'this is my hand') and judgement (i.e., 'this is not my hand') causes ambiguous body

<sup>1</sup>Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan. <sup>2</sup>Department of Information and Management Systems Engineering, Nagaoka University of Technology, Niigata, Japan. <sup>3</sup>The Institute of Statistical Mathematics, Research Center for Statistical Machine Learning, Tokyo, Japan. <sup>4</sup>These authors contributed equally: Takayuki Niizato and Yuta Nishiyama. <sup>Semail:</sup> t\_niizato@yahoo.co.jp; y-nishiyama@kjs.nagaokaut.ac.jp

ownership (ABO) rather than a specific sense of body ownership<sup>7</sup>. The fact that the results of the two processes are not necessarily consistent motivated us to consider that malleability of body image originates during the mismatched feeling-judgement process.

The free-energy principle (FEP; or Bayesian inference) is a promising theory for understanding the flexibility of our perceptual system<sup>10</sup>. According to the FEP, we can change our beliefs whenever confronted with unexpected situations (i.e., high saliency). The perceptual system updates prior hypotheses with the preferable posterior hypotheses. The FEP suggests that appropriately adjusting the hypotheses in response to the environment corresponds to the malleability of body image. The RHI is, therefore, related to the corruption of sensorimotor predictive cycles. Phenomenologically, this sensor-motor prediction error indicates a corruption of the feeling system. Most studies do not proceed further than this interpretation of corruption at the feeling level; however, as discussed earlier, the corruption of the feeling system does not provide an adequate explanation for ABO because we must also consider the top-down judgement level.

Tsakiris et al. investigated the interaction between bottom-up and top-down processes through interoceptive and exteroceptive perceptions<sup>11-13</sup>. They showed that the intensity of RHI is due to the mismatch between interoceptive and exteroceptive predictions. In their experiments, subjects who predicted their heartbeats more precisely experienced less illusion of ownership during the RHI than subjects who did so less precisely<sup>11</sup>. They speculated the existence of an interdependent relationship between interoceptive and exteroceptive predictions, i.e., that high interoceptive saliency suppresses exteroceptive saliency, resulting in ABO<sup>14</sup>.

Although Tsakiris<sup>14</sup> proposed a remarkable relationship between RHI subjectivity and FEP, FEP per se is still fraught with a difficulty: specifically, it assumes that judgement is a process for selecting the highest saliency among possible saliencies. As discussed in Ransom et al.<sup>15</sup>, these saliency selections cannot explain human affective behaviour for low saliency. This limitation implies that the saliency-driven explanation is insufficient for human experience. If we accept the saliency selection, then we agree that there is no conflict between feeling and judgement, because the judgement system mechanically picks up the highest saliency. In this case, the intensity of RHI is ascribed only to the feeling and judgement. In this case, we neglect the essential difference between feeling and judgement owing to the compression of two different systems into one sensor-motor system.

Integrated information theory (IIT) has been proposed to mathematically define human consciousness<sup>16-19</sup>. The basic concept of IIT is summarised in the proposition that the degree of consciousness can be measured as the difference between the whole system and the sum of its parts; that is, the irreducibility of the whole system to its parts is the key to understanding the experience of consciousness<sup>16,17</sup>. Notably, IIT also considers both feeling and judgement in the perceptual process. Feeling corresponds to the local process and represents the information process of the subsystem (i.e., the subsystems of the system). By contrast, judgement corresponds to the global process and represents the information process of the whole system. Importantly, IIT rejects any external evaluations as an explanation for consciousness. Therefore, there is no superiority between feeling and judgement is revealed only by subjective experience. Thus, IIT can mathematically evaluate body malleability induced by the conflict between feeling and judgement.

In a recent study that took physiological measurements during RHI, autonomic nervous activity mediating interoception did not change under an altered body–representation<sup>20</sup>. Thus, merely measuring autonomic activity may not be enough to reveal effects of both feeling and judgement on subjective experience during RHI. Instead, it would be effective to measure autonomic activity with central nervous activity mediating exteroception. We thus measured multiple physiological activities and examined its relationship with subjective experience using IIT. Our measurements took in both brain activities and internal bodily states, whereas IIT is usually applied only to the complex neural networks in the brain. We measured seven types of physiological data from the body (respiration, heartbeat, and skin conductance) and brain (electroencephalogram [EEG] with electrodes on midline frontal [Fz], vertex [Cz], midline parietal [Pz], and midline occipital [Oz] regions). Notably, affective stimuli such as apparent threat to the rubber hand evoking transient physiological responses<sup>21–26</sup> were not applied because we were interested in a wide range of physiological changes over time associated with the illusory body ownership during the classical RHI. Hence, we focused on these data to explore whether the interaction between interoceptive (body-related responses) and exteroceptive (part of brain activity) processes reflected the subjective experience of ABO during RHI, as highlighted by Tsakiris<sup>14</sup>. We predicted that the degree of conflict between body and brain, rather than that of the whole brain process, is enough to estimate experiences during RHI.

#### Results

**Minimum information partition (MIP).** We examined the average tendency of MIP during the experiments. MIP gives two measures for estimating the system's state: (i) MIP cut, which divides the two subsystems by the weakest information flow; and (ii)  $\Phi_{\text{MIP}}$ , which is the degree of information loss by the MIP cut.  $\Phi_{\text{MIP}}$  is positive by definition; otherwise (i.e., if it were equal to 0), the two subsystems would be completely independent.

This study used seven datasets: respiration (RES), heartbeat (ECG), and skin conductance (EDA) for body, and EEG from midline frontal (Fz), vertex (Cz), midline parietal (Pz), and midline occipital (Oz) regions for brain, for IIT analysis. We considered that these body–brain signals work as a system ( $S = \{\text{Res, ECG, EDA, Fz, Cz, Pz, Oz}\}$ : seven node's all connected network). If the rubber hand stimulus causes a difference in the system, the relation among these seven datasets will change. This difference also reflects the MIP information.

Figure 1A shows the time series of the mean  $\Phi_{\text{MIP}}$  for all 22 subjects. The  $\Phi_{\text{MIP}}$  drops at the stimulus phase in both synchronous (SYNC) and asynchronous (ASYNC) conditions (Fig. 1B). This decrease suggests that connections between subsystems become weaker than during pre- and post-stimulus phases. This result is interesting because similar changes occur in both conditions.



**Figure 1.** (A) Time series of mean  $\Phi_{\text{MIP}}$  for both conditions (blue: SYNC, red: ASYNC). The stimulus phase covers 300–600 s. The other regions indicate resting states (no stimulus). The computational time frame for IIT is 250 (2 s). A moving-average (sample length 150 steps) was calculated in the time series to make temporal changes of values easy to see (Niigata)  $\Phi_{\text{MIP}}$  of SYNC and ASYNC values drops significantly compared with each rest state (\*p < 0.05, \* \*p < 0.01, \* \*p < 0.005).

Next, we examined the details of the MIP cut. The most remarkable tendency was the cut between {Oz} and {RES, ECG, EDA, Fz, Cz, Pz}. More than 50% of MIP cuts were of this type (Fig. 2A,B); furthermore, during the stimulus phase, the frequency of this cut increased statistically. Although its frequency increased, the  $\Phi_{MIP}$  of the {Oz}-cut decreased. The implication of this tendency is suggestive. It is a well-known fact that the occipital region (Oz) is involved in the visual information process. The high frequency of {Oz}-cut reflects the discrepancy between visual information and tactile information. The low  $\Phi_{MIP}$  during the stimulus means that the visual process in Oz diverges from other information processes (Fig. 2C).

In the remaining patterns, the MIP-cut frequencies decreased after the stimulus was applied (Fig. 2A,B). The decreased frequency of the rest compensates for the increased frequency of the {Oz}-cut. Interestingly, in the first case we observed, the cut was rarely located between the body and brain systems. But in most cases, the cut was located in the brain system. This fact suggests that the body and brain systems are not the weakest links. The second most frequent  $\Phi_{MIP}$  cut is located between {ECG} and {RES, EDA, Fz, Cz, Pz, Oz}. According to this cut, the heartbeat (ECG) merges into the cycle of the brain–body system during the stimulus. Even if {ECG} was selected as the MIP-cut, its integrity decreased during the stimulus. This aspect resonates with Tsakiri's result<sup>11</sup>.

Finally, we highlight the fact that  $\Phi_{\text{MIP}}$  for SYNC and ASYNC conditions during RHI showed a statistical difference between the two conditions observed (Welch's *t* test: df = 13159, t-stat = 2.2561, p < 0.05). This result seems interesting because half of the subjects in our study reported a change of body ownership during both conditions. Many of the individual differences were perceived not in nature but in degree. This logic also applies to the next subsection.

**Main complexes analysis.** We examined the average tendency of the main complexes under SYNC and ASYNC conditions. The concept of the main complex accompanies some other concepts; namely, (i)  $\Phi_{\text{MIP}}^T$ , which is the degree of information loss caused by partitioning the system into independent subsystems; the manner of partitioning system *T* is determined by MIP:  $\Phi_{\text{MIP}}^T$ ; (ii) complex *T*, where all subsystems satisfy  $\Phi_{\text{MIP}}^T > 0$  and  $\Phi_{\text{MIP}}^T > \Phi_{\text{MIP}}^R$  for any superset *R* of *T*; and (iii) main complex *M*, where the complex *M* satisfies  $\Phi_{\text{MIP}}^M > \Phi_{\text{MIP}}^U$  for any subset *U* of *M*.

Because  $\Phi_{\text{MIP}}$  represents the irreducibility to the system's parts, the high  $\Phi_{\text{MIP}}$  values indicate the system's integrity. Notably, the system may include several main complexes. For a subsystem to be the main complex, it is sufficient that it does not contain other complexes (Lemma 4 in Kitazono et al.<sup>27</sup>). In our analysis, the body-brain system has up to ten or more main complexes. The integrity of the system can be estimated as the sum of  $\Phi_{\text{MIP}}^M$  of all main complexes M in S (denoted  $\sum_{M \in \mathcal{M}} \Phi^M_{\text{MIP}}$  where  $\mathcal{M}$  is a set of main complexes of system  $S = \{\text{Res, ECG, EDA, Fz, Cz, Pz, Oz}\}$ ).

Figure 3A shows that  $\sum_{M \in \mathcal{M}} \Phi_{MIP}^M$  increases only in the stimulus phase. In contrast to the decreasing  $\Phi_{MIP}^S$  for the whole system *S*, the integrities of the particular subsystems increased (Fig. 3B). While the whole system splits into relatively independent parts, specific subsystems compensate by their integrity. This tendency can also be observed in the number of complexes. The number of irreducible subsystems (parts) increased during the stimulus phase (Fig. S4). Furthermore, in contrast to the  $\Phi_{MIP}^S$ , an after-effect was observed in the main complex. This means that fluctuations in the system integrities remained even after the stimulus had ended.

The content of the most vital main complex (i.e.,  $\max_{M \in \mathcal{M}} \{\Phi_{MIP}^M\}$ : the main complex M with the highest  $\Phi_{MIP}^M$ ) is more suggestive. Figure 4A shows that the frequency of the main complex containing all elements (or  $\Phi_{MIP}^S$ ) dropped during the stimulus phase. Here, the value of  $\Phi_{MIP}^S$  indicates that the body-brain unit itself is irreducible to its parts. Metaphorically, the oneness of the body-brain system was corrupted during the stimulus. After the stimulus phase, the frequency of this main complex recovered.

After the stimulus phase, the frequency of this main complex recovered. The other main complexes compensated for the decreasing frequencies of the complexes *S* (Fig. 4B,C). The main complex {EDA, RES} is the most significant, showing skin conductance and respiration. The  $\Phi_{MIP}^{\text{[EDA,RES]}}$  also increased significantly during the stimulus (Fig. 4C). The integrity only involves the body parts. Another high frequent main complex is {Fz, Cz, EDA, RES}, containing {Fz, Cz} pairs. The frequency of this complex increased



**Figure 2.** (A) Differences in the  $\pi_{\text{MIP}}$  frequencies of pre- and intra-stimuli:  $\text{Fr}(\pi_{\text{MIP}}|\text{ST}) - \text{Fr}(\pi_{\text{MIP}}|\text{PRE})$ . The computational time frame for IIT is 250 (1 s). The positive values indicate that a certain type of  $\pi_{\text{MIP}}$  increases during the stimulus. Blue (red) bars show the frequency differences in the synchronous (asynchronous) condition. Blue (red) letters show data from the brain (body) system. (B) Image of MIP-cut location of the system *S*. {Oz}-, {Fz}-, {ECG}-cut are the subsets with the three largest differences shown in (A). (C)  $\Phi_{\text{MIP}}$  of each MIP-cut location. In all cases,  $\Phi_{\text{MIP}}$  drops during the stimulus (\*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.005). No after-effects were observed (see Tables S2 and S3 for other parameter settings).

during the stimulus (the  $\Phi_{\text{MIP}}^{\text{[Fz,Cz,EDA,RES]}}$  also increased: Welch's *t* test: df = 553, t-stat = -4.0359,  $p < 10^{-6}$ ). This might relate to how the subject attempts to match past and current information.

As was observed in the MIP analysis,  $\sum_{M \in \mathcal{M}} \Phi^M_{MIP}$  during the stimulus was statistically more significant in the SYNC condition than the ASYNC condition (Welch's t test: df = 13177, t-stat = – 2.3734, p < 0.05). Although both conditions changed the perception of subject's body ownership, the  $\Phi$  value reflected the difference between the two subjective reports of the participants.

**IIT 2.0 estimates the degree of ownership.** In this final subsection, we discuss how IIT 2.0 can estimate subjective reports of body ownership. The original purpose of IIT was to measure the degree of consciousness theoretically. Although the estimation of consciousness itself is still beyond our computational power, this aim could be attained through our small body-brain system consisting of seven time-series of physiological data.

We have thus far observed the average effects of (a)synchronous stimulus on each brain-body system. However, there seem to be individual differences in the subjective strength of RHI: some subjects responded negatively to questionnaire statements Q1–Q3 asking RHI experience (see "Methods" and Table S1) in the SYNC condition,



**Figure 3.** (A) Time series of mean  $\sum_{M \in \mathcal{M}} \Phi_{\text{MIP}}^M$  in both conditions (blue: SYNC, red: ASYNC). The stimulus phase is 300 – 600 s. The other regions are in resting phase (no stimulus). The computational time frame for IIT is 250 (1 s). Notably, the moving-average (sample length 150 steps) was calculated in the time series to make temporal changes of values easy to see. (B)  $\sum_{M \in \mathcal{M}} \Phi_{\text{MIP}}^M$  of SYNC and ASYNC values rises significantly compared with each rest phase (\*p < 0.05, \* \* p < 0.01, \* \*p < 0.005). In contrast to the MIP analysis, an after-effect was observed in both conditions.

and conversely, some responded positively in the ASYNC condition (Fig. S3). To know why such individual differences emerge, we tested whether individual differences in the experience of the RHI were associated with an index of IIT analysis. First, we averaged the subjective ratings of questions Q1–Q3 for each subject in each condition. This value has been used to represent the overall strength of the illusion, as an RHI index<sup>28</sup>. The RHI index includes the referral of touch over a fake hand rather than purely body ownership. This is appropriate for our purpose because it may express a degree of ABO well, the concurrent feeling of both fake and own hand. We used this index for the following analysis.

We begin by asking whether the integrated information and  $\Phi_{\text{MIP}}$  for each subject can account for individual differences in subjective reports. The answer is no. In both conditions,  $\sum_{M \in \mathcal{M}} \Phi_{\text{MIP}}^M$  (abbr.  $\sum \Phi_{\text{MIP}}$ ) has a low correlation (SYNC: Pearson's correlation test, n = 21, r = 0.28, p = 0.21. ASYNC: n = 21, r = 0.20, p = 0.38). Interestingly,  $\sum \Phi_{\text{MIP}}$  itself did not reflect the subjective reports. In our system, at least, the  $\sum \Phi_{\text{MIP}}$  is not the effective estimator of subjective reports.

Next, we calculated the difference in  $\sum \Phi_{MIP}$  between the SYNC and the ASYNC conditions. After subtracting the baseline (mean pre-stimulus phase for each subject), we calculated the difference across stimulus phases and the sum of its time series. Mathematically, this measure can be expressed for each subject *i* as:

$$\mathfrak{s}_{i}(\Phi_{\mathrm{MIP}}) = \sum_{t=300}^{600} \{ \sum_{M \in \mathcal{M}_{t}} (\bar{\Phi}_{\mathrm{MIP}}^{M,\mathrm{SYNC}} - \bar{\Phi}_{\mathrm{MIP}}^{M,\mathrm{ASYNC}}) \}$$
(1)

where  $\bar{\Phi}_{\text{MIP}}^{M,\text{SYNC}}$  means that the integrity of the main complex *M* at time *t* in the SYNC condition subtracted the baseline.

We found that differences in RHI indices between conditions positively correlated with  $s_i(\Phi_{\text{MIP}})$  (Fig. 5; Pearson's correlation test: n = 21, r = 0.56, p < 0.005). This shows that the  $\Phi_{\text{MIP}}$  gaps, rather than of  $\Phi_{\text{MIP}}$  itself, can estimate individual differences in subjective rating gaps between experimental conditions.

### Discussion

This study showed that IIT 2.0 could characterise the subjective experience during the rubber hand illusion-that is, ambiguous body ownership (ABO)—in terms of both general tendency and individual difference.

Before summarising the results, let us review the computational meaning of the application of IIT.  $\Phi_{\text{MIPS}}$  differ from the single or coupled physiological data (e.g., heartbeats, skin conductance, and temperature). These interceptives are sometimes related to the FEP paradigm<sup>20</sup>. However, the relationship between the autonomic system and subjective experience is still unproven (at least at the level of single physiological data). If their relationship should be proven somehow, it would be indirectly.

 $\Phi_{MIP}$  differs from those measures: IIT measures not a change of variables but the change of the system's interaction structure.  $\Phi_{MIP}$  deals with the interaction heterogeneity of the system rather than its covariance; it measures how entangled information exists inside the system during the stimuli.

Some simple examples are helpful. In complete random interaction systems,  $\Phi_{MIP}$  becomes zero because its mutual information is also zero. In perfectly correlated interaction systems,  $\Phi_{MIP}$  also becomes zero because we cannot discriminate between local and global behaviour. The high  $\Phi_{MIP}$  found in the system have a more heterogeneous interaction structure (for a more detailed comparison, see Median et al.<sup>29</sup>).

 $\Phi_{\text{MIP}}$  is a measure positioned within the whole system structure. Therefore,  $\Phi_{\text{MIP}}$  can be a more direct measure of the subjective experience of RHI than dealing directly with physiological data. Here we regarded the autonomic and central nervous systems as one system; the results showed that they were complexly intertwined and could not be treated in isolation, even though they vary on different timescales.



**Figure 4.** (A) Histogram of the difference in the max  $_{M \in \mathcal{M}} \{\Phi_{\text{MIP}}^M\}$  frequency for pre- and intra-stimuli: Fr( $M \in \mathcal{M}$ [ST)—Fr( $M \in \mathcal{M}$ [PRE). Because many main complexes exist, we analysed only the outliers (indexed 1, 2, 3). Blue (red) bars show the frequency in the synchronous (asynchronous) condition. (B) Image of the main complex for the system *S*. {EDA, RES}, *S*, and *S*-{ECG} are the three greatest differences. (C) max  $_{M \in \mathcal{M}} \{\Phi_{\text{MIP}}^M\}$  for each main complex. Only for  $M = \{\text{EDA}, \text{RES}\}$  does  $\Phi_{\text{MIP}}$  rises significantly during the stimulus. The other two values are negative throughout the three phases (\*p < 0.05, \*\*p < 0.01, \* \*\*p < 0.005. See Tables S4 and S5 for other parameter settings).

We found several insights by applying IIT 2.0 to the small body–brain system. The MIP reveals that the system's independence emerges during the stimulus phase for both experimental conditions. The location of the MIP cut indicates that the disjunctive parts are not body–brain pairs but intra-brain or intra-body pairs. Before the stimulus, the body–brain system works as one irreducible system. Although the *S* exists even in the stimulus phase, the degree of this irreducibility ( $\Phi_{MIP}^S$ ) decreases. In other words, the system loses the entire entanglement during RHI.

We found several insights by applying IIT 2.0 to the small body—brain system. The MIP revealed that the system's independence emerged during the stimulus phase in both experimental conditions. The location of the MIP cut indicated that the disjunctive parts were not body–brain pairs but intra-brain or intra-body pairs. Before the stimulus, the body—brain system worked as one irreducible system. Although *S* existed even in the stimulus phase, the degree of this irreducibility ( $\Phi_{MIP}^S$ ) decreased. In other words, the system lost the total entanglement during RHI.

Although the whole system *S* lost its integrity, the subsystem (main complexes) retained high integrity (i.e.,  $\sum \Phi_{MIP}$ ) during the stimulus. The entangled high parts, for example, were {Fz, Cz, RES, and ECG}. As MIP is



**Figure 5.** Relationship between the difference in mean subjective reports,  $\langle R_{SYNC} \rangle - \langle R_{ASYNC} \rangle$ , and  $\mathfrak{s}(\Phi_{MIP})$  (see Eq. 1) for each subject (Pearson's correlation test: r = 0.56, n = 21, p < 0.005; see Table S6 for other parameter settings))

located at Oz, the complexes, including Fz and Cz, also have a compensated meaning. In the stimulus phase, the brain might raise the integrity for the working memory (Fz) and motor function (Cz) instead of separating the inconsistent visual information (Oz). This relation is logical if we consider that the motor-perception with the subject's short-term memory (including multisensory integration at the ventral premotor cortex<sup>30,31</sup>) attempts to counterbalance the mismatch in the visual system. Body integrity (RES and EDA) was also observed in the stimulus phase. Notably, ECG (i.e., heartbeat) was excluded from this body integrity (the MIP cut is located at {ECG} in Fig. 2B). This relative independence of the heartbeat might be related to its use as an index of the malleability of body representation<sup>11</sup>.

Although our average IIT analysis explained the general tendencies, individual differences still existed. Some subjects experienced a much stronger feeling of owning the fake hand in the SYNC condition than in the ASYNC condition; but others did not show such a strong difference between conditions. Previous studies have tried to explain these various individual differences from the physiological perspective<sup>21</sup>, but with no consistent conclusions. Here, we found the positive correlation of the  $\sum \Phi_{MIP}$  value with subjective reports as focusing on the differences between conditions (SYNC minus ASYNC). This suggests that subjective evaluation is determined not by a single experience but rather by comparing different experiences. We also attempted the same analysis by applying the peak–end rule, which states that how a subject feels at their peak and at their end determines their judgement of a given experience<sup>32</sup>. The results were the same (Table S6). Furthermore, it is necessary to use time-varying data on the subjective rating to analyse the relationship between the detailed subjective rating (the onset of the illusion) and IIT. This is left for future work.

Finally, we comment on two critical issues on ABO from our analysis:

- 1. Average tendency of  $\Phi_{MIP}$  variations: We confirmed the graph similarities between the two experimental conditions (note that a statistical difference existed between them). This result might appear somewhat strange, because both conditions generate different phenomenological experiences. However, we should also be aware that the ASYNC condition is not a neutral experience. For instance, some researchers have pointed out that we should not miss that the ASYNC condition generates an "ambiguous sensation" during RHI<sup>33,34</sup>. Furthermore, Perez-Marcos et al. have reported that body image distortion can be elicited by the ASYNC condition but not by the SYNC condition<sup>35</sup>. Their results imply that the ASYNC condition also changes latent body cognition levels. All these results suggest that the ASYNC condition also contribute to change the phenomenological experiences under the radar of the questionnaires.
- 2. Individual subjective differences and  $\Phi_{MIP}$  s: We confirmed that  $\Phi_{MIP}$  itself was not associated with subjective reports in either condition. Individual differences in RHI experience have been pointed out from several aspects. There are many biases in the subjective reports. First, the anchoring effect is the most famous: past subject ratings influence current ratings<sup>36</sup>. Second, an order effect of the experimental conditions exists<sup>37</sup>. Third, hypothesis awareness also affects subjective ratings<sup>36,39</sup>: the experimental setting changes the subject's expectations. All these biases show that the experimental setting influences the subjective ratings without the subject's intention. Another possible reason is perceptual ability<sup>40,41</sup>. For instance, Costantini et al. have shown that the participant's temporal resolution ability changes the illusion strength in the RHI in the asynchronous condition<sup>42</sup>. Also, proprioceptive accuracy has been shown to associate with RHI strength<sup>43</sup>. Therefore, individual differences in RHI index could be due to a variety of factors; we may need more data to associate with the individual subjective difference from  $\Phi_{MIP}$  itself.

While there was no association between the RHI index and  $\Phi_{\text{MIP}}$  in either condition, we found that the difference in  $\Phi_{\text{MIP}}$  between conditions (i.e.,  $\mathfrak{s}(\Phi_{\text{MIP}})$ ) correlated with the difference in RHI indices between conditions, suggesting that  $\Phi_{\text{MIP}}$  reflects individual differences in RHI experience when comparing conditions. So what component of the RHI index is crucial for association with  $\mathfrak{s}(\Phi_{\text{MIP}})$ ?

The difference in Q3 between conditions (dQx, where x represents an item number) was correlated significantly with  $\mathfrak{s}(\Phi_{\text{MIP}})$ , and its coefficient was slightly smaller than the RHI index (Pearson's correlation test: n = 21, r = 0.53, p < 0.05). The dQ2 was also correlated significantly, and its coefficient was even smaller than that of dQ3 (n = 21, r = 0.44, p < 0.05). There was no significant correlation between dQ1 and  $\mathfrak{s}(\Phi_{\text{MIP}})$  (n = 21, r = 0.40, p = 0.07). The RHI index can be classified into "ownership" (Q3) and "referral of touch" (Q1 and Q2)<sup>28</sup>. Therefore, these results of correlation analysis suggest that  $\mathfrak{s}(\Phi_{\text{MIP}})$  reflects "ownership" rather than "referral of touch". In this sense, perhaps "ownership" distorts the "referral touch"<sup>44</sup>.

However, it might be too early to assume that  $\mathfrak{s}(\Phi_{\mathrm{MIP}})$  represents pure ownership. We found significant differences between conditions even in the control questions Q7 and Q8 (paired t-test, p < 0.05). Interestingly, dQ7 and dQ8 ratings were significantly correlated with dQ3 (i.e., ownership) (Pearson's correlation test, dQ7: n = 21, r = 0.67, p < 0.005; dQ8: n = 21, r = 0.45, p < 0.05). Calculating mean values of dQ1, dQ2, dQ3, dQ7, and dQ8 for each participant, we found that these values were correlated significantly with  $\mathfrak{s}(\Phi_{\mathrm{MIP}})$  and, importantly, had larger coefficients than the RHI index (Pearson's correlation test: n = 21, r = 0.59, p < 0.005).

How should we consider this stronger correlation caused by adding control questions to the RHI index? We should note that the participants gave negative answers (50) to Q7 and Q8 in almost all conditions (Q7: mean 57.3, SD 31.8 for SYNC; mean 43.8, SD 29.2 for ASYNC; Q8: mean 37.3, SD 27.7 for SYNC; mean 28.7, SD 26.0 for ASYNC). Remarkably, these ratings indicated that denial belief was weakened in the SYNC condition. The significant correlation of dQ3 with dQ7 and dQ8 indicated that participants who felt ownership of the dummy (maybe unconsciously) tended to give a neutral answer rather than a negative one to these questions. Q7 and Q8 were related to the physical properties and location, respectively, of the dummy hand. Participants might have had an ambiguous feeling of "It's both my hand and a dummy hand." Thus, we assumed that  $\mathfrak{s}(\Phi_{\rm MIP})$  measured the ambiguous sensation accompanying ownership, ABO, rather than ownership itself.

As mentioned in the Introduction,  $\Phi_{\text{MIP}}$  in IIT is used to measure the gap between feeling and judgement. Given these considerations, it may be correct to consider that what  $\Phi_{\text{MIP}}$  measures is not a "concrete" phenomenological state but rather a vague sensation that includes positive ratings (such as referral touch and ownership) and negative ratings observed in control questions, that is, ABO. In other words, ABO reflects entangled information for the single brain–body system.

### Methods

**Basic concept of IIT.** Let us first explain IIT concepts for the sake of readers unfamiliar with IIT. Although many concepts exist in IIT, we focus on three: minimum information partition (MIP), main complexes, and integrated information  $\Phi$ .

IIT deals with the intrinsic, rather than extrinsic, information; it only depends on inner variables. Typical information theories used in biology focus on the relation between external inputs and their results. In this setting, the system itself becomes a black box. The aim of IIT is to consider this black box in terms of "what the system is" rather than "what the system does"<sup>45</sup>. The basic concept of IIT expresses mutual information between the past and current system states; mathematically, this is expressed as  $I(X(t - \tau); X(t))$ .

However, this simple mutual information contains redundant information on the system's integrity, which should be measured as the system's irreducibility to its parts (subsystems). The information obtained from the subsystems must be subtracted from the entire information  $I(X(t - \tau); X(t))$ . The rest of the information represents the system's integrity; therefore, the problem for IIT lies in the determination of the system partitions.

This study applies IIT as the mismatch decoding proposed by Oizumi et al.<sup>18</sup> because  $\Phi^*$  satisfies the following properties. (i) Non-negativity: the lower bound of  $\Phi^*$  is zero; (ii) sensitivity to noise correlation:  $\Phi^*$  deals with external correlated noise<sup>29</sup>; and (iii) applicability to continuous variables. Sometimes, the IIT as the mismatch decoding is classified as "IIT 2.0" to distinguish it from IIT 3.0<sup>46–48</sup>. In this paper, we did not apply IIT 3.0 to the physiological data because IIT 3.0 deals only with discrete values.

Integrated information ( $\Phi^*$ ) and minimum information partition (*MIP*). As mentioned in the previous section, the mode of partitioning is essential to determining the system's integrity. MIP is the system partition where the system's integrity is at its minimum. The system's integrity in  $\Phi^*$  is given by

$$\Phi^* = I(X(t-\tau); X(t)) - I^*[X; \tau, \mathscr{P}_S],$$
(2)

where  $I(X(t - \tau); X(t))$  is the mutual information between the current X(t) and past states  $X(t - \tau); S$  is the set of all nodes of a given system;  $\mathcal{P}_S$  is the set of all bi-partitions (total  $2^{|S|} - 1$  partitions);  $I^*(X; \tau, \mathcal{P}_S)$  is a "hypothetical" mutual information, indicating the mismatched decoding in the partitioned probability distribution. More precisely,  $I^*(X; \tau, \mathcal{P}_S)$  is given as the partition  $\max_{\beta} \tilde{I}(\beta; X, \tau, \mathcal{P}_S)$  that minimises  $\Phi^*$  (see listed studies<sup>18,29</sup> for further details about this expression).

Because  $\Phi^*$  depends on the partition  $\pi \in \mathscr{P}_{S}$ , MIP is the partition that minimises the integrated information.

$$\pi_{\text{MIP}} := \arg\min_{\pi \in \mathscr{R}} \Phi^*(\pi) \tag{3}$$

The integrated information of  $\pi_{\text{MIP}}$  is expressed as  $\Phi^*(\pi_{\text{MIP}})$ . We simply denote  $\Phi^*(\pi_{\text{MIP}})$  as  $\Phi_{\text{MIP}}$ . Notably, if  $\Phi_{\text{MIP}}$  equals zero, the parts of the system are mutually independent; that is, there is no interaction between the parts. In this sense,  $\Phi_{\text{MIP}}$  characterises the system's irreducibility to its parts.

*Main complexes.* Generally, we can compute  $\Phi_{\text{MIP}}$  for any subsystem in the system, not only for the set *S*. We denote each  $\Phi_{\text{MIP}}$  for a subsystem *T* as  $\Phi_{\text{MIP}}^T$ , where  $T \subset S$ . A "complex" is a subsystem  $C(\subset S)$ , where  $\Phi_{\text{MIP}}^C > \Phi_{\text{MIP}}^T$  for all supersets of *S*. From this definition, we can define the main complexes as those with the local maximum  $\Phi_{\text{MIP}}^M$ .

Definition (main complex): A main complex is a complex *M* satisfying  $\Phi_{\text{MIP}}^M > \Phi_{\text{MIP}}^R$  for any subsystem  $R \subset M$ . The definition says that if two main complexes exist (say, *A* and *B*), then they are exclusive. IIT researchers consider these main complexes to be the system's information core that might be related to our conscious experience. The validity of this assumption remains to be demonstrated, although there is no doubt that such information cores play a vital role in living systems<sup>49-51</sup>.

**Application to RHI.** *Participants.* Twenty-two healthy male adults (mean age, 21.1 years; SD, 0.5) were recruited from among the students at Nagaoka University of Technology. All participants gave written consent to participate after receiving an explanation of the procedures involved. The study was conducted according to the principles of the Declaration of Helsinki and was approved by the Ethics Committee of Nagaoka University of Technology.

*Procedure.* Participants sat on a chair with their left arm on a table with the palm facing down. The experimenter stood opposite the participant across the table. A realistic-looking left rubber hand was placed to the right of the participant's real left hand. The distance between the index finger of the rubber hand and that of the real hand was 15 cm. Two grey towels covered the space from the participant's shoulder to the participant's wrist and the proximal end of the artificial hand. A 45 cm  $\times$  60 cm standing screen between the rubber hand and the real hand prevented participants from viewing their own left hand (see Fig. S1).

Two experimental conditions were investigated in our study: synchronous and asynchronous stroking. The experiments comprised two blocks corresponding to these conditions. The sequence of the blocks was randomised and counterbalanced across participants. Each block lasted for 15 min, with a few minutes of inter-block intervals: 5-min rest periods (Rest Phase: Pre and Post) before and after a 5-min stroking period (Stimulus Phase). The rubber hand and unseen hand were stroked by two brushes synchronously or asynchronously during the stroking period. A trained experimenter indicated the predetermined pattern and frequency (40 bpm) of stimulation using a metronome through an earphone. In the asynchronous condition, the dummy hand was touched immediately after touching the hidden real hand by the metronome; that is, every visual stimulus followed a tactile stimulus in the asynchronous condition. Thus, the tactile stimulation was identical for both conditions. After the second rest period, the participants completed a questionnaire about their subjective experiences during the stroking period.

*Measurement.* Physiological measurement. Throughout each experimental block, EEG, electrocardiogram (ECG), respiration airflow (RES), and electrodermal activity (EDA) were recorded using a bio-amplifier (Polymate Pro MP 6000, Miyuki Giken Corp., Tokyo, Japan) at a sampling rate of 250 Hz and with a 24-bit resolution, using a 50-Hz notch filter. In particular, for the EEG, four signals (Fz, Cz, Pz, Oz) were measured at the locations of the four electrodes on the scalp according to the international 10-20 system for measuring the four EEG signals (Fz, Cz, Pz, Oz); here, we used a high-frequency 50 Hz filter and a low-frequency 1 Hz filter.

Questionnaire. The questionnaire comprised nine statements from the original RHI study<sup>1</sup> presented on a computer screen in random order (see Table S1). Participants indicated their responses by clicking on a visual-analogue scale ranging from strong agreement (100) to strong disagreement (0), including neither agreement nor disagreement (50). All the questions and instructions were presented in Japanese. Questionnaire data from one participant were excluded from the analysis because they were lost owing to mechanical issues.

Data setting for IIT 2.0 application. Oizumi et al. proposed approximation methods for the computational problem of IIT<sup>18,19,52,53</sup>. This study applied their "Practical  $\Phi$  Toolbox for MATLAB" to the physiological data; we considered only seven types of physiological data for the 22 subjects. The exhaustive method (computing all possible MIPs) can be applied in realistic computation time for such a small system.

Another parameter is time delay  $\tau$  (IIT has only two constraints: partition  $\pi$  and time delay  $\tau$ ). To choose the suitable parameter, we computed  $\Phi_{\text{MIP}}$  for  $\tau$  from 0 to 250 frames. The time delay  $\tau$  is the point where the mean is the maximum  $\Phi_{\text{MIP}}$  for all subjects. We found that this  $\tau$  is approximately 50 frames (i.e. 0.2 s). This value seems to be suitable because the value is the same as the human response time (Fig. 2S).

The time series of each set of raw data was 225,000 frames (250 frames per second). Because we applied a 50-frame delay (0.2 s) for the computation of IIT 2.0, sufficient data series for IIT 2.0 are required. We attempted 250 frames (1 s), 500 frames (2 s), and 750 frames (4 s) for IIT 2.0 computations—the time window shifts 250 frames each. The first set of blank data for some frames (250, 500, 750 frames) was set at zero (only the first three steps at most for 750 frames), which sum up to 900 datasets for each subject. The results for all frame samples were almost the same (Tables S2–S6).

We extracted all measures related to integrated information for 900 steps (1 s for one step). The first third (0–300 steps) and the last third (600–900 steps) correspond to the pre- and post-stimulus resting states. The middle section (300–600 steps) is where the synchronous or asynchronous stimulus was applied. For all statistical tests, we detrended the obtained  $\Phi$  time series because our focus was on the variation due to the stimulus for our rubber hand experiments. The negative value of  $\Phi_{\text{MIP}}$  is due to this treatment.

# Data availability

All data are available from the corresponding author upon reasonable request.

Received: 7 February 2022; Accepted: 20 October 2022 Published online: 26 October 2022

#### References

- 1. Botvinick, M. & Cohen, J. Rubber hands 'feel' touch that eyes see. *Nature* **391**, 756. https://doi.org/10.1038/35784 (1998).
  - Lenggenhager, B., Tadi, T., Metzinger, T. & Blanke, O. Video ergo sum: Manipulating bodily self-consciousness. Science 317, 1096–1099 (2007).
  - 3. Blanke, O. & Arzy, S. The out-of-body experience: Disturbed self-processing at the temporo-parietal junction. *Neuroscientist* 11, 16–24 (2005).
  - Blanke, O., Landis, T., Spinelli, L. & Seeck, M. Out-of-body experience and autoscopy of neurological origin. *Brain.* 127, 243–258. https://doi.org/10.1093/brain/awh040 (2004).
  - Gallagher, S. & Zahavi, D. The phenomenological mind: An introduction to philosophy of mind and cognitive science (year2007).
     Synofzik, M., Vosgerau, G. & Newen, A. Beyond the comparator model: A multifactorial two-step account of agency. *Conscious*.
  - Cogn. https://doi.org/10.1016/j.concog.2007.03.010 (2008).
    7. De Vignemont, F. Embodiment, ownership and disownership. Conscious. Cogn. https://doi.org/10.1016/j.concog.2010.09.004 (2011).
  - David, A. B. & Ataria, Y. The body image-body schema/ownership-agency model for pathologies: four case studies. In Body Schema and Body Image https://doi.org/10.1093/oso/9780198851721.003.0020 (2021).
  - 9. Ataria, Y. Body Disownership in Complex Posttraumatic Stress Disorder (2018).
  - Friston, K. J., Stephan, K. E., Montague, R. & Dolan, R. J. Computational psychiatry: The brain as a phantastic organ. Lancet Psychiatry. https://doi.org/10.1016/S2215-0366(14)70275-5 (2014).
  - Tsakiris, M., Tajadura-Jiménez, A. & Costantini, M. Just a heartbeat away from one's body: Interoceptive sensitivity predicts malleability of body-representations. Proc. R. Soc. B Biol. Sci. https://doi.org/10.1098/rspb.2010.2547 (2011).
  - Apps, M. A. & Tsakiris, M. The free-energy self: A predictive coding account of self-recognition. Neurosci. Biobehav. Rev. https:// doi.org/10.1016/j.neubiorev.2013.01.029 (2014).
  - 13. de Vignemont, F. & Alsmith, A. The Subject's Matter. Self Consciousness and the Body (2017).
  - Tsakiris, M. The material me: unifying the exteroceptive and interoceptive sides of the bodily self. In *The Subject's Matter* https:// doi.org/10.7551/mitpress/10462.003.0019 (2018).
  - Ransom, M. et al. Affect-biased attention and predictive processing. Cognition https://doi.org/10.1016/j.cognition.2020.104370 (2020).
  - Balduzzi, D. & Tononi, G. Qualia: The geometry of integrated information. PLoS Comput. Biol. https://doi.org/10.1371/journal. pcbi.1000462 (2009).
  - Tononi, G., Boly, M., Massimini, M. & Koch, C. Integrated information theory: From consciousness to its physical substrate. *Nat. Rev. Neurosci.* https://doi.org/10.1038/nrn.2016.44 (2016).
  - Oizumi, M., Amari, S.-I., Yanagawa, T., Fujii, N. & Tsuchiya, N. Measuring integrated information from the decoding perspective. PLOS Comput. Biol. 12, 1–18. https://doi.org/10.1371/journal.pcbi.1004654 (2016).
  - Oizumi, M., Tsuchiya, N. & Amari, S. I. Unified framework for information integration based on information geometry. Proc. Natl. Acad. Sci. USA https://doi.org/10.1073/pnas.1603583113 (2016).
  - Critchley, H. D., Botan, V. & Ward, J. Absence of reliable physiological signature of illusory body ownership revealed by fine-grained autonomic measurement during the rubber hand illusion. *PLoS ONE* https://doi.org/10.1371/journal.pone.0237282 (2021).
  - Armel, K. C. & Ramachandran, V. S. Projecting sensations to external objects: Evidence from skin conductance response. Proc. R. Soc. B Biol. Sci. https://doi.org/10.1098/rspb.2003.2364 (2003).
  - Honma, M., Koyama, S. & Osada, Y. Double tactile sensations evoked by a single visual stimulus on a rubber hand. *Neurosci. Res.* 65, 307–311 (2009).
  - Petkova, V. I. & Ehrsson, H. H. If i were you: Perceptual illusion of body swapping. PLoS ONE 3, 1–9. https://doi.org/10.1371/ journal.pone.0003832 (2008).
  - Slater, M., Spanlang, B., Sanchez-Vives, M. V. & Blanke, O. First person experience of body transfer in virtual reality. *PLoS ONE* 5, 1–9. https://doi.org/10.1371/journal.pone.0010564 (2010).
  - Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J. & Passingham, R. E. Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. Proc. Natl. Acad. Sci. 104, 9828–9833. https://doi.org/10.1073/pnas.0610011104 (2007).
  - González-Franco, M., Peck, T. C., Rodríguez-Fornells, A. & Slater, M. A threat to a virtual hand elicits motor cortex activation. *Exp. Brain Res.* https://doi.org/10.1007/s00221-013-3800-1 (2014).
  - Kitazono, J., Kanai, R. & Oizumi, M. Efficient search for informational cores in complex systems: Application to brain networks. *Neural Netw.* https://doi.org/10.1016/j.neunet.2020.08.020 (2020).
  - Reader, A. T., Trifonova, V. S. & Ehrsson, H. H. The relationship between referral of touch and the feeling of ownership in the rubber hand illusion. *Front. Psychol.* https://doi.org/10.3389/fpsyg.2021.629590 (2021).
  - Mediano, P. A., Seth, A. K. & Barrett, A. B. Measuring integrated information: Comparison of candidate measures in theory and simulation. *Entropy* https://doi.org/10.3390/e21010017 (2019).
  - Ehrsson, H. H., Spence, C. & Passingham, R. E. That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. Science https://doi.org/10.1126/science.1097011 (2004).
  - Gentile, G., Björnsdotter, M., Petkova, V. I., Abdulkarim, Z. & Ehrsson, H. H. Patterns of neural activity in the human ventral premotor cortex reflect a whole-body multisensory percept. *NeuroImage* https://doi.org/10.1016/j.neuroimage.2015.01.008 (2015).
  - Kahneman, D., Fredrickson, B. L., Schreiber, C. A. & Redelmeier, D. A. When more pain is preferred to less: Adding a better end. Psychol. Sci. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x (1993).
  - Lewis, E. & Lloyd, D. M. Embodied experience: A first-person investigation of the rubber hand illusion. *Phenomenol. Cogn. Sci.* https://doi.org/10.1007/s11097-010-9154-2 (2010).
  - Valenzuela-Moguillansky, C., Bouhassira, D. & O'Regan, J. K. The role of body awareness in pain: An investigation using the rubber hand illusion. J. Conscious. Stud. 18(9), 110–142 (2011).
  - Perez-Marcos, D. et al. Selective distortion of body image by asynchronous visuotactile stimulation. Body Image https://doi.org/ 10.1016/j.bodyim.2017.11.002 (2018).
  - Sherif, M., Taub, D. & Hovland, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. J. Exp. Psychol. https:// doi.org/10.1037/h0048784 (1958).
  - 37. Lush, P. Order effects in the rubber hand illusion. PsyArXiv https://doi.org/10.31234/osf.io/amsrp (2021).
  - Lush, P. et al. Trait phenomenological control predicts experience of mirror synaesthesia and the rubber hand illusion. Nat. Commun. https://doi.org/10.1038/s41467-020-18591-6 (2020).

- Lush, P., Seth, A. K. & Dienes, Z. Hypothesis awareness confounds asynchronous control conditions in indirect measures of the rubber hand illusion. R. Soc. Open Sci. https://doi.org/10.1098/rsos.210911 (2021).
- Haans, A., Kaiser, F. G., Bouwhuis, D. G. & Ijsselsteijn, W. A. Individual differences in the rubber-hand illusion: Predicting selfreports of people's personal experiences. *Acta Psychol.* https://doi.org/10.1016/j.actpsy.2012.07.016 (2012).
- Marotta, A., Tinazzi, M., Cavedini, C., Zampini, M. & Fiorio, M. Individual differences in the rubber hand illusion are related to sensory suggestibility. *PLoS ONE* https://doi.org/10.1371/journal.pone.0168489 (2016).
- Costantini, M. et al. Temporal limits on rubber hand illusion reflect individuals' temporal resolution in multisensory perception. Cognition https://doi.org/10.1016/j.cognition.2016.08.010 (2016).
- Horváth, Á. *et al.* Proprioception but not cardiac interoception is related to the rubber hand illusion. *Cortex* https://doi.org/10. 1016/j.cortex.2020.08.026 (2020).
- Kalckert, A., Perera, A. T. M., Ganesan, Y. & Tan, E. Rubber hands in space: The role of distance and relative position in the rubber hand illusion. *Exp. Brain Res.* https://doi.org/10.1007/s00221-019-05539-6 (2019).
- Albantakis, L. & Tononi, G. The intrinsic cause-effect power of discrete dynamical systems-from elementary cellular automata to adapting animats. *Entropy* https://doi.org/10.3390/e17085472 (2015).
- Oizumi, M., Albantakis, L. & Tononi, G. From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. PLoS Comput. Biol. https://doi.org/10.1371/journal.pcbi.1003588 (2014).
- 47. Tononi, G. & Koch, C. Consciousness: Here, There and Everywhere? https://doi.org/10.1098/rstb.2014.0167 (2015).
- Albantakis, L., Marshall, W., Hoel, E. & Tononi, G. What caused what? A quantitative account of actual causation using dynamical causal networks. *Entropy* https://doi.org/10.3390/e21050459 (2019).
- Marshall, W., Kim, H., Walker, S. I., Tononi, G. & Albantakis, L. How causal analysis can reveal autonomy in models of biological systems. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* https://doi.org/10.1098/rsta.2016.0358 (2017).
- Sheneman, L., Schossau, J. & Hintze, A. The evolution of neuroplasticity and the effect on integrated information. *Entropy* https:// doi.org/10.3390/e21050524 (2019).
- Niizato, T. et al. Finding continuity and discontinuity in fish schools via integrated information theory. PLoS ONE https://doi.org/ 10.1371/journal.pone.0229573 (2020).
- Kitazono, J., Kanai, R. & Oizumi, M. Efficient algorithms for searching the minimum information partition in Integrated Information Theory. *Entropy* https://doi.org/10.3390/e20030173 (2018).
- Hidaka, S. & Oizumi, M. Fast and exact search for the partition with minimal information loss. *PLoS ONE* https://doi.org/10.1371/ journal.pone.0201126 (2018).

# Author contributions

T.N. and Y.N. designed the experiments. Y.N., T.K., T.O., and T.Y. conducted the experiments. T.N., Y.N., and K.S. analysed and interpreted the data. T.N. and Y.N. wrote the manuscript. All authors reviewed the manuscript.

# Competing interests

The authors declare no competing interests.

# Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-22927-1.

Correspondence and requests for materials should be addressed to T.N. or Y.N.

Reprints and permissions information is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022