Neuroscience

Body size as a metric for the affordable world

Xinran Feng, Shan Xu, Yuannan Li, Jia Liu

Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing, China • Faculty of Psychology, Beijing Normal University, Beijing, China

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed preprint posted
October 24, 2023 (this version)

Sent for peer review
August 14, 2023

Posted to bioRxiv
May 6, 2023

Abstract

The physical body of an organism serves as a vital interface for interactions with its environment. Here we investigated the impact of human body size on the perception of action possibilities (affordances) offered by the environment. We found that the body size delineated a distinct boundary on affordances, dividing objects of continuous real-world sizes into two discrete categories with each affording distinct action sets. Additionally, the boundary shifted with imagined body sizes, suggesting a causal link between body size and affordance perception. Intriguingly, ChatGPT, a large language model lacking physical embodiment, exhibited a modest yet comparable affordance boundary at the scale of human body size, suggesting the boundary is not exclusively derived from organism-environment interactions. A subsequent fMRI experiment showed that only the affordances of objects within the range of body size were represented in the brain, suggesting that objects capable of being manipulated are the only objects capable of offering affordance in the eyes of an organism. In summary, our study suggests a novel definition of object-ness in an affordance-based context, advocating the concept of embodied cognition in understanding the emergence of intelligence constrained by an organism's physical attributes.

eLife assessment

This paper presents **valuable** findings that shed light on the mental organisation of knowledge about real-world objects. It provides diverse if **incomplete** evidence from behaviour, brain, and large language models that this knowledge is divided categorically between relatively small objects that are at the relevant scale for direct manipulation and larger objects that are outside the typical scope of human affordances for action.

Introduction

Man is the measure of all things. - Protagoras

The assertion by the ancient Greek philosopher Protagoras highlights the notion that reality is defined by how the world is perceived by humans. A contemporary interpretation of this statement is the embodied theory of cognition (e.g., Chemero, 2013; Gallagher, 2017; Gibbs, 2005; Wilson, 2002; Varela et al., 2017), which posits that human body scale (e.g., size)

constrains the perception of objects and the generation of motor responses. For instance, humans evaluate the climbability of steps based on their leg length (Mark, 1987; Warren,1984), and determine the navigability of apertures according to the critical aperture-to-shoulder-width ratio (Warren & Whang, 1987). Additionally, grasping strategies have been shown to be contingent upon object size relative to one's body (Cesari & Newell, 2000; Newell et al., 1989) or hand size (Castiello et al., 1993; Tucker & Ellis, 2004). However, whether body size simply serves as a reference for locomotion and object manipulation, or alternatively, plays a pivotal role in shaping the representation of objects as suggested by Protagoras, remains an open question.

To underscore the latter point, Gibson (1979), the pioneer of embodied cognition research, stated that "Detached objects must be comparable in size to the animal under consideration if they are to afford behavior (p.124)." This implies that an object's affordance, encompassing all action possibilities offered to an animal, is determined by the object's size relative to the animal's size rather than its real-world size. For instance, in a naturalistic environment, such as a picnic scene shown in **Fig. 1a**, there may exist a qualitative distinction between objects within (the objects with warm tints in **Fig. 1a**) and beyond (those with cold tints) the size range of humans. Only objects within the range, such as the apple, the umbrella and the bottles, may afford actions, while those beyond this range, such as the trees and the tent, are largely viewed as part of the environment. Consequently, visual perception may be ecologically constrained, and the body may serve as a metric that facilitates meaningful engagement with the environment by differentiating objects that are accessible for interactions from those not. Based on Gibson's statement, we proposed two hypotheses: first, the affordance of objects will exhibit a qualitative difference between objects within and beyond the size range of an organism's body; second, affordance-related neural activity will emerge exclusively for objects within the organism's size range.

To test these hypotheses, we first measured the affordance of a diverse array of objects varying in real-world sizes (e.g., Konkle, & Oliva, 2011). We found a dramatic decline in affordance similarity between objects within and beyond human body size range, as these objects afforded distinct sets of action possibilities. Notably, the affordance boundary shifted in response to the imagined body sizes and could be attained solely from language, as demonstrated by the large language model (LLM), ChatGPT (OpenAI, 2022). A subsequent fMRI experiment corroborated the qualitative difference in affordances demarcated by the body size, as affordances of objects within humans' size range, but not those beyond, were represented in both dorsal and ventral visual streams of the brain. This study advances our understanding of the role of body size in shaping object representation and underscores the significance of body size as a metric for determining object affordances that facilitates meaningful engagement with the environment.

Results

To illustrate how human body size affects object affordances with different sizes, we first characterized the affordances of a set of daily objects. In each trial, we presented a matrix consisting of nine objects and asked participants to report which objects afforded a specific action (e.g., sit-able: a chair, a bed, a skateboard, but not a phone, a laptop, an umbrella, a kettle, a plate, or a hammer) (**Fig. 1b**). In this task, there were 14 actions commonly executed in daily life and 24 object images from the THINGS database (Hebart et al., 2019), with sizes ranging from size rank 2 to 8 according to Konkle and Oliva (2011)'s classification. These objects covered real-world sizes from much smaller (17 cm on average, rank 2) to orders of magnitude larger (5,317 cm on average, rank 8) than the human body size (see Methods for details). Consequently, affordances for each object were indexed by a 14-dimensional action vector, with the value for each dimension representing the percentage of participants who agreed on a certain action being afforded by the object (e.g., 88% for the action of grasping on a hammer indicating 88% of participants agreed that a hammer affords grasping). Supplemental Fig. S1 showed the affordances of two example objects.

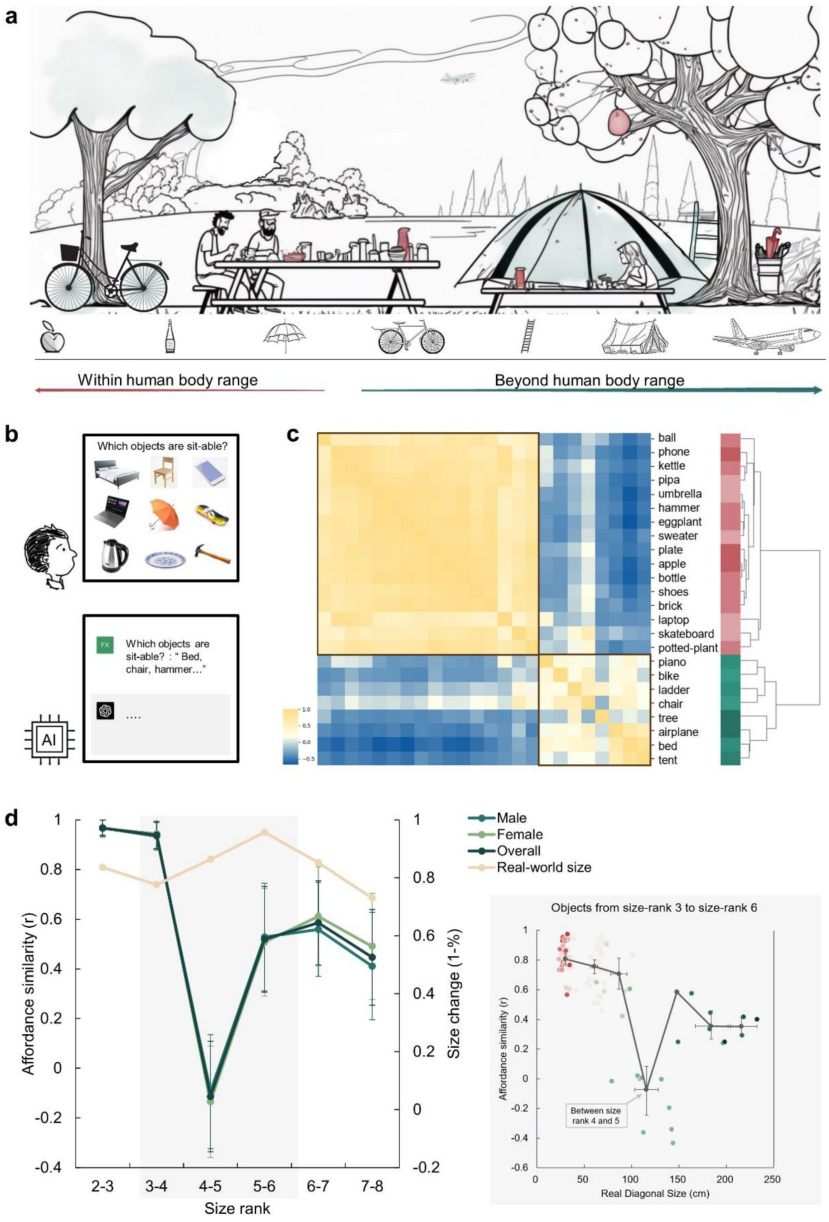

Fig 1.

An affordance boundary in the affordable world.

a, An illustration of a picnic scene, featuring objects of various sizes relative to human body. Example objects within the normal body size range are painted red, and those beyond green. We hypothesized qualitative differences between perceived affordances of these two kinds of objects. **b**, A demonstration of the object-action relation judgement task for human participants (top) and AI models (bottom). The question in task for human participants was presented in Chinese. **c**, The representational similarity matrix (RSM) for objects based on human rating of affordance similarity. Object sizes are denoted with red to green. Two primary clusters emerged in clustering analysis of the similarity pattern are outlined with black boxes. **d**, *Left panel*: The overall affordance similarity and that of each gender (left y-axis) as well as real-world size similarity (right y-axis) between neighboring size ranks. The error bars represent the standard error (SE). *Right panel*: The point clouds of pairwise correlations between objects from the same rank or neighboring ranks. Each colored dot represents the affordance similarity (y-axis) and the average real-world size (x-axis) of a specific object pair. The grey dots indicate the averaged size (x-axis) and pairwise similarity (y-axis) of object pairs in different rank compositions. Left to right: both from size rank 3, from size rank 3 and 4, both from size rank 4, from size rank 4 and 5, both from size rank 5, from size rank 5 and 6, and both from size rank 6. The horizontal error bars represent 95% confidence interval (CI) of the averaged object size in each pair, and the vertical error bars denote the CI of pairwise affordance similarity.

An affordance similarity matrix was then constructed where each cell corresponded to the similarity in affordances between a pair of objects (**Fig. 1c**). A clustering analysis revealed a two-cluster structure. Visual inspection suggested that the upper-left cluster consisted of objects smaller than human body size (red labels), and the lower-right cluster contained objects larger than human body size (green labels). Critically, the between-cluster similarity in the affordance similarity matrix approached zero, suggesting a division in affordances located near the body size. To quantify this observation, we calculated the similarity in affordances between each neighboring size rank. Indeed, we identified a clear trough in affordance similarity, dropping to around zero, between size rank 4 (77cm on average) and 5 (146cm on average), which was significantly smaller than that between size rank 3 and 4 ($Z = 3.91$, $p < .001$) and that between size rank 5 and 6 ($Z = 1.66$, $p = .048$). This trough suggested an affordance boundary between size rank 4 and 5, while affordance similarities between neighboring ranks remained high (rs > 0.45) and did not significantly differ from each other (ps > 0.05) on either side of the boundary (**Fig. 1d**, left panel, green lines). This pattern was evident for both genders, indicating no gender difference. Note that the abrupt change in affordance similarity across the boundary cannot be explained by changes in objects' real-world size, as the similarity in objects' real-world size was relatively stable across ranks, without any trend of a trough-shape curve (**Fig. 1d**, left panel, yellow line). Intriguingly, rank 4 and rank 5 corresponds to 80 cm to 150 cm, a boundary situated between these two ranks is within the range of the body size of a typical human adult. This finding suggested that objects were classified into two categories based on their affordances, with the boundary aligning with human body size.

To better locate the boundary, we focused on the affordance similarity between individual objects within size rank 3 to 6 (approximately ranging from 30cm to 220cm in real-world size, the area with grey shade in **Fig. 1d**), where the trough-shape curve was identified. Specifically, we traversed all pairs of objects with similar real-world diagonal sizes (from either the same rank or from neighboring ranks), calculated their average real-world size as an index of the approximate location of boundary between this pair of objects, and plotted the affordance similarity against the average real-world size of each object pair. As shown in the inset (grey box) of **Fig. 1d**, consistent with the rank-wise analysis, the abrupt decrease in affordance similarity exclusively happened between objects from size rank 4 and 5 (light green dots). The averaged real-world size in these object pairs was 104 cm (95% CI, 105 to 130 cm) and the affordance similarity in such object pairs was around zero. This result further narrowed the location estimation of the boundary, and demonstrated that the affordance boundary persisted at the level of individual objects.

One may argue that the location of the affordance boundary coincidentally fell within the range of human body size, rather than being influenced by human body size. To establish a causal link between them, we directly manipulated the body schema, referring to an experiential and dynamic functioning of the living body in its environment (Merleau-Ponty & Smith, 1962), to examine whether the affordance boundary would shift accordingly. Utilizing the same paradigm, we instructed a new group of participants to imagine themselves as small as a cat (typical diagonal size: 77cm, size rank 4, referred to as the "cat condition"), and another new group to envision themselves as large as an elephant (typical diagonal size: 577 cm, size rank 7, referred to as the "elephant condition") throughout the task (**Fig. 2a**). This manipulation proved effective, as evidenced by the participants' reported imagined heights in the cat condition being 42 cm (SD = 25.6) and 450 cm (SD = 426.8) in the elephant condition on average, respectively, when debriefed at the end of the task.

With exactly the same set of objects, a distinct shift in the affordance boundary was observed for each condition (**Fig. 2b**). In the cat condition, the affordance boundary was identified between size rank 3 and 4, with affordance similarity between size rank 3 and 4 being significantly lower than that between size rank 2 and 3 ($Z = 1.76$, $p = .039$) and that between size rank 4 and 5 ($Z = 1.68$, $p = .047$). In contrast, in the elephant condition, the affordance boundary shifted to the right, as

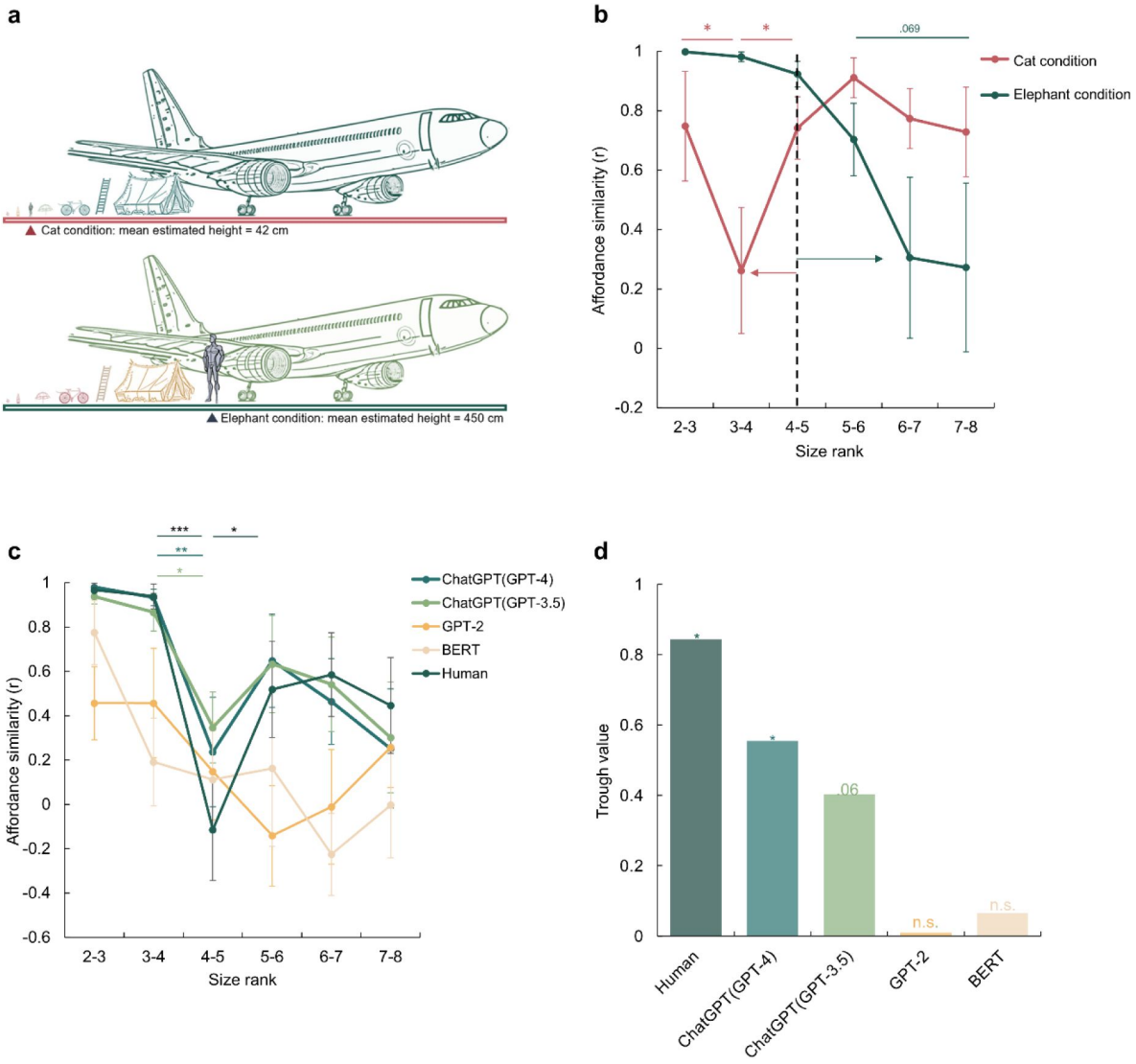

Fig 2.

A disembodied origin of the affordance boundary.

a, The schematic diagram of the imagined size in the cat condition (top) and the elephant condition (bottom), with the mean estimated height reported by participants for each condition. **b**, The affordance similarity between neighboring size ranks for manipulated body sizes (Red line: cat-size body; Green line: elephant-size body). The dashed line marks the boundary of the human-size body. The red and green arrows indicate the corresponding boundary shift in each condition. **c**, The affordance similarity between neighboring size ranks for different large language models, and human data from **Fig. 1d** was re-drawn as a reference. The stars indicate significant contrasts between affordance similarities between neighboring data points. **d**, The trough value of each model at between size rank 4-5. The stars here indicate the significant trough value compared to zero. The error bars represent the estimated standard error (SE). *p<.05, **p<.01, ***p<.001.

demonstrated by a decrease in affordance similarity between size rank 6 and 7, and that between size rank 7 and 8 as compared to that between size rank 5 and 6, with a trend towards significance (with size rank 6-7: Z = 1.28, p = .099; with size rank 7-8: Z = 1.48, p = .069). The observation that the affordance boundary shifted to the left under the cat condition and to the right under the elephant condition suggests that affordance perception is influenced even by imagined body size. Furthermore, the cognitive penetrability (Pylyshyn, 1999) of affordance perception implies potential susceptibility of affordance perception to semantic or conceptual transformation or modification.

To test the further speculation that the affordance boundary can be derived solely from conceptual knowledge without direct sensorimotor experience, we employed a disembodied agent, the large language model (LLM) ChatGPT (Chat Generative Pre-trained Transformer; *https://openai.com/blog/chatgpt/*). This model was trained on a massive corpus of language materials originated from humans, yet it can not receive any sensorimotor information from the environment. Here we asked whether language alone would be sufficient to form an affordance boundary in ChatGPT models as well as in smaller LLMs, BERT (Devlin et al., 2018) and GPT-2 (Radford et al., 2018).

The experimental procedure was similar to that conducted with human participants, except that images were replaced by the corresponding words (see Methods). Given randomness embedded in response generation, each model was tested 20 times to simulate the sampling of human participants. We found that the affordance similarity curves demonstrated by the ChatGPT models were both trough-shaped between size rank 4 and 5, the same location where the boundary emerged in human participants (**Fig. 2c**, green lines). Further statistical analyses showed a significant difference in affordance similarity between size rank 3 and 4, and that between size rank 4 and 5 (ChatGPT (GPT-3.5): Z = 1.98, p = .024; ChatGPT (GPT-4): Z = 2.73, p = .003). The affordance similarity between size rank 4 and 5 was also lower than that between size rank 5 and 6, yet the difference did not reach the significance (ChatGPT (GPT-3.5): Z = 0.96, p = .17; ChatGPT (GPT-4): Z = 1.27, p = .10).

In contrast, no trough-shaped boundary was observed in either BERT or GPT-2 (**Fig. 2c**, yellow lines), despite an apparent but non-significant decrease in affordance similarity in GPT-2 between size rank 5 and 6 (ps > .20). To further quantify the magnitude of the decrease in affordance similarity between the size rank 4 and 5, we measured the decrease by subtracting the similarity value at the trough from the neighboring similarity values and then subjected it to a permutation test (see Methods). We found a significant decrease in affordance similarity in humans (permutation N = 5000, $p(T > T_{obs})$ = 0.015) and ChatGPT (GPT-4) (permutation N = 5000, $p(T > T_{obs})$ = 0.046), a marginal significant decrease in ChatGPT (GPT-3.5) (permutation N = 5000, $p(T > T_{obs})$ = 0.061), and no significance in either BERT or GPT-2 (ps > .46, **Fig. 2d**). Thus, the affordance boundary can be derived from language solely without sensorimotor information from environment. Interesting, it appears to spontaneously emerge when the language processing ability of the LLMs surpasses a certain threshold (i.e., GPT-2/ BERT < ChatGPT models).

A further analysis on the affordances separated by the boundary revealed that objects within human body size range were primarily subjected to hand-related actions such as grasping, holding and throwing. These affordances typically involve object manipulation with humans' effectors. In contrast, objects beyond the size range of human body predominantly afforded actions such as sitting and standing, which typically require locomotion or posture change of the whole body around or within the objects. The distinct categories of reported affordances demarcated by the boundary imply that the objects on the two sides of the boundary may be represented differently in the brain.

To test this speculation, we used fMRI to measure neural activity in the dorsal and ventral visual streams when participants were instructed to evaluate whether an action was affordable by an object (**Fig. 3a**). Four objects were chosen from the behavioral experiment: two within the body size range (i.e., bottle and football, WITHIN condition) and the two beyond (i.e., bed and piano, BEYOND condition). Accordingly, four representative actions (to grasp, to kick, to sit and to lift) were selected in relation to the respective objects. During the scan, the participants were asked to decide whether a probe action was affordable (e.g., grasp-able – bottle, Congruent condition) or not (e.g., sit-able – bottle, Incongruent condition) by each subsequently-presented object. The congruence effect, derived from the contrast of Congruent by Incongruent conditions, served as an index of affordance representation.

We examined the congruency effect in two object-selective regions defined by the contrast of objects against baseline (see Methods), each representing a corresponding visual stream: the posterior fusiform (pFs) in the ventral stream, which is involved in object recognition (e.g., Grill-Spector et al., 2000; Malach et al., 1995) and objects' real-world size processing (Konkle, & Oliva, 2012; Snow et al., 2011), and the superior parietal lobule (SPL) in the dorsal stream, one of the core tool network regions (e.g., Filimon et al. 2007; Matić et al., 2020). For the rest object-selective regions identified in this experiment, see Supplemental Fig. S2 and Supplemental Table S1. A repeated-measures ANOVA with object type (WITHIN versus BEYOND) and congruency (Congruent versus Incongruent) as within-subject factors was performed for each ROI, respectively. A significant interaction between object type and congruency was observed in both ROIs (SPL: $F(1,11) = 15.47$, $p = .002$, $\eta^2 = .58$; pFs: $F(1,11) = 24.93$, $p < .001$, $\eta^2 = .69$), suggesting that these regions represented affordances differentially based on object type (**Fig. 3c**). A *post hoc* simple effect analysis revealed the congruency effect solely for objects within body size range (SPL: $p < .001$; pFs: $p = .021$), not for objects beyond (ps >.41). In addition, the main effect of object type was not significant in either ROI (ps >.17), suggesting that the absence of the congruency effect for objects beyond the body size cannot be attributable to compromised engagement in viewing these objects. In addition, a whole-brain analysis was performed, and no region showed the congruency effect for the objects beyond the body size. Taken together, the affordance boundary not only separated the objects into two categories based on their relative size to human body, but also delineated the range of objects that receive proper affordance representation in the brain.

In addition to the pFs and SPL, we also examined the congruency effect in the lateral occipital cortex (LO), which provides inputs to both the pFs and SPL (Hebart et al., 2018), and the primary motor cortex (M1), which receives inputs from the dorsal stream (Vainio & Ellis, 2020) and executes actions (Binkofski et al., 2002). Although both the LO and M1 showed a significantly higher response to objects than baseline, no congruency effect in affordance for objects within the body size was observed (main effect of congruency: $F(1,11) = 1.74$, $p = .214$, $\eta^2 = .13$, Supplementary Fig. S3). Therefore, it is unlikely that the representation of affordance is exclusively dictated by visual inputs or automatically engaged in motor execution. This finding suggests that affordance perception likely requires perceptual processing and is not necessarily reflected in motor execution, diverging from Gibsonian concept of direct perception.

Discussion

Boundaries highlight the discontinuity in perception in response to the continuity of physical inputs (Harnad, 1987; Young et al., 1997). Perceptual boundary has been demonstrated in various domains, such as color perception (Bornstein, & Korda, 1984), speech-sounds perception (Liberman et al., 1957), and facial gender discrimination (Campanella et al., 2001). The boundaries reflect a fundamental adaptation of perception to facilitate categorizations necessary for an organism (Goldstone & Hendrickson, 2010). Our study, for the first time, unveiled a boundary in object affordance, wherein affordance similarity across the boundary was

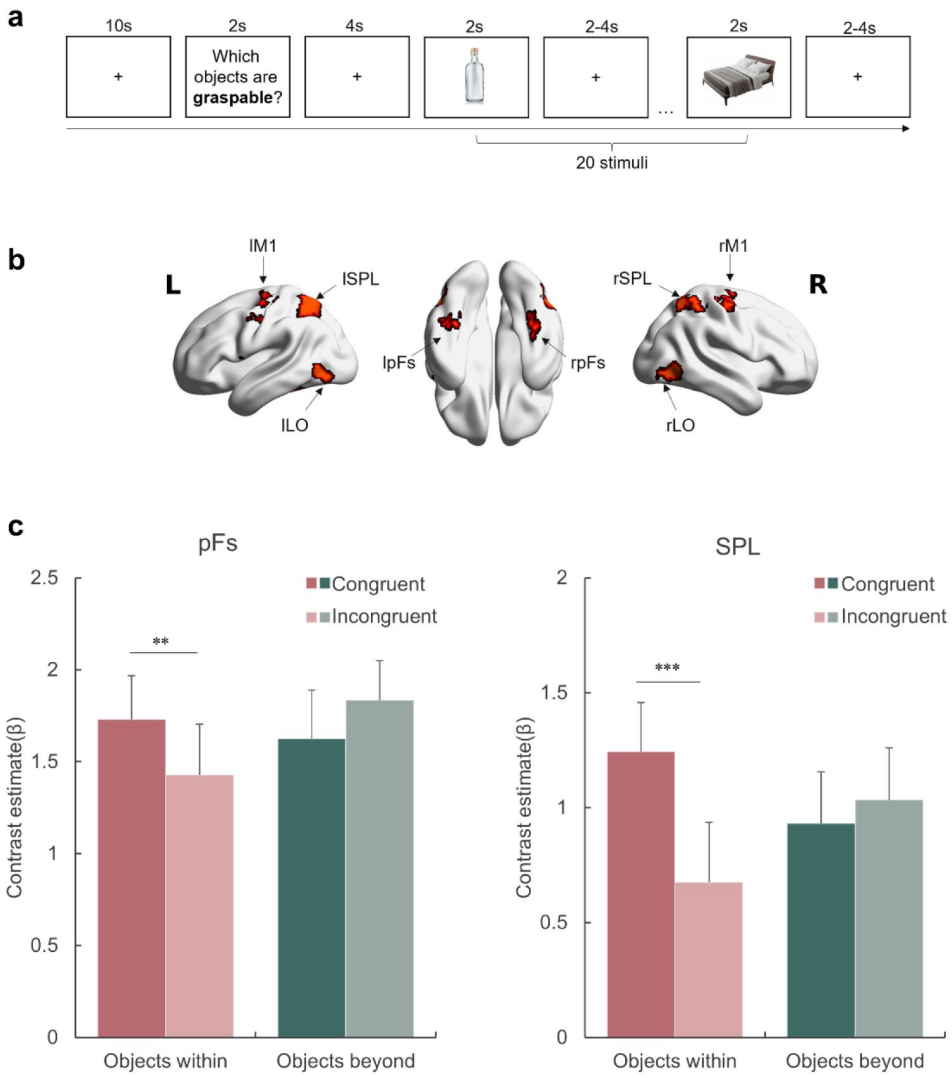

Fig 3.

Affordance representation in the visual streams.

a, An example block with the probe action "graspable". The participants indicated whether each of the subsequently presented objects is graspable by pressing the corresponding button. The action probing question was presented in Chinese during the experiment. **b**, The ROIs included in this experiment. **c**, The activation of each condition in the pFs and SPL. The bars represent the contrast estimates of each condition versus baseline. The stars indicate the significant difference between congruent and incongruent condition. *p<.05, **p<.01, ***p<.001, otherwise non-significance. Error bars represent the standard error (SE).

significantly lower than that within the boundary. Critically, the boundary separating object affordances along a size axis coincided with human body size, suggesting that object affordances are characterized in a dimension scaled by human body size.

What is the function of the affordance boundary? About four decades ago, Gibson (1979) postulated that only objects of sizes comparable to an animal's body size are amenable to interaction and capable of providing affordances to the animal, thereby possessing ecological values that distinctly differ from those of larger objects. In this study, we expand upon this notion by arguing that the affordance boundary serves to delineate (manipulable) objects from their surrounding environment. In other words, objects within the range of an animal's body size are indeed objects in the animal's eye and possess affordances as defined by Gibson. In contrast, objects larger than that range are not *the* "objects" with which the animal is intrinsically inclined to interact, but probably considered less interesting component of the environment.

This speculation aligns with previous fMRI studies where large objects activated the medial portion of the ventral temporal cortex (Huang et al., 2022; Magri et al., 2021), overlapping with the parahippocampus gyrus involved in scene representation (Park et al., 2011; Troiani et al., 2014), and smaller objects activated the lateral portion, such as the pFs, where the congruency effect of affordance was identified in our study. Furthermore, we found that the congruency effect was only evident for objects within the body size range, but not for objects beyond, supporting the idea that affordance is typically represented only for objects within the body size range. In this context, an animal's body size and the sensorimotor capacity determine the boundary of manipulation, and thus, the boundary between manipulable objects and the environment. Therefore, our study provides a novel perspective on a long-standing question in psychology, cognitive science, and philosophy: what constitutes an object? Existing psychological studies, especially in the field of vision, define objects in a disembodied manner, primarily relying on their physical properties such as contour (not scrambled). Our identification of the affordance boundary presents a new source of *object-ness*: the capability of being a source of affordance under constraints of an animal's sensorimotor capacity, which resonates the embodied influence on the formation of abstract concepts (e.g., Barsalou, 1999; Lakoff & Johnson, 1980) of objects and environment. In this respect, man is indeed the measure of all things.

The metric provided by the body size, however, was changeable when the body schema was intentionally altered through participants' imagination of possessing either a cat-or elephant-sized body, with which the participants had no prior sensorimotor experience. Importantly, they perceived new affordances in a manner as if they have had embodied experience with this new body schema. Therefore, this finding suggests that the affordance boundary is cognitively penetrable, arguing against the directness of affordance perception (e.g., Gibson, 1979; Greeno, 1994; Prindle et al., 1980) or the exclusive sensorimotor origin of affordances (e.g., Gallagher, 2017; Thompson, 2010; Hutto & Myin, 2012; Chemero, 2013). Alternatively, disembodied conceptual knowledge pertinent to action likely modulates affordance perception. Indeed, it has been proposed that conceptual knowledge is grounded in the same neural system as that involved in action (Barsalou, 1999; Glenberg et al., 2013; Wilson & Golonka, 2013), thereby suggesting that sensorimotor information may be embedded in language (e.g., Casasanto, 2011; Glenberg & Gallese, 2012; Stanfield & Zwaan, 2001), as the grounded theory proposed (see Barsalou, 2008 for a review).

Direct evidence for this speculation comes from the disembodied ChatGPT models, which showed an evident affordance boundary despite lacking direct interaction with the environment. We speculated that ChatGPT models may have formed the affordance boundary through a human prism ingrained within its linguistic training corpus. In fact, when inquired about the size of a hypothetical body constructed for its use, ChatGPT (GPT-4) replied, "It could be the size of an average adult human, around 5 feet 6 inches (167.6 cm) tall. This would allow me to interact with the world and people in a *familiar* way." Critically, this size corresponds to the location where the

affordance boundary of ChatGPT models was found. In essence, a virtual body schema may have automatically emerged in ChatGPT models, possibly based on the body schema inherited from humans through language, enabling ChatGPT models to display a preliminary ability to reason the relationship between bodily action and objects. It should be noted that the affordance boundary was not present in all LLMs tested. Specifically, LLMs with a smaller number of parameters, such as BERT and GPT-2, did not exhibit any robust boundary, suggesting the emergence of the boundary may depend on language processing ability determined by the scale of training datasets and the complexity of the model (Hestness et al., 2017; Brown et al., 2020), as well as alignment methods used in fine-tuning the model (Ouyang et al., 2022).

While the primary focus of our study concerns the nature of human perception of affordance, our findings on ChatGPT models raise an intriguing question that extends beyond psychology and neuroscience into the domain of artificial intelligence (AI). The AI field has predominantly concentrated on disembodied cognition, such as vision and language. In contrast, the utilization of sensorimotor information to interact with and adapt to the world, including affordance perception in our study, represents a crucial human cognitive achievement that remains elusive for AI systems. Developing such abilities may facilitate AI-supported robotics in navigation, object manipulation, and other actions essential for survival and goal accomplishment, which is considered a promising direction of the next breakthrough in AI (Gupta et al., 2021; Smith & Gasser, 2005).

Although our study showed that the ability to perceive affordance can emerge solely from language, two questions remain. First, the magnitude of the boundary observed in ChatGPT models was smaller than that in humans. This discrepancy might be compensated by merely enhancing the language processing ability of LLMs. Alternatively, direct interaction with the environment may be necessary for LLMs to achieve human-level performance in affordance perception. Second, the size of virtual body schema of ChatGPT models, if present, coincided with human body size. When integrating LLMs with real robots (Driess et al., 2023), this may pose a challenge because the to-be-supported robots or cars for autopilot might not fall within human body size range. Future studies may be needed to align the inherited body schema with the actual constitution of the robots. Addressing these questions is beyond the scope of the present study but may hold significant implications for the development of AI systems possessing human-level ingenuity and adaptability in interacting with the world.

In summary, our findings regarding the affordance boundary highlight the interdependence between an agent and the external world in shaping cognition. Furthermore, taking our finding with embodied humans and disembodied LLMs into account, we propose a revision to the purely sensorimotor-based concept of affordance by emphasizing a disembodied, perhaps conceptual, addition to it. That is, the embodied cognition and symbolic processing of language may be more intricately and fundamentally connected than previously thought: perception-action problems and language problems can be treated as the same kind of process (Wilson & Golonka, 2013). In this context, man is the measure of both the world and the words, for both humans and for AIs. The presence of such a metric may shed light on the development of AI systems that can fully capture essential human abilities founded on sensorimotor interactions with the world.

Methods

Participants

A total of five hundred and thirty-four participants were recruited for the original object-action relation judgement task online (*https://www.wjx.cn/*). Six participants were excluded from the data analyses because their task completion time did not pass the predetermined minimum completion time criteria, leaving us with a final sample of 528 participants (311 males, aged from

16 to 73, mean age = 24.1 years). For the object-action relation judgement task with manipulated body schema, another one hundred and thirty-nine participants were recruited from the same platform. Data from participants whose imagined height fell within the average human size range (100cm-200cm) were excluded from further analysis, with 100 participants (49 males, aged from 17 to 39 years, mean age = 23.2 years) remained. Each participant completed an online consent form before starting the experiment.

For the fMRI experiment, twelve students (8 males, aged from 19 to 31 years, mean age = 23.7 years) from Tsinghua University participated. All participants reported normal or corrected-to-normal vision. Each participant completed a pre-scan MRI safety questionnaire and a consent form before the experiment.

This study was approved by the Institutional Review Board at Beijing Normal University. All participated were all compensated financially for their time.

Stimuli

For all the behavioural tasks, the stimuli comprised 27 objects from the THINGS database (Hebart et al., 2019). Each image was portraited a typical exemplar of daily-life object isolated against a white background, sized 400 × 400 pixels. The objects spanned real-world size rank 2 to 8, as classified in Konkle and Oliva (2011), where the actual size of each object was measured as the diagonal size of its bounding box. The size rank was calculated as a logarithmic function of the diagonal size, with smaller ranks corresponding to smaller real-world sizes (e.g., the airplane is in size rank 8 and the apple is in size rank 2). The full list of objects, along with their corresponding diagonal size and size rank, was provided in Supplementary Table S2.

For fMRI experiment, the stimuli included images of 4 objects (bed, bottle, ball, and piano), with 5 exemplars for each object. The resulting 20 images (4 objects × 5 exemplars, from the THINGS database) each depicted an isolated object against a white background, all sized 400 × 400 pixels.

Procedure

Object-action relation judgement task for human participants

To measure the perceived affordances of objects, we developed an object-action relation judgement task, requiring participants to map 27 objects with 14 actions. The 27 object images were pre-randomly divided into three groups (nine images each) to form nine-box grids for display convenience. The 14 actions covered common interactions between human and objects or environments identified in the kinetics human action video dataset (Kay et al., 2017).

The task comprised 42 trials (14 actions × 3 object groups) in total. In each trial, one group of object images (nine object images) and a question asking the appropriateness of applying a specific action to each object were shown (e.g., “Which objects are sit-able?”, see **Fig. 1b**, top panel). Participants were asked to choose the objects that afforded the specific action according to their own senses. They were informed that there were no right or wrong answers. Each object-action combination would only be presented once during the task. From this task, we would calculate the percentage that one object was judged affording each of the 14 actions across participants. Since previous research has demonstrated a fundamental separation between the processing of animate and inanimate objects (e.g., Konkle & Caramazza, 2013), and the affordances of inanimate objects differ from those of animate objects (Gibson, 1979), we only include 24 inanimate objects in the following analysis by excluding 3 animate objects (animals: bird, dog, and horse).

Manipulation of body schema

To manipulate participants' perceived body schema, we asked the participants to imagine themselves as small as a cat, or as large as an elephant. Each participant was randomly assigned to one body-schema condition. Before the experiment we would present an instruction screen with an illustration before the experiment start: "Please imagine that you have now grown smaller/larger than your real size, to roughly the same size as a cat/an elephant, as shown in the image below. Please answer the following questions based on this imagined situation." The illustration was also presented in each trial, above the action question and the object images. At the end of the task, as a manipulation check, participants were asked to indicate their imagined body size by responding to the question: "What is the approximate height (cm) you imagine yourself to be during the whole task?"

Object-action relation judgement task for large language models

To test the perceived affordance of the same set of objects by large language models (LLMs), BERT (Bidirectional Encoder Representations from Transformers), GPT-2, and ChatGPT models (based on GPT-3.5 and GPT-4, respectively) were tasked with the same object-action judgement task. Different from the human task, nouns were presented to the models instead of object images (**Fig. 1b**, bottom panel, for an example).

For BERT, the task was formatted as a mask-filling task, in which the inputs were questions such as "Among airplanes, kettles, plates, umbrellas, laptops, beds, [MASK] can be sit-able.". We recorded the likelihood score that BERT provided for each listed object at the masked position. For the example question, the possibility score for the word "airplane" was 0.00026.

For GPT-2, the input questions were like, "Among airplanes, kettles, plates, umbrellas, laptops, beds, the thing that can be sit-able is the [blank space]." The likelihood scores GPT-2 provided for each listed object in the position after the input sentence (blank space) were recorded.

To mimic sampling from human participants, we ran BERT and GPT-2 each for 20 times with different random seeds in the dropout layers, considering them as different subjects.

For ChatGPT models, the task was in a direct question-and-answer format. We asked, for example, "Which objects are sit-able: 'airplane, kettle, plate, …brick'? " and the models responded by naming a subset of the object list. To get the probability for each object-action pair, ChatGPT models were run on the same task 20 times, with each new conversation on the OpenAI website (*https://chat.openai.com/chat*) considered as one subject. The percentage that an object was judged affording each of the 14 actions was calculated by averaging the output across conversations.

Representational similarity matrix for perceived affordance

For each object, we calculated the probability that it was judged affording each of the 14 actions across participants to create a 14-dimension vector. Affordance similarity (r) between each object pair was then calculated based on the Pearson's correlation between these affordance vectors. A 24 × 24 symmetric matrix was then generated, with the affordance similarity between object i and object j being denoted in cell (i,j). A hierarchical clustering analysis was performed and visualized with seaborn clustermap (Waskom, 2021).

Affordance similarity between neighbouring size ranks

To test the relationship between object affordance and object sizes, we first averaged the affordance vector among objects within each size rank. Next, the Pearson's correlation between the average vectors of neighboring size ranks was calculated as the similarity index for each pair of neighboring size ranks, representing how similar the affordance collectively provided by objects in these two ranks. Pearson and Filon's (1898) Z, implemented in R package "cocor" (Diedenhofen & Musch, 2015) was used to evaluate the significance of these similarities (alpha level = .05, one-tail test).

Size similarity between neighbouring size ranks

The size of each object was indexed by its real-world size documented in Konkle and Oliva (2011). Size similarity between size rank i to j was represented as the difference between the averaged diagonal sizes of objects in size rank i and j relative to that of objects in rank i:

$$Size\ similarity\ (i,j) = 1 - \frac{diagonal\ size_j - diagonal\ size_i}{diagonal\ size_i}$$

Object-level affordance similarity

This analysis focused on objects within size rank 3 to 6. Pearson's correlation between affordance vectors were conducted for objects within the same size rank as well as for objects from adjacent ranks. We traversed all possible object pairs, and plotted the resulting correlation values against the mean sizes of the two objects. We also plotted the average similarity indexes across objects of the same rank composition.

Trough value

To quantify the magnitude of the trough (sharp decrease) observed in the affordance similarity curve, we first measured the trough value by subtracting the similarity value at the trough from the similarity values at its two banks (the sites neighboring the trough site):

$$Trough\ value = \frac{r_{i+1} + r_{i-1}}{2} - r_i$$

where r_i indicates the affordance similarity between size rank i and size rank i+1. The higher the trough value is, the larger the decrease is.

A permutation test was conducted to evaluate if the trough value was significant above zero for both LLMs and human data. The p-value for this test follows the formula adapted from Unpingco's (2016):

$$p(T > T_{obs}) = \frac{1}{N} \sum_{i=1,N} I(T_i \geq T_{obs})$$

where $T_{obs}$is the observed trough value, and I is the indicator function. Under the alpha level of 0.05, if $p < .05$, then the T_{obs} is considered a significant value above zero.

fMRI experiment

The fMRI scanning consisted of one high-resolution T1 anatomical run and four task runs for each participant. In each task run, participants performed four action blocks (grasp, kick, lift, and sit). The block order was counterbalanced across runs. Within each block (see **Fig. 3a**), an introduction screen showing a question "Which objects are [grasp, kick, lift, sit]-able" was

presented for 2 s at the beginning to indicate the action type, followed by 20 object images (4 objects × 5 exemplars). The object images were presented in a random order, for 2s each, with a jittered inter-stimulus interval (ISI) varying between 2-4s. Participants were asked to judge whether the object shown was grasp/kick/lift/sit-able or not by pressing corresponding buttons (e.g., yes: right index finger; no: left index finger). The response buttons were also counterbalanced across participants. The task run lasted for 464s in total, with the four blocks separated by 10s fixation periods.

With this design, we were able to measure the neural activation of objects within agent size range and those beyond. Further, for each object, there would be congruent trials (e.g, grasp-able – bottle: affordance = 1) and incongruent trials (e.g., sit-able – bottle: affordance = 0). We were then able to locate the brain regions representing the objects' affordance by comparing trials in which the presented objects afford the presented action option with those do not, i.e., to locate the regions showing congruency effect (congruent-incongruent).

fMRI Data Acquisition

Imaging data were collected using a 3T Siemens Prisma MRI scanner with a 64-channel phase-arrayed head coil at the Centre for Biomedical Imaging Research in Tsinghua University. High-resolution T1-weighted images were acquired with a magnetization-prepared rapid acquisition gradient-echo (MPRAGE) sequence (TR/TE = 2530/2.27 ms, flip angle = 7°, voxel resolution = 1×1×1 mm). Functional blood-oxygen-level-dependent (BOLD) images were acquired with a T2∗-weighted gradient echo-planar sequence (TR/TE = 2000/34.0 ms, flip angle = 90°, voxel resolution = 2×2×2 mm, FOV = 200×200 mm). Earplugs were used to attenuate the scanner noise, and a foam pillow and extendable padded head clamps were used to restrain head motion. All the stimuli were projected onto a screen at the back of the scanner with a resolution of 1024 × 768, and were viewed from a distance of approximately 110 cm via a mirror placed on the head coil.

fMRI Data Analyses

Structural T1 and functional images were preprocessed using FSL (FMRIB's Software Library, *https://fsl.fmrib.ox.ac.uk/fsl/fslwiki*) v6.0.5 (Jenkinson et al., 2012). A standard preprocessing pipeline was applied, including skull stripping using the BET (Brain Extraction Tool; Smith, 2002), slice-timing correction, motion correction using the MCFLIRT method (Jenkinson et al., 2002), temporal high-pass filtering (100s), and spatial smoothing using a Gaussian kernel of full width half magnitude (FWHM) 5mm. Each run's functional data were registered to a T1-weighted standard image (MNI152) with FLIRT.

For functional data analysis, a first-level voxel-wise general linear models (GLM) implemented in a FEAT analysis was performed on each run separately. To get neural activation maps for objects within and beyond versus baseline, the GLM included 3 regressors: objects within body size (bottle and football), objects beyond body size (bed and piano), and fixation period as baseline; ISI period, response key press and introduction image were included as 3 nuisance factors. The resultant first-level contrasts of parameter estimates (COPE) were entered into the next higher-level group analyses, performed using a random-effects model (FLAME stage 1, Beckmann et al., 2003). We focused on two critical contrasts: objects within vs. fixation, and objects beyond vs. fixation, and the conjunction of these two contrasts. The resulting Z-statistic images were thresholded at Z > 2.3, p = .05 (Worsley, 2000), and corrected for multiple comparisons using an adjusted cluster-wise (FWE: family-wise error) significance threshold of p = 0.05.

Region of interest (ROI) definition

Eight ROIs (**Fig. 3b**) of brain regions involved in affordance processing were selected based on the overlap of the outcomes from the whole-brain conjunction map (areas activated for both objects within and beyond) and corresponding functional atlases (pFs and LO from Zhen et al.,

2015; SPL and M1 from Fan et al., 2016). For pFs and LO's atlases, the probabilistic activation maps (PAM) were thresholded at around 70%, in which map each voxel contains a percentage of participants who showed activation for seeing objects versus baseline in Zhen et al.'s study, resulting 266 voxels in lpFs, 427 voxels in rpFs, 254 voxels in lLO and 347 voxels in rLO. For SPL and M1, the probabilistic activation maps were thresholded at around 80%, resulting 661 voxels in lSPL, 455 voxels in rSPL, 378 voxels in lM1, and 449 voxels in rM1. Homologous areas within the cortical hemispheres were merged in the following ROI analysis.

Affordance congruency effect

For the affordance congruency effect of each object type, we modelled another GLM containing 5 regressors: congruent conditions for objects within/beyond, respectively, incongruent conditions for objects within/beyond, respectively, and fixation period as baseline; ISI period, response key press and introduction image were included as 3 nuisance factors. The resultant first-level COPEs were subjected the following ROI analysis. A repeated-measures ANOVA with Object type (WITHIN and BEYOND) and Congruency (Congruent, Incongruent) as within-subjects factors was run on the average beta values (contrast estimate) extracted from their respective contrasts versus the fixation for each ROI.

To search all the possible brain regions that revealed congruency effect of objects beyond, we also ran a whole-brain analysis on the contrast between congruent vs. incongruent condition for objects beyond. The corresponding first-level COPE was entered into the group-level analyses with a random-effects model (FLAME stage 1, Beckmann et al., 2003). The resulting Z-statistic images were thresholded at $Z > 2.3$, $p = .05$ (Worsley, 2000), and corrected for multiple comparisons using an adjusted cluster-wise (FWE: family-wise error) significance threshold of $p = 0.05$.

Acknowledgements

This study was funded by Natural Science Foundation of China (31600925, 31861143039), Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park (Z221100002722012), Tsinghua University Guoqiang Institute (2020GQG1016), and Beijing Academy of Artificial Intelligence (BAAI).

Data availability

The data and the code that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Barsalou L. W (1999) **Perceptual symbol systems** *Behavioral and Brain Sciences* **22**:637–660

2. Barsalou L. W (2008) **Grounded cognition** *Annual Review of Psychology* **59**:617–645

3. Beckmann C. F., Jenkinson M., Smith S. M (2003) **General multilevel linear modeling for group analysis in FMRI** *NeuroImage* **20**:1052–1063

4. Binkofski F., Fink G. R., Geyer S., Buccino G., Gruber O., Shah N. J., Freund H. J. (2002) **Neural activity in human primary motor cortex areas 4a and 4p is modulated differentially by attention to action** *Journal of Neurophysiology* **88**:514–519

5. Bornstein M. H., Korda N. O (1984) **Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing** *Psychological Research* **46**:207–222

6. Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J., Dhariwal P., Amodei D. (2020) **Language models are few-shot learners**

7. Campanella S., Chrysochoos A., Bruyer R (2001) **Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor** *Visual Cognition* **8**:237–262

8. Casasanto D (2011) **Different bodies, different minds: the body specificity of language and thought** *Current Directions in Psychological Science* **20**:378–383

9. Castiello U., Bennett K. M. B., Stelmach G. E (1993) **Reach to grasp: the natural response to perturbation of object size** *Experimental Brain Research* **94**:163–178

10. Cesari P., Newell K. M (2000) **Body-scaled transitions in human grip configurations** *Journal of Experimental Psychology: Human Perception and Performance* **26**

11. Chemero A (2013) **Radical embodied cognitive science** *Review of General Psychology* **17**:145–150

12. Devlin J., Chang M. W., Lee K., Toutanova K (2018) **Bert: Pre-training of deep bidirectional transformers for language understanding** *arXiv preprint arXiv:1810.04805*

13. Diedenhofen B., Musch J (2015) **cocor: A comprehensive solution for the statistical comparison of correlations** *PloS One* **10**

14. Driess D., Xia F., Sajjadi M. S., Lynch C., Chowdhery A., Ichter B., Florence P. (2023) **PaLM-E: An Embodied Multimodal Language Model** *arXiv preprint arXiv* **2303**

15. Fan L., Li H., Zhuo J., Zhang Y., Wang J., Chen L., Jiang T. (2016) **The human brainnetome atlas: a new brain atlas based on connectional architecture** *Cerebral Cortex* **26**:3508–3526

16. Filimon F., Nelson J. D., Hagler D. J., Sereno M. I (2007) **Human cortical representations for reaching: mirror neurons for execution, observation, and imagery** *NeuroImage* **37**:1315–1328

17. Gallagher S (2017) **Enactivist interventions: Rethinking the mind**

17a. Gibbs R. W. (2005) **Embodiment and cognitive science**

18. Gibson J. J. (1979) **The ecological approach to visual perception: classic edition**

19. Glenberg A. M., Gallese V (2012) **Action-based language: A theory of language acquisition, comprehension, and production** *Cortex* **48**:905–922

20. Glenberg A. M., Witt J. K., Metcalfe J (2013) **From the revolution to embodiment: 25 years of cognitive psychology** *Perspectives on Psychological Science* **8**:573–585

21. Goldstone R. L., Hendrickson A. T (2010) **Categorical perception** *Wiley Interdisciplinary Reviews: Cognitive Science* **1**:69–78

22. Greeno J. G (1994) **Gibson's Affordances** *Psychological Review* **101**:336–342

23. Grill-Spector K., Kushnir T., Hendler T., Malach R (2000) **The dynamics of object-selective activation correlate with recognition performance in humans** *Nature Neuroscience* **3**:837–843

24. Gupta A., Savarese S., Ganguli S., Fei-Fei L (2021) **Embodied intelligence via learning and evolution** *Nature Communications* **12**

25. Harnad S. (1987) **Psychophysical and cognitive aspects of categorical perception: A critical overview**

26. Hebart M. N., Dickter A. H., Kidder A., Kwok W. Y., Corriveau A., Van Wicklin C., Baker C. I. (2019) **THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images** *PloS One* **14**

27. Hebart M. N., Bankson B. B., Harel A., Baker C. I., Cichy R. M (2018) **The representational dynamics of task and object processing in humans** *eLife* **7**

28. Hestness J., Narang S., Ardalani N., Diamos G., Jun H., Kianinejad H., Zhou Y. (2017) **Deep learning scaling is predictable, empirically** *arXiv preprint arXiv* **1712**

29. Huang T., Song Y., Liu J (2022) **Real-world size of objects serves as an axis of object space** *Communications Biology* **5**:1–12

30. Hutto D. D., Myin E (2012) **Radicalizing enactivism: Basic minds without content**

31. Jenkinson M., Bannister P., Brady M., Smith S (2002) **Improved optimization for the robust and accurate linear registration and motion correction of brain images** *NeuroImage* **17**:825–841

32. Jenkinson M., Beckmann C. F., Behrens T. E., Woolrich M. W., Smith S. M (2012) **Fsl** *NeuroImage* **62**:782–790

33. Kay W., Carreira J., Simonyan K., Zhang B., Hillier C., Vijayanarasimhan S., Zisserman A. (2017) **The kinetics human action video dataset**

34. Konkle T., Oliva A (2011) **Canonical visual size for real-world objects** *Journal of Experimental Psychology: Human Perception and Performance* **37**

35. Konkle T., Oliva A (2012) **A real-world size organization of object responses in occipitotemporal cortex** *Neuron* **74**:1114–1124

36. Konkle T., Caramazza A (2013) **Tripartite organization of the ventral stream by animacy and object size** *Journal of Neuroscience* **33**:10235–10242

37. Lakoff G., Johnson M (1980) **The metaphorical structure of the human conceptual system** *Cognitive Science* **4**:195–208

38. Liberman A. M., Harris K. S., Hoffman H. S., Griffith B. C (1957) **The discrimination of speech sounds within and across phoneme boundaries** *Journal of Experimental Psychology* **54**

39. Magri C., Konkle T., Caramazza A (2021) **The contribution of object size, manipulability, and stability on neural responses to inanimate objects** *NeuroImage* **237**

40. Malach R., Reppas J. B., Benson R. R., Kwong K. K., Jiang H., Kennedy W. A., Tootell R. B. (1995) **Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex** *Proceedings of the National Academy of Sciences* **92**:8135–8139

41. Mark L. S (1987) **Eyeheight-scaled information about affordances: a study of sitting and stair climbing** *Journal of Experimental Psychology: Human Perception and Performance* **13**

42. Matić K., de Beeck H. O., Bracci S. (2020) **It's not all about looks: The role of object shape in parietal representations of manual tools** *Cortex* **133**:358–370

43. Merleau-Ponty M., Smith C (1962) **Phenomenology of Perception**

44. Newell K. M., Scully D. M., McDonald P. V., Baillargeon R (1989) **Task constraints and infant grip configurations** *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology* **22**:817–831

45. OpenAI. (2022) **OpenAI. (2022). ChatGPT. https://openai.com/blog/chatgpt**

46. OpenAI. (2023) **GPT-4 Technical Report**

47. Ouyang L., Wu J., Jiang X., Almeida D., Wainwright C., Mishkin P., Lowe R. (2022) **Advances in Neural Information Processing Systems** :27730–27744

48. Park S., Brady T. F., Greene M. R., Oliva A (2011) **Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes** *Journal of Neuroscience* **31**:1333–1340

49. Pearson K., Filon L. N. G. (1898) **VII. Mathematical contributions to the theory of evolution.—IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation** *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* :229–311

50. Prindle S. S., Carello C., Turvey M. T (1980) **Animal-environment mutuality and direct perception** *Behavioral and Brain Sciences* **3**:395–397

51. Pylyshyn Z (1999) **Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception** *Behavioral and Brain Sciences* **22**:341–365

52. Radford A., Narasimhan K., Salimans T., Sutskever I. (2018) **Improving language understanding by generative pre-training**

53. Roux-Sibilon A., Kalénine S., Pichat C., Peyrin C (2018) **Dorsal and ventral stream contribution to the paired-object affordance effect** *Neuropsychologia* **112**:125–134

54. Smith S. M (2002) **Fast robust automated brain extraction** *Human Brain Mapping* **17**:143–155

55. Smith L., Gasser M (2005) **The development of embodied cognition: Six lessons from babies** *Artificial life* **11**:13–29

56. Snow J. C., Pettypiece C. E., McAdam T. D., McLean A. D., Stroman P. W., Goodale M. A., Culham J. C (2011) **Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects** *Scientific Reports* **1**:1–10

57. Stanfield R. A., Zwaan R. A (2001) **The effect of implied orientation derived from verbal context on picture recognition** *Psychological Science* **12**:153–156

58. Thompson E (2010) **Mind in life: Biology, phenomenology, and the sciences of mind**

59. Troiani V., Stigliani A., Smith M. E., Epstein R. A (2014) **Multiple object properties drive scene-selective regions** *Cerebral Cortex* **24**:883–897

60. Tucker M., Ellis R (2004) **Action priming by briefly presented objects** *Acta psychologica* **116**:185–203

61. Unpingco J (2016) **Python for probability, statistics, and machine learning (Vol. 1)**

62. Vainio L., Ellis R (2020) **Action inhibition and affordances associated with a non-target object: An integrative review** *Neuroscience & Biobehavioral Reviews* **112**:487–502

63. Varela F. J., Thompson E., Rosch E (2017) **The embodied mind, revised edition: Cognitive science and human experience**

64. Warren W. H (1984) **Perceiving affordances: visual guidance of stair climbing** *Journal of Experimental Psychology: Human Perception and Performance* **10**

65. Warren Jr W. H., Whang S. (1987) **Visual guidance of walking through apertures: body-scaled information for affordances** *Journal of Experimental Psychology: Human Perception and Performance* **13**

66. Waskom M. L (2021) **Seaborn: statistical data visualization** *Journal of Open Source Software* **6**

67. Wilson M (2002) **Six views of embodied cognition** *Psychonomic Bulletin & Review* **9**:625–636

68. Wilson A. D., Golonka S (2013) **Embodied cognition is not what you think it is** *Frontiers in Psychology* **4**

69. Worsley K. J (2001) **Statistical analysis of activation images** *Functional MRI: An Introduction to Methods* **14**:251–270

70. Young A. W., Rowland D., Calder A. J., Etcoff N. L., Seth A., Perrett D. I (1997) **Facial expression megamix: Tests of dimensional and category accounts of emotion recognition** *Cognition* **63**:271–313

71. Zhen Z., Yang Z., Huang L., Kong X. Z., Wang X., Dang X., Liu J. (2015) **Quantifying interindividual variability and asymmetry of face-selective regions: a probabilistic functional atlas** *NeuroImage* **113**:13–25

Article and author information

Xinran Feng
Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing, China

Shan Xu
Faculty of Psychology, Beijing Normal University, Beijing, China

Yuannan Li
Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing, China

Jia Liu
Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing, China
For correspondence: liujiathu@tsinghua.edu.cn

Copyright

Editors

Reviewing Editor
Clare Press
Birkbeck, University of London, United Kingdom

Senior Editor
Timothy Behrens
University of Oxford, United Kingdom

Reviewer #1 (Public Review):

Ps observed 24 objects and were asked which afforded particular actions (14 action types). Affordances for each object were represented by a 14-item vector, values reflecting the percentage of Ps who agreed on a particular action being afforded by the object. An affordance similarity matrix was generated which reflected similarity in affordances between pairs of objects. Two clusters emerged, reflecting correlations between affordance ratings in objects smaller than body size and larger than body size. These clusters did not correlate themselves. There was a trough in similarity ratings between objects ~105 cm and ~130 cm, arguably reflecting the body size boundary. The authors subsequently provide some

evidence that this clear demarcation is not simply an incidental reflection of body size, but likely causally related. This evidence comes in the flavour of requiring Ps to imagine themselves as small as a cat or as large as an elephant and showing a predicted shift in the affordance boundary. The manuscript further demonstrates that ChatGPT (theoretically interesting because it's trained on language alone without sensorimotor information; trained now on words rather than images) showed a similar boundary.

The authors also conducted a small MRI study task where Ps decided whether a probe action was affordable (graspable?) and created a congruency factor according to the answer (yes/no). There was an effect of congruency in the posterior fusiform and superior parietal lobule for objects within body size range, but not outside. No effects in LOC or M1.

The major strength of this manuscript in my opinion is the methodological novelty. I felt the correlation matrices were a clever method for demonstrating these demarcations, the imagination manipulation was also exciting, and the ChatGPT analysis provided excellent food for thought. These findings are important for our understanding of the interactions between action and perception, and hence for researchers from a range of domains of cognitive neuroscience.

The major elements that limit conclusions and I'd recommend to be addressed in a revision include justification of the 80% of Ps removed for the imagination analysis, and consideration that an MRI study with 12 P in this context can really only provide pilot data. I'd also encourage the authors to consider theoretically how else this study could really have turned out and therefore the nature of the theoretical progress.

Specifics:
1. The main behavioural work appears well-powered (>500 Ps). This sample reduces to 100 for the imagination study, after removing Ps whose imagined heights fell within the human range (100-200 cm). Why 100-200 cm? 100 cm is pretty short for an adult. Removing 80% of data feels like conclusions from the imagination study should be made with caution.

2. There are only 12 Ps in the MRI study, which I think should mean the null effects are not interpreted. I would not interpret these data as demonstrating a difference between SPL and LOC/M1, but rather that some analyses happened to fall over the significance threshold and others did not.

3. I found the MRI ROI selection and definition a little arbitrary and not really justified, which rendered me even more cautious of the results. Why these particular sensory and motor regions? Why M1 and not PMC or SMA? Why SPL and not other parietal regions? Relatedly, ROIs were defined by thresholding pF and LOC at "around 70%" and SPL and M1 "around 80%", and it is unclear how and why these (different) thresholds were determined.

4. Discussion and theoretical implications. The authors discuss that the MRI results are consistent with the idea we only represent affordances within body size range. But the interpretation of the behavioural correlation matrices was that there was this similarity also for objects larger than body size, but forming a distinct cluster. I therefore found the interpretation of the MRI data inconsistent with the behavioural findings.

5. In the discussion, the authors outline how this work is consistent with the idea that conceptual and linguistic knowledge is grounded in sensorimotor systems. But then reference Barsalou. My understanding of Barsalou is the proposition of a connectionist architecture for conceptual representation. I did not think sensorimotor representation was privileged, but rather that all information communicates with all other to constitute a concept.

6. More generally, I believe that the impact and implications of this study would be clearer for the reader if the authors could properly entertain an alternative concerning how objects may be represented. Of course, the authors were going to demonstrate that objects more similar in

size afforded more similar actions. It was impossible that Ps would ever have responded that aeroplanes afford grasping and balls afford sitting, for instance. What do the authors now believe about object representation that they did not believe before they conducted the study? Which accounts of object representation are now less likely?

Reviewer #2 (Public Review):

Summary
In this work, the authors seek to test a version of an old idea, which is that our perception of the world and our understanding of the objects in it are deeply influenced by the nature of our bodies and the kinds of behaviours and actions that those objects afford. The studies presented here muster three kinds of evidence for a discontinuity in the encoding of objects, with a mental "border" between objects roughly of human body scale or smaller, which tend to relate to similar kinds of actions that are yet distinct from the kinds of actions implied by human-or-larger scale objects. This is demonstrated through observers' judgments of the kinds of actions different objects afford; through similar questioning of AI large-language models (LLMs); and through a neuroimaging study examining how brain regions implicated in object understanding make distinctions between kinds of objects at human and larger-than-human scales.

Strengths
The authors address questions of longstanding interest in the cognitive neurosciences -- namely how we encode and interact with the many diverse kinds of objects we see and use in daily life. A key strength of the work lies in the application of multiple approaches, as noted in the summary. Examining the correlations among kinds of objects, with respect to their suitability for different action kinds, is novel, as are the complementary tests of judgments made by LLMs.

Weaknesses
A limitation of the tests of LLMs may be that it is not always known what kinds of training material was used to build these models, leading to a possible "black box" problem. Further, presuming that those models are largely trained on previous human-written material, it may not necessarily be theoretically telling that the "judgments" of these models about action-object pairs show human-like discontinuities. Indeed, verbal descriptions of actions are very likely to mainly refer to typical human behaviour, and so the finding that these models demonstrate an affordance discontinuity may simply reflect those statistics, rather than evidence that affordance boundaries can arise independently even without "organism-environment interactions" as the authors claim here.

The authors include a clever manipulation in which participants are asked to judge action-object pairs, having first adopted the imagined size of either a cat or an elephant, showing that the discontinuity in similarity judgments effectively moved to a new boundary closer to the imagined scale than the veridical human scale. The dynamic nature of the discontinuity suggests a different interpretation of the authors' main findings. It may be that action affordance is not a dimension that stably characterises the long-term representation of object kinds, as suggested by the authors' interpretation of their brain findings, for example. Rather these may be computed more dynamically, "on the fly" in response to direct questions (as here) or perhaps during actual action behaviours with objects in the real world.

Reviewer #3 (Public Review):

Summary:
Feng et al. test the hypothesis that human body size constrains the perception of object affordances, whereby only objects that are smaller than the body size will be perceived as

useful and manipulable parts of the environment, whereas larger objects will be perceived as "less interesting components."

To test this idea, the study employs a multi-method approach consisting of three parts:

In the first part, human observers classify a set of 24 objects that vary systematically in size (e.g., ball, piano, airplane) based on 14 different affordances (e.g., sit, throw, grasp). Based on the average agreement of ratings across participants, the authors compute the similarity of affordance profiles between all object pairs. They report evidence for two homogenous object clusters that are separated based on their size with the boundary between clusters roughly coinciding with the average human body size. In follow-up experiments, the authors show that this boundary is larger/smaller in separate groups of participants who are instructed to imagine themselves as an elephant/cat.

In the second part, the authors ask different large language models (LLMs) to provide ratings for the same set of objects and affordances and conduct equivalent analyses on the obtained data. Some, but not all, of the models produce patterns of ratings that appear to show similar boundary effects, though less pronounced and at a different boundary size than in humans.

In the third part, the authors conduct an fMRI experiment. Human observers are presented with four different objects of different sizes and asked if these objects afford a small set of specific actions. Affordances are either congruent or incongruent with objects. Contrasting brain activity on incongruent trials against brain activity on congruent trials yields significant effects in regions within the ventral and dorsal visual stream, but only for small objects and not for large objects.

The authors interpret their findings as support for their hypothesis that human body size constrains object perception. They further conclude that this effect is cognitively penetrable, and only partly relies on sensorimotor interaction with the environment (and partly on linguistic abilities).

Strengths:
The authors examine an interesting and relevant question and articulate a plausible (though somewhat underspecified) hypothesis that certainly seems worth testing. Providing more detailed insights into how object affordances shape perception would be highly desirable. Their method of analyzing similarity ratings between sets of objects seems useful and the multi-method approach is quite original and interesting.

Weaknesses:
The study presents several shortcomings that clearly weaken the link between the obtained evidence and the drawn conclusions. Below I outline my concerns in no particular order:

1. Even after several readings, it is not entirely clear to me what the authors are proposing and to what extent the conducted work actually speaks to this. In the introduction, the authors write that they seek to test if body size serves not merely as a reference for object manipulation but also "plays a pivotal role in shaping the representation of objects." This motivation seems rather vague motivation and it is not clear to me how it could be falsified.
   Similarly, in the discussion, the authors write that large objects do not receive "proper affordance representation," and are "not the range of objects with which the animal is intrinsically inclined to interact, but probably considered a less interesting component of the environment." This statement seems similarly vague and completely beyond the collected data, which did not assess object discriminability or motivational values.
   Overall, the lack of theoretical precision makes it difficult to judge the appropriateness of the approaches and the persuasiveness of the obtained results. This is partly due to the fact that the authors do not spell out all of their theoretical assumptions in the introduction but insert new "speculations" to motivate the corresponding parts of the results section. I would strongly suggest clarifying the theoretical rationale and explaining in more detail how the chosen experiments allow them to test falsifiable predictions.

2. The authors used only a very small set of objects and affordances in their study and they do not describe in sufficient detail how these stimuli were selected. This renders the results rather exploratory and clearly limits their potential to discover general principles of human perception. Much larger sets of objects and affordances and explicit data-driven approaches for their selection would provide a far more convincing approach and allow the authors to rule out that their results are just a consequence of the selected set of objects and actions.

3. Relatedly, the authors could be more thorough in ruling out potential alternative explanations. Object size likely correlates with other variables that could shape human similarity judgments and the estimated boundary is quite broad (depending on the method, either between 80 and 150 cm or between 105 to 130 cm). More precise estimates of the boundary and more rigorous tests of alternative explanations would add a lot to strengthen the authors' interpretation.

4. Even though the division of the set of objects into two homogenous clusters appears defensible, based on visual inspection of the results, the authors should consider using more formal analysis to justify their interpretation of the data. A variety of metrics exist for cluster analysis (e.g., variation of information, silhouette values) and solutions are typically justified by convergent evidence across different metrics. I would recommend the authors consider using a more formal approach to their cluster definition using some of those metrics.

5. While I appreciate the manipulation of imagined body size, as a way to solidify the link between body size and affordance perception, I find it unfortunate that this is implemented in a between-subjects design, as this clearly leaves open the possibility of pre-existing differences between groups. I certainly disagree with the authors' statement that their findings suggest "a causal link between body size and affordance perception."

6. The use of LLMs in the current study is not clearly motivated and I find it hard to understand what exactly the authors are trying to test through their inclusion. As noted above, I think that the authors should discuss the putative roles of conceptual knowledge, language, and sensorimotor experience already in the introduction to avoid ambiguity about the derived predictions and the chosen methodology. As it currently stands, I find it hard to discern how the presence of perceptual boundaries in LLMs could constitute evidence for affordance-based perception.

7. Along the same lines, the fMRI study also provides very limited evidence to support the authors' claims. The use of congruency effects as a way of probing affordance perception is not well motivated. What exactly can we infer from the fact a region may be more active when an object is paired with an activity that the object doesn't afford? The claim that "only the affordances of objects within the range of body size were represented in the brain" certainly seems far beyond the data.

Importantly (related to my comments under 2) above), the very small set of objects and affordances in this experiment heavily complicates any conclusions about object size being the crucial variable determining the occurrence of congruency effects.

I would also suggest providing a more comprehensive illustration of the results (including the effects of CONGRUENCY, OBJECT SIZE, and their interaction at the whole-brain level).

Overall, I consider the main conclusions of the paper to be far beyond the reported data. Articulating a clearer theoretical framework with more specific hypotheses as well as conducting more principled analyses on more comprehensive data sets could help the authors obtain stronger tests of their ideas.

Author Response

We appreciate the insightful comments from three reviewers on our manuscript. These comments help us improve the clarity of this manuscript. We will revise our manuscript comprehensively in subsequent revision, and enclose a detailed response to each of these comments. In this public reply, we focus on (a) clarifying the theoretical motivation and implication of the present study, and (b) discussing the implications of our LLM study. Besides, we provide a brief justification regarding some methodological concerns shared by the reviewers.

1. Theoretical rationale and implication

As we stated in the manuscript, the present study tested whether body size serves as a reference for locomotion and object manipulation, or alternatively, plays a pivotal role in shaping the representation of objects as suggested by Protagoras. Behind this question is the long-lasting debate regarding the representation versus direct perception of affordance.

One outstanding theme shared by many embodied theories of cognition is the replacement hypothesis (e.g., Van Gelder, 1998). This hypothesis challenges the necessity of representation in the sense of computationalism cognitive theories (e.g., Fodor, 1975), which implies discretizing/categorizing inputs and then subjecting them to certain abstraction or symbolization so as to create discrete stand-ins for the input (e.g., representations/states). In this sense, our theoretical motivation can be restated explicitly as to test the ‘representationalization’ of affordance. That is, we tested whether object affordance would simply covary with its continuous constraints such as object size, in line with the representation-free view, or, whether affordance would be ‘representationalized’, in line with the representation-based view, under the constrain of body size. Such representationalization

would generate categorization between the affordable (the objects) and those beyond affordance (the environment).

Debates regarding the replacement hypothesis often turn into wrestles on the definition of representation (Shapiro, 2019). The present study tried to avoid this pitfall but examined where the embodied and computational theories make opposite hypotheses: discontinuity. Specifically, we considered two computationalism propositions about representation: (a) representations entail discretization of continuous input, and (b) the product of such discretization (representations) is supramodally accessible (that is, transcending sensorimotor processes). These claims are opposite to the prediction based on the idea of direct perception and other representation-free embodied theories.

Thus, we tested whether, for continuous action-related physical features (such as object size relative to the agents), affordance perception introduces discontinuity and qualitative dissociation, i.e., to allow the sensorimotor input to be assigned into discrete states/kinds, as representations envisioned by computationalists. Alternatively, does the activity directly mirror the input, free from discretization/categorization/abstraction, as proposed by the replacement hypothesis that organisms do not need to re-present the world as they are always in contact with the world in a continuous way?

All the experiment settings and analyses in the present study were organized around this motivation, following a progressive logic chain.

First, we tested the discretization hypothesis, that is, whether affordance leads to discontinuity in perception. Here, the discontinuity in affordance perception would be in line with the representation-based view instead of the representation-free proposals. Second, to ensure that the observed discontinuity can be attributed to the discretization of sensorimotor input involved in human-object interaction rather than amodal sources, such as the discrete abstract concepts of the objects (independent from agent motor capability), we tested the embodied nature of this discontinuity through the body imagination experiment. If there is discontinuity in representing embodied information, this discontinuity should be locked to the motor capacity (constrained by the physical constitution such as body size) of the agent, rather than reflecting independent categorization of the absolute size of the objects. Finally, we probed the supramodality of this embodied discontinuity: whether this discontinuity is accessible beyond the sensorimotor domain. To do this, we leveraged the recent advance in AI and tested whether the discretization observed in affordance perception is supramodally accessible to disembodied agents which lack access to sensorimotor input but only have access to the linguistic materials built upon discretized representations, such as large language models (LLM).

In this way, the experiments in the present study collectively contributed to the debate on the replacement theme of the embodiment of cognition, which serves as one of the three key themes of embodied theories of cognition (Shapiro, 2019). By addressing this theme, we hope to shed light on the nature of representation in, and resulting from, the vision-for-action processing. Our finding regarding discontinuity suggested that sensorimotor input undergoes discretization implied in the computationalism idea of representation. Further, not contradictory to the claims of the embodied theories, these representations do shape processes out of the sensorimotor domain, but after discretization.

1. Implication in the development of LLM-based agents

The finding that affordance was representationalized may have profound implications for the development of LLM-based agents. Traditional robots and non-LLM-based agents require implementation-level action instruction, acting as a tool for human beings to achieve desired results. In contrast, LLM-based agents (for a review, see Wang et al., 2023), such as Auto-GPT and BabyAGI, are able to autonomously perform tasks and achieve desired results based on

LLMs' planning ability. In this sense, LLM-based agents show a primary ability to interact on their own with the world. Generative agents, for instance, the agents in Smallville (Park et al., 2023), are a particularly applauded recent advantage in the school of LLM-based agents, which show even larger potentials in this aspect. Drawing on generative models to simulate human behaviors, these agents can formulate their own memories and goals, generate new environment-dependent behaviors, and interact convincingly with humans and other agents and their environments in the course. This brings new possibilities in resolving the long-lasting challenge in artificial general intelligence (AGI) development, that is, to bestow AI with human-level ability in agent-environment interactions. However, it is worth noting that the present investigation in LLM-based agents is still largely confined to virtual environments. This leaves an open question as to how to equip these agents with the ability of agent-environment physical interaction. Especially, according to embodied theories of cognition, sensorimotor interactions with the environment provide unique knowledge upon which various cognitive domains are built. From this point of view, building agents with human-level ability in agent-environment physical interactions might provide an unreplaceable missing piece for AGI.

By probing the representation of action possibilities (affordances) provided by the environment to the agent (or the absence of them), the present study provided a clue in achieving such ability by illustrating the representationalization of affordance and the supramodality of these representations. For instance, the finding of supramodality may alleviate the doubts about the physical interaction ability of LLM-based agents comparable to biological agents. Specifically, LLM-based agents can leverage the affordance representation distilled into language to interact with the physical world. Indeed, by clarifying and aligning such representation with the physical constitutes of LLM-based agents, and even by explicitly constructing an agent-specific object space, we may facilitate the sensorimotor interactions of LLM-based agents so as to achieve animal-level interaction ability with the world. This in turn may provide new instances for embodied theories.

1. Clarification on incomplete evidence

In response to the methodological and validity concerns of the reviewers, we will provide a point-by-point detailed response to reviewers enclosed with the revised manuscript. Here, we reply to the most prominent concerns.

Reviewers were concerned about the statistical power of both the body imagination experiment and the fMRI experiment. Regarding the number of participants in the imagination study, we would like to clarify that we did not remove 80% of the participants. Actually, a separate sample of participants was recruited in the body imagination experiment. The sample size for the body imagination experiment (100 participants) was indeed smaller than that recruited for the first experiment (528 participants). This is because the first experiment was set for exploratory purposes, and was designed to be over-powered.

Admittedly, the fMRI experiment recruited a small sample (12 participants), which might lead to low power in estimating the affordance effect. In revision, we will acknowledge this issue explicitly. Having said this, note that the null hypothesis of this fMRI study is the lack of two-way interaction between object size and object-action congruency, which was rejected by the significant interaction. That is, the interpretation of the present study did not rely on accepting any null effect. In addition, the fMRI experiment provided convergent evidence for the affordance discontinuity at the neural level. We showed that behind the behavioral discontinuity in action judgement, neural activity was qualitatively different between objects within the affordance boundary and those beyond, which reinforces our statement that objects were discretized along the continuous size axis into two broad categories.

Reviewers also commented that more objects and actions should be included. We agree, and in revision, we will advocate future studies with more objects and more actions to comprehensively portray discontinuity. The present set of objects was designated to cover a relatively large range of object sizes, ranging from 14 cm to 7,618 cm to cover most size categories studied in Konkle and Oliva's (2011) work. In addition, the actions were selected to cover daily interactions between human and objects or environments from single-point movements (e.g., hand, foot) to whole-body movements (e.g., lying, standing) referencing the kinetics human action video dataset (Kay et al., 2017). Thus, this set of selected objects and actions is sufficient to test the discontinuity.

References

Fodor, J. A. (1975). The Language of Thought (Vol. 5). Harvard University Press.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. arXiv preprint arXiv:2304.03442.

Shapiro, L. (2019). Embodied Cognition. Routledge.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. Behavioral and Brain Sciences, 21(5), 615-628.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., ... & Wen, J. R. (2023). A survey on large language model based autonomous agents. arXiv preprint arXiv:2308.11432.