Psychonomic Bulletin & Review (2023) 30:2230–2239
https://doi.org/10.3758/s13423-023-02301-6

BRIEF REPORT

No need to integrate action information during coarse semantic processing of man-made tools

Wenyuan Yu[1,2,3] · Long Ni[1,4] · Zijian Zhang[1,2] · Weiqi Zheng[5] · Ye Liu[1,2]

Accepted: 26 April 2023 / Published online: 23 May 2023
© The Psychonomic Society, Inc. 2023

Abstract

Action representation of man-made tools consists of two subtypes: structural action representation concerning how to grasp an object, and functional action representation concerning the skilled use of an object. Compared to structural action representation, functional action representation plays the dominant role in fine-grained (i.e., basic level) object recognition. However, it remains unclear whether the two types of action representation are involved differently in the coarse semantic processing in which the object is recognized at a superordinate level (i.e., living/non-living). Here we conducted three experiments using the priming paradigm, in which video clips displaying structural and functional action hand gestures were used as prime stimuli and grayscale photos of man-made tools were used as target stimuli. Participants recognized the target objects at the basic level in Experiment 1 (i.e., naming task) and at the superordinate level in Experiments 2 and 3 (i.e., categorization task). We observed a significant priming effect for functional action prime-target pairs only in the naming task. In contrast, no priming effect was found in either the naming or the categorization task for the structural action prime-target pairs (Experiment 2), even when the categorization task was preceded by a preliminary action imitation of the prime gestures (Experiment 3). Our results suggest that only functional action information is retrieved during fine-grained object processing. In contrast, coarse semantic processing does not require the integration of either structural or functional action information.

Keywords Action representation · Coarse-to-fine · Functional action · Structural action

Introduction

How are object concepts represented in the human brain? Over the past few decades, we have witnessed an accumulation of evidence for the role of sensory and motor systems in conceptual representation. Many semantic theories share the assumption that object conceptual processing depends on a hierarchical structure: Modality-specific information is progressively integrated into a more modality-independent representation (Patterson et al., 2007). For manipulable objects, such as man-made tools, accumulating evidence indicates that action information plays a role in the conceptual representation (Helbig et al., 2006; Symes et al., 2008; Vainio et al., 2008).

The dominant role of functional action representation in object recognition

For manipulable objects, we can either grasp-then-move (structural action) them or use them to fulfill their functions (functional action). It often occurs that the manner of

✉ Ye Liu
liuye@psych.ac.cn

1 State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing 100101, People's Republic of China

2 Department of Psychology, University of Chinese Academy of Sciences, Beijing 101408, People's Republic of China

3 Research Center for Applied Mathematics and Machine Intelligence, Research Institute of Basic Theories, Zhejiang Lab, Hangzhou 311121, People's Republic of China

4 Present Address: Department of Psychology, University of Pennsylvania, Philadelphia, PA 19106, USA

5 School of Psychology, Beijing Sport University, Beijing 100084, People's Republic of China

grasping an object (e.g., *gripping a calculator*) could be very different from the manner of using it (e.g., *pressing its buttons*) (Binkofski & Buxbaum, 2013; Buxbaum & Kalénine, 2010; Buxbaum et al., 2006). A number of studies have demonstrated the functional distinctions between the two types of action representation, based on which Buxbaum and Kalénine (2010) proposed the Two Action Systems (2AS) Theory. Under the 2AS Theory, structural action representation relies on processing online-input structural properties (e.g., size, shape, orientation, weight). Its activation is fast-rising and rapidly decaying. In contrast, functional action representation relies on retrieving object features stored in the long-term memory. As a result, its activation arises more slowly and is sustained for longer (Lee et al., 2013, 2018; Sakreida et al., 2016).

Previous studies have indicated that functional action representation plays a dominant role in recognition of manipulable objects as compared with structural action representation. Behavioral studies using the priming paradigm showed that the priming effect was more robust for functional than structural action in the fine-grained object processing (Ni et al., 2019a). The execution of functional but not structural action was facilitated by the preceding object name associated with a congruent action (Bub & Masson, 2012; Bub et al., 2008; Masson et al., 2008). Furthermore, eye-tracking (Lee et al., 2013) and event-related potential (ERP) (Lee et al., 2018) studies found that the activation of functional action representation arose more slowly but was sustained for longer than that of structural action representation, suggesting that the retrieval of functional action knowledge requires retrieving information from the semantic memory. However, these studies supporting the dominant role of functional action representation focused on fine-grained semantic processing, i.e., naming an object at its basic level. The roles of the two types of action representation in coarse semantic processing remain to be elucidated.

Coarse and fine-grained semantic processing of object concepts

According to the distributed model account, concepts are composed of features that vary in their distinctiveness. When only less-distinctive (shared) features are accessed, objects can be recognized at a superordinate level, indicative of coarse semantic processing. In contrast, when the more-distinctive features are accessed, objects can be recognized at a basic level, indicative of fine-grained semantic processing (McRae et al., 2005; Taylor et al., 2011, 2012). Behavioral evidence showed that the less-distinctive and more-distinctive features facilitated object categorization and naming, respectively (Taylor et al., 2012). Moreover, neuroimaging studies showed that the less-distinctive features were represented in the posterior ventral stream at the early processing stage, and the more-distinctive features were represented in the anterior ventral stream at the late processing stage (Clarke, 2015; Clarke et al., 2013; Tyler et al., 2013). Therefore, conceptual processing progresses from a coarse to a fine-grained analysis (Clarke et al., 2013; Taylor et al., 2012). The goal of coarse processing is to recognize an object as a member of a category. In contrast, the goal of fine-grained processing is to recognize an object as a specific object.

Are the two types of action representation involved in coarse semantic processing?

Functional action is a more distinctive feature as it is shared only by a limited number of objects. For example, the gesture "*typing*" is a distinctive and diagnostic feature of a keyboard. Thus, functional action representation might not be involved in coarse semantic processing that requires accessing only less-distinctive features (Clarke et al., 2015; McRae et al., 2005).

We considered two possibilities regarding the role of structural action in coarse semantic processing. The first possibility is that structural action representation is only involved in the coarse semantic processing because structural actions tend to be shared by a larger number of objects (e.g., we hold a bat and a hammer with the same grasping action). This hypothesis is supported by studies showing that structural gestures facilitated the categorization of target objects associated with similar structural gestures (Borghi et al., 2007; Godard et al., 2019; Vainio et al., 2008). However, the interpretation of the priming effect in these studies is still in dispute. Borghi et al. (2007) found that the structural action priming effect was observed only after preliminary motor training about gestures. More evidence is needed to support the notion that structural action representation is involved in semantic processing where no motor response is required. Vainio et al. (2008) displayed a series of hand gesture photos to create dynamic grasping gestures, with the last one being partially transparent and superimposed on the target object. Therefore, the priming effect observed in this study could be caused by visual interference between the gesture and the object images rather than the facilitation of structural action representation. Godard et al. (2019) reported that a congruent grasping action facilitated the categorization of manipulable artificial but not natural objects. The result was interpreted as indicating that the grasping action primes facilitated the recognition of artificial objects by activating their associated functional action representations. However, the conclusion is premature without disentangling structural and functional actions.

An alternative hypothesis is that structural action representation is not involved in any stages of semantic processing. Instead of being part of the conceptual representation, the structural action is more likely to be an instance of

affordance offered by objects in the current context and perceived directly, thus requiring no pre-existing knowledge (Chong & Proctor, 2020; Gibson, 1979).

Present study

The present study aimed to dissociate the roles of structural and functional action representations in object naming (fine-grained semantic processing) and object categorization (coarse semantic processing). Using a priming paradigm, we tested whether object naming (Experiment 1) and object categorization (Experiment 2) can be influenced by a preceding hand gesture. To enhance the potential priming effect, we asked participants to imitate the priming gestures before performing the categorization task in Experiment 3.

Method

Power analysis

We performed a simulation-based power analysis to estimate the sample size required to detect the main effect of action congruency and the interaction between action type and action congruency. We used *mixedpower* R package (Kumle et al., 2021) on the data published in our prior study (Ni et al., 2019a, Experiment 3). This experiment presented a prime gesture and a target object sequentially and found a significant main effect of action congruency and a significant interaction between action congruency and action type. We first fitted the linear mixed model on the data, where action type, action congruency, and their interaction were fixed effects, and the random intercepts of participants and items were included. We then estimated the power on a range of different sample sizes (i.e., 18, 24, 32, 36, 48) using the function *R2power*(), which allows us to simulate sample sizes for one random variable (i.e., items whose number was 42) and simultaneously change the level of another random variable (i.e., participants). We defined a *t*-value of 2 as the threshold of significance (Baayen et al., 2008). The simulation process was run with 1,000 repetitions. Apart from data-based simulation, we also determined the smallest effect size of interest (SESOI) by reducing all beta coefficients for fixed effects by 15%. We adopted a strategy resulting in > 80% power for the interested effects (i.e., the main effect of action type and the interaction between action type and action congruency). Data-based simulation suggested that we would need at least 24 participants to detect the interested effects, and SESOI simulation suggested that we would need at least 30 participants to detect the interested effects.

Participants

Eighty-four undergraduate and graduate students (average age 21.3 ± 2 years, 50 females), naïve to the experiment purpose, participated in the present study (Experiment 1: 30 participants, average age 21.3 ± 2.2 years, 16 females; Experiment 2: 30 participants, average age 21.4 ± 2.6 years, 19 females; Experiment 3: 24 participants, average age 21.3 ± 2.3 years, 15 females). All participants had a normal or corrected-to-normal vision and were right-handed. Each participant only participated in one of the three experiments. They gave informed consent before the experiment and were paid after they completed the experiment. The study adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences.

Stimuli and design

Stimuli Priming stimuli were 45 movie clips (1,280 × 720 pixels) with a duration of 2,500 ms, including 11 depicting structural gestures and 34 depicting functional gestures. All the movie clips showed a right hand performing an action to use (functional action) or grasp (structural action) an unseen object in front of a white background. We invited a female hand model to mimic structural or functional actions towards target objects and recorded her hand actions using a camera. As opposed to our prior study (Ni et al., 2019a) and others (Borghi et al., 2007; Godard et al., 2019; Vainio et al., 2008) that used only two structural gestures (i.e., power and precise grasp gestures), we used 11 structural gestures. Specifically, structural gestures in the present study matched the objects in their volume, shape, orientation, and weight. Structural gestures included *power grasp gestures with palm down/inward* (suitable for grasping large round objects), *handle-grasp gestures with palm down/forward/inward* (suitable for grasping large elongated objects), *precise grasp gestures with palm down/inward* (suitable for grasping small objects), a *power book-hold gesture with palm inward* (suitable for holding heavy flat objects), a *precise book-hold gesture* with *palm inward* (suitable for holding light flat objects), a *pillar-grasp gesture with palms of two hands inward* (suitable for lifting heavy cylindrical objects), and a *lift gesture with two hands* (suitable for lifting heavy large objects). The functional gestures were more varied in two aspects than those in our prior study (Ni et al., 2019a). First, besides one-hand gestures, the functional gestures depicted in movie clips contained two-hand gestures, such as fluting and typing. These gestures were suitable for objects requiring two hands to manipulate, such as the flute and keyboard. Second, our functional gestures were more distinctive and closely matched with the target objects. For example, in our prior study (Ni et al., 2019a), the bottle opener and compass were

paired with identical functional gestures, i.e., twisting the unseen object with the thumb, index finger, and middle finger. We paired the bottle opener and compass with similar but different functional gestures to account for the difference in thickness between the top parts of the two objects. Specifically, we used a twisting gesture for both objects, but in the gesture paired with the bottle opener, the distance between the thumb and index finger was farther apart than in the gesture paired with the compass (see details in the Online Supplementary Materials (OSM)).

The target stimuli were 42 man-made tool photos selected from the China Image Set (Ni et al. 2019b) and transformed into grayscale images. All stimuli were displayed on a CRT screen with a resolution of 1,024 × 768 at a refresh rate of 100 Hz. The priming videos were presented with an average visual angle of 13.29° horizontally and 7.44° vertically. The target object photos were presented with an average visual angle of 6.6° at a viewing distance of 85 cm.

Design We paired each target object (e.g., a calculator) with a congruent (grasping) and an incongruent (pinching) structural action, as well as a congruent (pressing with index finger) and an incongruent (hammering) functional action. Thus, there were a total of 84 structural prime-target pairs that were divided into four blocks (two blocks contained 22 trials and another two blocks contained 20 trials) and a total of 84 functional prime-target pairs that were also divided into four blocks (two blocks contained 22 trials and another two blocks contained 20 trials).

In each block, half of the pairs were (functional or structural) action congruent and the other half were action incongruent. To prevent observers from seeing an object twice within the same block, we paired each target object with a congruent action in one block and an incongruent action in another. Prime-target pairs in each block were randomly presented. Each block comprised only one type of (structural or functional) action pairs. Half of the participants completed the four structural action blocks first, with the other half completing the four functional action blocks first.

Action congruency ratings Thirty participants who did not participate in Experiments 1–3 rated the structural and functional action congruency of the prime-target pairs on a 6-point Likert scale, in which "1" indicated a poor match between the structural/functional gestures displayed in the movie clip and the action implied by the target object, and "6" indicated a perfect match between them. As expected, the congruency rating was significantly higher for the congruent (4.71 ± 0.635) than the incongruent (2.20 ± 0.837) structural action target-prime pairs, $t(1,41) = 14.467$, $p < 0.001$. It was also significantly higher for the congruent than the incongruent functional action prime-target pairs, $t(1,41) = 54.623$, $p < 0.001$ (see Fig. 1).

We conducted two additional norming experiments with participants who had not previously taken part in the action congruency rating. Seventeen participants (19.4 ± 0.94 years of age, 15 females) rated the similarity between target objects' congruent structural and functional gestures on a 6-point Likert scale ("1" = low similarity; "6" = high similarity). The results showed that the average similarity of target objects' congruent structural and functional gestures was very low (1.88 ± 0.46), indicating that the object's congruent functional gestures were distinct from their congruent structural gestures. Sixteen participants (19.8 ± 2.54 years of age, 15 females, nine of them also participated in the first norming experiment) rated to what extent a gesture was to *grasp* or to *use* an object on a 6-point Likert scale ("1" = absolutely not to grasp/use an object; "6" = definitely to grasp/use an object). Structural gestures scored high on the *to-grasp* dimension (5.44 ± 0.24) but low on *the to-use* dimension (2.43 ± 0.36). The results were reversed for the functional gestures (*to-grasp* dimension: 2.4 ± 1.037; *to-use* dimension: 5.1 ± 0.83) (see details in OSM).

Natural object filler trials In the categorization task of Experiments 2 and 3, we added an equal number of filler

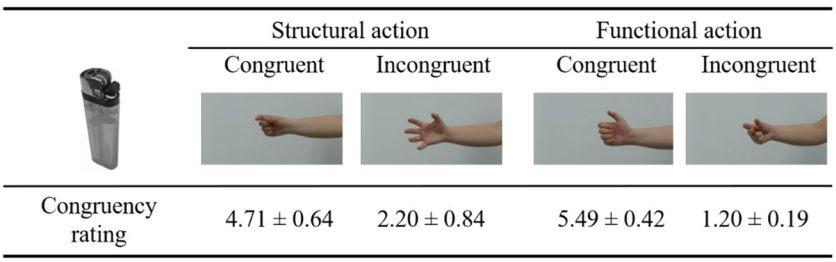


Fig. 1 Example pairs of target objects and their corresponding congruent vs. incongruent functional and structural action hand gestures and their action congruency ratings. (mean ± standard deviation). The congruency rating was significantly higher for congruent than incongruent functional and structural action prime-target pairs

pairs in which the targets were replaced with 42 natural objects (i.e., vegetables and fruits). These natural object photos were also selected from the China Image Set (Ni et al. 2019b) and transformed into grayscale images. Each natural target object was paired with four priming stimuli: two structural actions and two functional actions. The priming stimuli were the same set of action movie clips from Experiment 1. In addition, the movie clips were randomly paired with natural object photos. Each movie clip was paired with the natural target objects with the same frequency as they were paired with the man-made target objects.

Thus, in Experiments 2 and 3, there were a total of 168 structural prime-target pairs (half were filler pairs) that were divided into four blocks (two blocks contained 44 trials and another two contained 40 trials) and a total of 168 functional prime-target pairs (half were filler pairs) that were also divided into four blocks (two blocks contained 44 trials and another two contained 40 trials).

Procedure

The trial procedure was similar to our prior study (Ni et al., 2019a). Each trial began with a fixation cross shown on the screen center for 500 ms. The priming video was then presented for 2,500 ms and replaced by a blank screen of 70 ms. A target object was then displayed for 80 ms and immediately replaced by a white noise mask (see Fig. 2). Participants were asked to respond once they saw the target object.

In Experiment 1, participants were asked to vocally name the target objects as quickly and accurately as possible. Naming latencies were recorded by a voice key of E-prime 2.0 automatically. The experimenter recorded the naming accuracy with a button box connected to the computer. Participants were reminded to refrain from producing any meaningless sounds that may trigger the voice key, which could result in invalid response latencies.

In Experiments 2 and 3, participants were asked to judge whether the target object was man-made or natural by pressing the corresponding key as quickly and accurately as possible. Half of the participants pressed the left and the right button to report man-made and natural target objects, respectively, and the other half responded with the opposite hand-to-category arrangement. Reaction times and accuracy were recorded using E-prime 2.0.

In addition, Experiment 3 began with an action imitation task, in which participants imitated all the functional and structural hand actions that were later displayed in the priming videos. Specifically, participants first watched each randomly presented movie clip depicting a hand action and then imitated it without time pressure. The experimenter provided standard oral instructions to ensure accurate imitations and corrected their actions if necessary. The oral instructions did not contain any information on the experiment's purpose or the experimental stimuli.

Before the formal experiment, participants named (Experiment 1) or categorized (Experiments 2 and 3) all the target objects twice with feedback to ensure they were familiarized with the target objects and could recognize and name them correctly. To familiarize participants with the trial procedure, they also completed ten (Experiment 1) or 20 (Experiments 2 and 3) practice trials in which the target objects were not included in the experiment. No feedback was given in the experiment session. The three experiments took approximately 30, 45, and 55 min to complete, respectively.

Data analysis

For each participant's data, we removed trials with wrong responses (Experiment 1: 4.35%; Experiment 2: 2.16%; Experiment 3: 1.98%), trials with responses triggered by meaningless sounds (only in Experiment 1, 0.06%), and trials with reaction times (RTs) below and above three standard deviations (Experiment 1: 1.47%; Experiment 2: 2.03%; Experiment 3: 1.76%) of the average response latency. We only analyzed RTs, as the accuracy of each participant reached the ceiling (over 90%) in the three experiments. In addition, since natural target objects were randomly paired with movie clips regardless of their action congruency, these filler trials in Experiments 2 and 3 were removed from the following analysis.

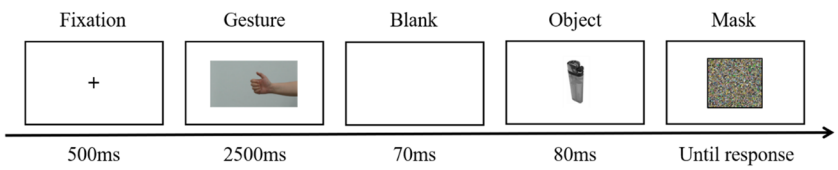


Fig. 2 Trial procedure. Each trial started with a fixation cross for 500 ms, followed by a priming hand gesture presented for 2,500 ms and then replaced by a blank screen of 70 ms. A target object was then presented for 80 ms and then replaced by a white noise mask. In Experiment 1, participants were required to speak out the name of the target object. In Experiments 2 and 3, participants were required to judge whether the target object was a man-made or natural object by pressing the corresponding keys on a reaction box. Participants were asked to respond as quickly and accurately as possible once the target appeared on the screen

We fitted the linear mixed-effects (LME) model to RTs. The model included action type (structural and functional action), action congruency (congruent and incongruent), task (naming and categorization), and their interaction as fixed effects, and participants and items (target objects) as random effects. Action type and action congruency were within-subject factors, and task was a between-subject factor. To test the fixed effect of task, we fitted one model to data from Experiments 1 (naming task) and 2 (categorization task) jointly, and another model to data from Experiments 1 and 3 (categorization task with a preliminary gesture imitation) jointly. In addition, since the difference in action congruency rating score between the congruent and incongruent structural gesture-target pairs was smaller than that of the functional gesture-target pairs, we added the fixed effect of action congruency rating score in the LME model to control for the potential confounding effect of the rating score.

We used the packages *lmerTest* (Kuznetsova et al., 2017) and *emmeans* (Lenth et al., 2019) in R Project for Statistical Computing environment, Version 4.1.2 (R Core Team, 2021). The LME analysis was carried out with *lmer()* function using restricted maximum likelihood. Models included random intercepts of participants and items. Random slopes were not included in models because they led to convergence issues. We used *emmeans()* function for further simple effect tests. The R code for all analyses can be found in Text S1 in the supplementary materials.

Results and discussion

LME for data in Experiments 1 and 2

There were significant fixed effects of action type (β = -63.939, SE = 11.085, t = -5.768, $p < .001$) and task (β = 203.742, SE = 38.623, t = 5.275, $p < .001$). However, the fixed effects of action congruency (β = 16.078, SE = 23.971, t = 0.671, p = .502) and action congruency rating score (β = 2.402, SE = 5.048, t = 0.476, p = .634) were not significant. In addition, the fixed effect of interaction between task, action type, and action congruency was significant (β = -50.449, SE = 20.880, t = -2.416, p = .016). The results of simple effect tests (Table 1) showed that in the naming task, the responses to objects paired with congruent functional gestures were faster than responses to objects paired with incongruent functional gestures (Fig. 3). In contrast, the naming responses to objects were not influenced by structural gestures. Moreover, neither structural nor functional gestures influenced the responses to objects in the categorization task (Fig. 4, left panel).

In the categorization task, the responses were slower when the target objects were paired with functional than structural gestures (β = 67.62, SE = 7.36, t = 9.190, $p < .001$). In contrast, there was no difference in the latency of naming the target objects paired with structural and functional gestures (β = 7.39, SE = 7.45, t = 0.993, p = .321).

Table 1 Contrasts for the model with the interaction between task, action type, and action congruency on reaction times (RTs) in Experiments 1 and 2

Contrast	β	SE	df	t	p
Naming: Functional congruent-incongruent	-60.68	24.1	9270	-2.523	.012
Naming: Structural congruent-incongruent	-2.86	16.5	9428	-0.174	.862
Categorization: Functional congruent-incongruent	-16.08	24.0	9268	-0.670	.503
Categorization: Structural congruent-incongruent	-8.71	16.4	9426	-0.532	.595

Degrees of freedom were corrected by the Kenward-Roger method

To explore whether the slower response to objects paired with functional gestures in the categorization task was due to the action representation, we fitted the LME model to RTs from natural object filler trials. The model included action type as the fixed factor and participants and items as the independent random effects on the intercept. The analysis showed that the categorization responses were still significantly slower to natural objects paired with functional than structural gestures, even though these natural objects were unrelated to the structural and functional gestures (β = -42.512, SE = 5.518, t = -7.705, $p < .001$) (Fig. 4, right panel). We hypothesized that the slower response to the functional action prime-target pairs may be attributed to a stronger forward masking effect caused by the functional action movies that contained more complex visual features (Alam et al., 2014).

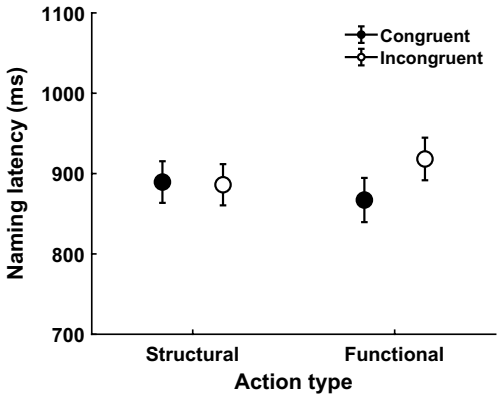


Fig. 3 The mean naming latencies in the four conditions averaged across 30 participants in Experiment 1. Error bars indicate 1 standard error above and below the means

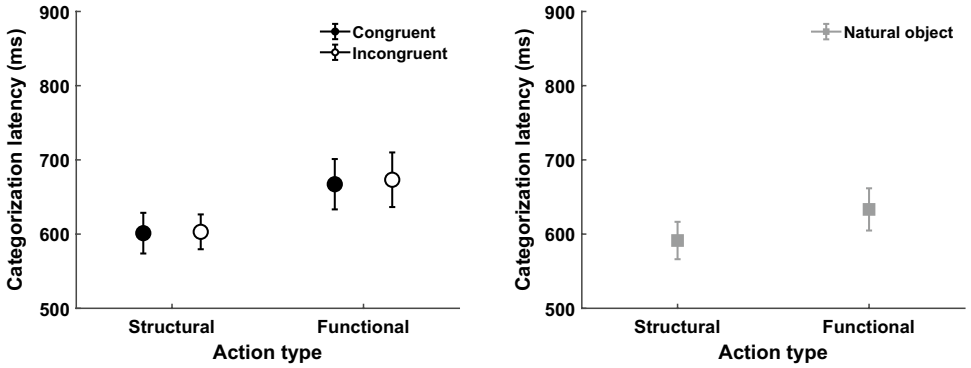


Fig. 4 Left panel: Mean latencies of categorizing *man-made* objects in the four conditions averaged across 30 participants in Experiment 2. Right panel: Mean latencies across 30 participants of categorizing *natural* objects that were paired with structural or functional gestures in the natural object filler trials. Error bars indicate 1 standard error above and below the means

Table 2 Contrasts for the model with the interaction between task, action type, and action congruency on reaction time (RT) in Experiments 1 and 3

Contrast	β	*SE*	*df*	*t*	*p*
Naming: Functional congruent-incongruent	-59.57	22.2	8495	-2.684	.007
Naming: Structural congruent-incongruent	-2.14	15.1	8552	-0.142	.887
Categorization: Functional congruent-incongruent	-12.13	22.6	8504	-0.536	.592
Categorization: Structural congruent-incongruent	-9.25	15.6	8562	-0.591	.554

Degrees of freedom were corrected by the Kenward-Roger method

LME for data in Experiments 1 and 3

There were significant fixed effects of action type (β = -47.443, *SE* = 10.844, *t* = -4.375, $p < .001$) and task (β = 243.473, *SE* = 36.187, *t* = 6.728, $p < .001$). However, the fixed effects of action congruency (β = 12.128, *SE* = 22.609, *t* = 0.536, $p = .592$) and action congruency rating score (β = 2.170, *SE* = 4.709, *t* = 0.461, $p = .645$) were not significant. In addition, the fixed effect of interaction between task, action type, and action congruency was significant (β = -54.552, *SE* = 19.509, *t* = -2.796, $p = .005$). The results of simple effect tests (Table 2) showed that in the naming task, the responses to objects paired with congruent functional gestures were faster than those paired with incongruent functional gestures (Fig. 3). In contrast, structural gestures did not influence the naming responses to objects. And neither structural nor functional gestures influenced the responses to objects in the categorization task (Fig. 5, left panel).

The categorization responses to objects paired with functional gestures were slower than those paired with structural gestures (β = 48.88, *SE* = 7.25, *t* = 6.745, $p < .001$). In contrast, there was no difference in the latency of

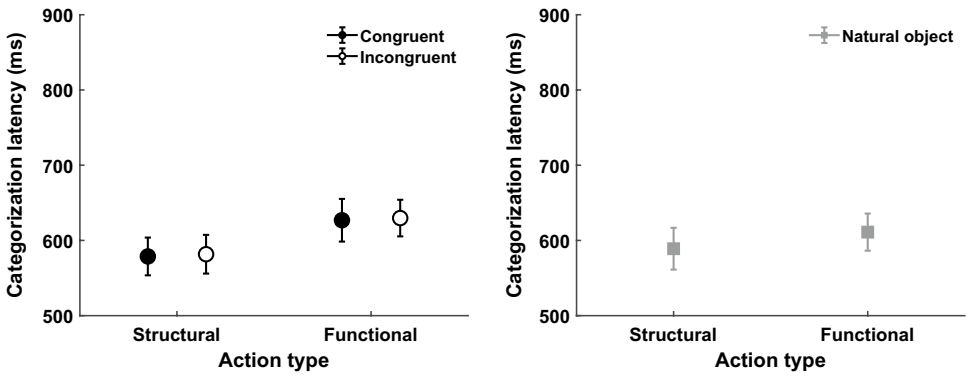


Fig. 5 Left panel: Mean categorization latencies on *man-made* objects in the four conditions averaged across 24 participants in Experiment 3. Right panel: Mean categorization latencies on *natural* objects when objects were paired with structural and functional gesture average across 24 participants in Experiment 3. Error bars indicate 1 standard error above and below the means

naming the target objects paired with structural and functional gestures ($\beta = 7.58$, $SE = 6.57$, $t = 1.154$, $p = .248$).

Once again, the LME model analysis on the natural object filler trials showed that the categorization responses to natural objects paired with functional gestures were still significantly slower than those paired with structural gestures, even though the natural objects were unrelated to the structural and functional gestures ($\beta = -21.98$, $SE = 5.82$, $t = -3.778$, $p < .001$) (Fig. 5, right panel).

General discussion

We investigated whether structural and functional action representations are involved in coarse (categorization task) and fine-grained (naming task) semantic processing. We found a functional action priming effect in the naming task (Experiment 1). In the categorization task, neither the functional nor structural action priming effect was observed (Experiment 2). In Experiment 3, we included an action imitation task to enhance the potential action priming effect by pre-activating the corresponding action system. Participants imitated all the priming gestures before the categorization task to better recall the action experience when viewing gestures in priming movie clips. Despite these efforts, we found no action-priming effect in the categorization task.

Functional action representation is involved in fine-grained semantic processing

Results of the present study support the notion that functional action representation is an intrinsic part of an object's conceptual knowledge. Furthermore, our findings reveal that the involvement of functional action representation in object processing depends on the levels of semantic processing. Specifically, it is the fine-grained but not coarse semantic processing that requires accessing functional action representation. This finding is consistent with our previous study (Ni et al., 2019a) along with a few others that employed tasks requiring fine-grained semantic processing, such as object naming (Helbig et al., 2006; Kiefer et al., 2011), word-picture matching (Campanella & Shallice, 2011), noun comprehension (Bub & Masson, 2012; Bub et al., 2008), and sentence comprehension (Masson et al., 2013; Masson et al., 2008). The involvement of functional action representation in fine-grained semantic processing can be explained by the fact that functional actions tend to be more- distinctive features that are unique or shared only by a few objects.

The absence of functional action representation in the categorization task is consistent with the findings of a development study that reported that the congruent functional actions facilitated object naming across all age groups. Interestingly, the functional action priming effect in object categorization has been observed in children but not adults, and the effect decreases with age in children (Mounoud et al., 2007), suggesting that functional action representation plays a role in coarse semantic processing during early development. An alternative explanation for the absence of a functional action priming effect in the categorization task is that functional gestures may take longer to influence semantic processing, and the short categorization latencies may prevent the effect from being observed.

The shorter categorization latencies compared to naming latencies provide support for the alternative interpretation. However, the level of semantic processing required to access more distinctive features depends on the task demands (Clarke et al., 2013; Taylor et al., 2012). In the categorization task, semantic processing may stop once the less distinctive features required for object categorization are accessed, which could explain the absence of a functional priming effect. It is important to note that the less distinctive features may not include structural action representations, as our findings showed no structural action priming effect. Thus, functional action representation may require more time to be accessed, but only when fine-grained processing is needed.

Structural action representation is not involved in semantic processing

The findings that the structural action priming effect was observed in neither the naming nor the categorization task support our second hypothesis that structural action representation is not involved in any stages of semantic processing. Our findings are in line with the 2AS theory that posits that structural action relies on processing the online inputs of objects' structural attributes. The representation of structural action is only dependent on the "rapidly decaying sensorimotor memory" and thus not an intrinsic part of the conceptual knowledge (Binkofski & Buxbaum, 2013; Buxbaum & Kalénine, 2010; Buxbaum et al., 2006). Therefore, we consider here that structural action representation is more likely an instance of affordance that can be directly offered by the incoming visual properties (Osiurak et al., 2017).

Our results nevertheless contradict some of the previous studies (Borghi et al., 2007; Godard et al., 2019; Vainio et al., 2008). The discrepancy between our and previous studies may originate from the priming gesture stimuli. Previous studies used only two structural gestures (i.e., power and precise grasp gestures) that matched the objects only in volume. Due to repeated presentations, the two gestures could be easily transformed into semantic representation, i.e., grasping large and small objects. The action priming effect observed in those studies might be a semantic priming effect caused by the facilitation of semantic information derived from gestures. In contrast, the 11 structural gestures used in the present study

matched the objects in volume, shape, orientation, and weight, and thus were hardly transformed into semantic representation. Therefore, the structural gestures in the present study activated affordance that is independent of semantic representation, resulting in the absence of structural action priming effect.

Alternatively, the transient temporal dynamics of structural action representation (Lee et al., 2018) may prevent it from being observed in behavioral data. High temporal resolution methods like EEG or MEG can be used to confirm whether the structural action priming effect can be observed in coarse and fine-grained semantic processing. Furthermore, future developmental studies are needed to explore the roles of two action representations in the development of conceptual knowledge, as previous research suggests that action information facilitates object naming in children (Kalenine et al., 2009).

When to integrate action-modality information in coarse-to-fine semantic processing?

Embodied cognition theory suggests that conceptual representation, particularly for concrete object concepts, is grounded in sensory and motor experiences and stored in the high-order sensorimotor and associated cortex. Modality-specific representations are then integrated into an area that is proposed to be the anterior temporal lobe (Martin, 2016; Patterson et al., 2007; Pulvermuller, 2013). On the other hand, the conceptual distribution account proposes that semantic processing starts with coarse processing followed by fine-grained processing (Clarke, 2015). The coexistence of coarse-to-fine and modality-specific-to-independent hierarchical structures is supported by the finding that the anterior temporal lobe is the key region of both abstract concept representation and cross-modal integration (Taylor et al., 2009). It is uncertain, however, at which stage modality-specific action representation is integrated into the conceptual representation under the coarse-to-fine processing model. In this study, we combine the view of the coarse-to-fine hierarchical model and modality-specific-to-independent hierarchical model on conceptual representation. Our findings reveal that action information, a modality-specific feature, is involved in the conceptual representation of manipulable objects. Moreover, the action-modality feature only becomes involved when the object processing reaches the fine-grained semantic level, extending the account of coarse-to-fine-grained semantic processing.

Conclusion

The current study examined the influence of preceding functional and structural actions on object naming and categorization. Our results show that a preceding congruent functional action can facilitate object naming but not object categorization, while a preceding congruent structural action has no influence on object recognition in either naming or categorization tasks. These findings suggest that functional action representation is involved in fine-grained semantic processing, but neither type of action representation is involved in coarse semantic processing. Our results suggest that functional action representation is a more-distinctive feature and an essential component of man-made tool concepts. Conversely, the structural action representation may not be an intrinsic part of object conceptual knowledge.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-023-02301-6.

Authorship contribution Wenyuan Yu: Conceptualization, methodology, investigation, software, formal analysis, writing – original draft.
Ni Long: Software, formal analysis, writing – original draft
Zijiang Zhang: Investigation, validation, software
Weiqi Zheng: Investigation, validation; writing – original draft
Ye Liu: Conceptualization, resources, writing – review and editing.

Funding This research was supported by the National Natural Science Foundation of China, and the German Research Foundation (NSFC62061136001/DFGTRR-169).

Data availability The materials used in the current study and the preprocessed data generated during the current study are available via the Open Science Framework at https://osf.io/tvns3/.

Declarations

Ethics approval The study adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication No identifying information of any participant is included in this article.

Conflict of interest None.

References

Alam, M. M., Vilankar, K. P., Field, D. J., & Chandler, D. M. (2014). Local masking in natural images: A database and analysis. *Journal of Vision, 14*(8), 22.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Binkofski, F., & Buxbaum, L. J. (2013). Two action systems in the human brain. *Brain and Language, 127*(2), 222–229.

Borghi, A. M., Bonfiglioli, C., Lugli, L., Ricciardelli, P., Rubichi, S., & Nicoletti, R. (2007). Are visual stimuli sufficient to evoke motor information? *Studies with hand primes. Neuroscience Letters, 411*(1), 17–21.

Bub, D. N., & Masson, M. E. J. (2012). On the Dynamics of Action Representations Evoked by Names of Manipulable Objects. *Journal of Experimental Psychology-General, 141*(3), 502–517.

Bub, D. N., Masson, M. E. J., & Cree, G. S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition, 106*(1), 27–58.

Buxbaum, L. J., & Kalénine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Sciences, 1191*(1), 201–218.

Buxbaum, L. J., Kyle, K. M., Tang, K., & Detre, J. A. (2006). Neural substrates of knowledge of hand postures for object grasping and functional object use: Evidence from fMRI. *Brain Research, 1117*, 175–185.

Campanella, F., & Shallice, T. (2011). Manipulability and object recognition: is manipulability a semantic feature? *Experimental Brain Research, 208*(3), 369–383.

Chong, I., & Proctor, R. W. (2020). On the Evolution of a Radical Concept: Affordances According to Gibson and Their Subsequent Use and Development. *Perspectives on Psychological Science, 15*(1), 117–132.

Clarke, A. (2015). Dynamic information processing states revealed through neurocognitive models of object semantics. *Language Cognition and Neuroscience, 30*(4), 409–419.

Clarke, A., Taylor, K. I., Devereux, B., Randall, B., & Tyler, L. K. (2013). From Perception to Conception: How Meaningful Objects Are Processed over Time. *Cerebral Cortex, 23*(1), 187–197.

Clarke, A., Devereux, B. J., Randall, B., & Tyler, L. K. (2015). Predicting the Time Course of Individual Objects with MEG. *Cerebral Cortex, 25*(10), 3602–3612.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception: Classic Edition.* Boston, MA: Houghton Mifflin.

Godard, M., Wamain, Y., & Kalenine, S. (2019). Do manufactured and natural objects evoke similar motor information? The case of action priming. *Quarterly Journal of Experimental Psychology, 72*(12), 2801–2806.

Helbig, H. B., Graf, M., & Kiefer, M. (2006). The role of action representations in visual object recognition. *Experimental Brain Research, 174*(2), 221–228.

Kalenine, S., Bonthoux, F., & Borghi, A. M. (2009). How action and context priming influence categorization: A developmental study. *British Journal of Developmental Psychology, 27*, 717–730.

Kiefer, M., Sim, E.-J., Helbig, H., & Graf, M. (2011). Tracking the time course of action priming on object recognition: Evidence for fast and slow influences of action on perception. *Journal of Cognitive Neuroscience, 23*(8), 1864–1874.

Kumle, L., Vo, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of statistical software, 82*, 1–26.

Lee, C.-L., Middleton, E., Mirman, D., Kalenine, S., & Buxbaum, L. J. (2013). Incidental and context-responsive activation of structure- and function-based action features during object identification. *Journal of Experimental Psychology-Human Perception and Performance, 39*(1), 257–270.

Lee, C.-L., Huang, H.-W., Federmeier, K. D., & Buxbaum, L. J. (2018). Sensory and semantic activations evoked by action attributes of manipulable objects: Evidence from ERPs. *Neuroimage, 167*, 331–341.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Package 'emmeans' (Version 1.1). Retrieved Nov 5, 2021, from https://github.com/rvlenth/emmeans

Martin, A. (2016). GRAPES-Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic Bulletin & Review, 23*(4), 979–990.

Masson, M. E. J., Bub, D. N., & Newton-Taylor, M. (2008). Language-based access to gestural components of conceptual knowledge. *Quarterly Journal of Experimental Psychology, 61*(6), 869–882.

Masson, M. E. J., Bub, D. N., & Lavelle, H. (2013). Dynamic evocation of hand action representations during sentence comprehension. *Journal of Experimental Psychology-General, 142*(3), 742–762.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*(4), 547–559.

Mounoud, P., Duscherer, K., Moy, G., & Perraudin, S. (2007). The influence of action perception on object recognition: A developmental study. *Developmental Science, 10*(6), 836–852.

Ni, L., Liu, Y., & Yu, W. (2019a). The dominant role of functional action representation in object recognition. *Experimental Brain Research, 237*(2), 363–375.

Ni, L., Liu, Y., Yu, W., & Fu, X. (2019b). The China Image Set (CIS): A new set of 551 colored photos with Chinese norms for 12 psycholinguistic variables. *Frontiers in Psychology, 10*, 2631.

Osiurak, F., Rossetti, Y., & Badets, A. (2017). What is an affordance? 40 years later. *Neuroscience and Biobehavioral Reviews, 77*, 403–417.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976–987.

Pulvermuller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences, 17*(9), 458–470.

R Core Team. (2021). R: A language and environment for statistical computing (Version 4.1.2). Vienna, Austria: *R Foundation for Statistical Computing*. Retrieved Nov 5, 2021, from https://www.R-project.org/

Sakreida, K., Effnert, I., Thill, S., Menz, M. M., Jirak, D., Eickhoff, C. R., … Binkofski, F. (2016). Affordance processing in segregated parieto-frontal dorsal stream sub-pathways. *Neuroscience and Biobehavioral Reviews, 69*, 89–112.

Symes, E., Tucker, M., Ellis, R., Vainio, L., & Ottoboni, G. (2008). Grasp preparation improves change detection for congruent objects. *Journal of Experimental Psychology-Human Perception and Performance, 34*(4), 854–871.

Taylor, K. I., Stamatakis, E. A., & Tyler, L. K. (2009). Crossmodal integration of object features: Voxel-based correlations in brain-damaged patients. *Brain, 132*, 671–683.

Taylor, K. I., Devereux, B. J., & Tyler, L. K. (2011). Conceptual structure: Towards an integrated neurocognitive account. *Language and Cognitive Processes, 26*(9), 1368–1401.

Taylor, K. I., Devereux, B. J., Acres, K., Randall, B., & Tyler, L. K. (2012). Contrasting effects of feature-based statistics on the categorisation and basic-level identification of visual objects. *Cognition, 122*(3), 363–374.

Tyler, L. K., Chiu, S., Zhuang, J., Randall, B., Devereux, B. J., Wright, P., … Taylor, K. I. (2013). Objects and categories: Feature statistics and object processing in the ventral stream. *Journal of Cognitive Neuroscience, 25*(10), 1723–1735.

Vainio, L., Symes, E., Ellis, R., Tucker, M., & Ottoboni, G. (2008). On the relations between action planning, object identification, and motor representations of observed actions and objects. *Cognition, 108*(2), 444–465.

Open practices statement None of the experiments was preregistered.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.