Cognition 116 (2010) 394–408

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

On the nature of hand-action representations evoked during written sentence comprehension

Daniel N. Bub *, Michael E.J. Masson **

Department of Psychology, University of Victoria, P.O. Box 3050 STN CSC, Victoria, BC, Canada V8W 3P5

ARTICLE INFO

Article history:
Received 17 August 2009
Revised 3 June 2010
Accepted 3 June 2010

Keywords:
Hand-action representations
Motor resonance
Sentence context

ABSTRACT

We examine the nature of motor representations evoked during comprehension of written sentences describing hand actions. We distinguish between two kinds of hand actions: a functional action, applied when using the object for its intended purpose, and a volumetric action, applied when picking up or holding the object. In Experiment 1, initial activation of both action representations was followed by selection of the functional action, regardless of sentence context. Experiment 2 showed that when the sentence was followed by a picture of the object, clear context-specific effects on evoked action representations were obtained. Experiment 3 established that when a picture of an object was presented alone, the time course of both functional and volumetric actions was the same. These results provide evidence that representations of object-related hand actions are evoked as part of sentence processing. In addition, we discuss the conditions that elicit context-specific evocation of motor representations.

1. Introduction

Sentences that describe physical actions on objects are commonplace in everyday language. Despite the apparent ordinariness of such sentences, a number of provocative questions arise regarding the mental processes that take place during their comprehension. These questions concern the interaction between language and other cognitive systems, in particular the visual and motor systems that deal with objects and the actions they afford. It seems quite reasonable to assume that knowing the full meaning of a word denoting a physical object must in part include a representation of what such an object looks like. The same requirement applies to many actions denoted by verbs. Jackendoff (1987) notes that the difference between verbs such as *walk*, *run*, *jog*, *lope* and *sprint* cannot adequately be captured by a set of binary algebraic features. Instead, to understand the difference, we require a representation of three-dimensional form that includes distinctions in gate and speed as part of the verbs' lexical entries.

The meaning of certain words, then, incorporates details of an object's physical appearance and movement. To what extent, however, do we understand a garden-variety sentence describing an action, such as *John grasped the cell phone*, by directly appealing to our experience of what a cell phone looks like, and to the actions we make when picking it up? A substantial amount of evidence shows that motor or perceptual representations are automatically evoked during or shortly after the processing of sentences describing actions or objects (e.g., Borreggine & Kaschak, 2006; Glenberg & Kaschak, 2002; Kaschak & Borreggine, 2008; Zwaan, Stanfield, & Yaxley, 2002). It is clear from these demonstrations that sentence comprehension does indeed yield motor and/or perceptual "resonance" (Zwaan & Taylor, 2006). Much less clear, though, is how these motor-based representations stand in relationship to the computation of sentence meaning.

In this article, we consider the nature of motor representations evoked by written sentences that describe

* Corresponding author.
** Corresponding author. Tel.: +1 250 721 7536; fax: +1 250 721 8929.
E-mail addresses: dbub@uvic.ca (D.N. Bub), mmasson@uvic.ca (M.E.J. Masson).

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.06.001

simple hand actions with objects (e.g., *Mary picked up the calculator*). We begin by introducing and commenting on two relatively neglected questions that are fundamental to an adequate understanding of motor resonance as part of sentence processing. Objects evoke multiple possible actions, only a subset of which is selected on the basis of intentions or context. The evocation of action representations in sentence comprehension will presumably involve the same or a similar form of competition. A word like *cup*, for example, indicating a manipulable object, evokes multiple possible motor affordances. *To what extent does selection of the contextually-relevant affordance occur in constructing the meaning of a sentence?* Our second question concerns language comprehension as a dynamic process that unfolds in real time. *How does the phenomenon of motor resonance stand in relationship to the dynamics of sentence comprehension?*

These two questions serve as background to motivate a set of experiments on the nature of motor representations evoked by sentences describing hand actions applied to manipulable objects. We seek to better understand the relationship between motor resonance and sentence comprehension. One possibility is that the phenomenon of motor resonance is driven by a literal interpretation of the sentence. In other words, a sentence referring to picking up a calculator entails one kind of action, but a sentence about using that object entails another. The resulting action representations, then, are part and parcel of the literal meaning of the sentence.

There are, however, more complex possibilities. First, motor resonance may enable predictions about changes in real-world states implied by the event structures described in the sentence. This predictive aspect of motor resonance is consistent with the notion of mental simulation favored by Altmann and Mirkovic (2009). On this account, a sentence like *John picked up the pencil* may evoke an anticipation of writing as the final goal of the action rather than more immediate effects produced on the way to that goal (see Searle, 1983, for the distinction between "prior intentions" and "intentions in action", referring to the distinction between an intention that requires a number of steps to completion and an intention directed toward an immediately accessible goal). In that case, motor resonance will reflect functional hand actions at the end of the sentence, even though the literal meaning simply refers to a precision grip for lifting.

A second alternative is that action representations evoked during sentence comprehension may play only a peripheral role in the construction of meaning. Jackendoff (2002) argues that modality-specific representations that include information about the shapes of objects and the nature of the actions applied to them serve to capture details that cannot easily be represented by an abstract conceptual structure of word meaning. For example, the distinction between the sentences *John swaggered into the room* and *John marched into the room* is best captured by referring to our embodied knowledge of different styles of walking. At another level (the interface between meaning and syntax), the sentences are equivalent in that both refer to John entering the room by walking. Motor resonance may reflect fleeting activation of knowledge associated with the details of an action (e.g., the form of walking implied by the verb) that do not contribute to the final representation of the sentence in memory unless the particular action is embellished by subsequent text. Alternatively, resonance effects may occur as a by-product of processing the core aspects of word meaning divorced from sentence context (e.g., a calculator refers to a device for computing numbers and we usually operate it by pressing its keys). Thus, a sentence that refers to picking up a calculator may evoke a functional action representation (pressing keys) because the word *calculator* denotes an object that is used in a particular way regardless of the verb accompanying it.

2. Motor resonance and the selection of contextually-relevant affordances

We interact with objects in a variety of ways, depending on our intentions. Consider a common object like a cell phone. If we wish to pick up the phone and move it to a new location, one set of actions is required; if we intend instead to make a call, another set of movements must occur. Intuition suggests that we program the actions to an object *after* we decide on what to do. The cell phone evokes an inverted closed-grasp once we decide to lift and move it. The same object affords a set of key presses given our intention to telephone a friend.

Despite its commonsense appeal, the view that behavior can be thought of as a sequence of *intention-then-action* routines is not fully consistent with recent neurophysiological evidence. Cisek (2007) argued persuasively, for example, that the motor system uses sensory information from an object to continuously specify a number of potential actions, whereas other procedures determine which action will be selected and executed at any given moment (see also Fagg & Arbib, 1998). For example, Cisek and Kalaska (2002) and Cisek and Kalaska (2005) trained monkeys to reach toward a target location defined by a color cue. If two color cues were presented together, signalling mutually incompatible reaching directions, *both* actions were simultaneously represented by neural signals in the premotor cortex, until a further cue identified one of these alternatives as the target action. At that point, premotor activity associated with the selected direction increased, while activity associated with the alternative direction sharply diminished. Baumann, Fluet, and Scherberger (2009) similarly trained monkeys to produce one of two possible grasp actions to an object, cued by color. If the object was presented before the color cue, both actions were represented by neural activity in premotor cortex. Once the color cue occurred, activity associated with the relevant action increased while neural representation of the alternative grasp faded away.

If behavior invariably demands a resolution of competing possible action representations evoked by an object, consider the word *cell phone* in a sentence such as *In order to clear the shelf, John lifted the cell phone*. The evidence we have described implies that when *cell phone* initially is encountered in the sentence it will evoke a number of possible action representations (driven by past experiences associated with the object), including a representation of both closed-grasp and key-press actions. The meaning of the sentence serves as the context that increases the activation of relevant motor representations (the closed-grasp hand action corresponding to lifting the cell phone) while decreasing the activation of those representations that do not fit the behavior implied by the sentence (poking keys). Motor resonance should correspondingly reflect the dynamics of this selection process; multiple action representations should become active in response to a word denoting a manipulable object, followed by selection of the relevant action representations determined by the meaning of the sentence.

3. Motor resonance and the dynamics of sentence comprehension

Studies looking for motor resonance in comprehension have demonstrated the evocation of motor representations at various time points (early as well as late) in sentence processing. Zwaan and Taylor (2006) required subjects to read a sentence one frame at time, with each frame comprising between one and three words. Subjects made each frame appear by rotating a small knob (1 in. in diameter) mounted on a keyboard, clockwise or counterclockwise (direction varied between the first and second half of the experiment). Critical sentences described manual rotation (e.g., *To quench his thirst the marathon runner eagerly opened the water bottle*), and rotation of the knob either matched or mismatched the implied direction of the action in the sentence (counter-clockwise for opening a water bottle). Motor resonance was demonstrated if reading times were faster on matching than mismatching trials. This effect was indeed obtained, but only in response to the main verb (*open*, in the example); no match effect occurred in the preverb region, the first postverb region, or the sentence-final region. Borreggine and Kaschak (2006) report a similarly short-lived effect of motor resonance in sentence comprehension; evidence for motor resonance was obtained during the course of on-line sentence processing but no effects were observed after the end of the sentence.

Other demonstrations, by contrast, show more persistent motor resonance effects, enduring well beyond the last word of a sentence. For example, Zwaan, Madden, Yaxley, and Aveyard (2004) presented subjects with auditory sentences such as *The shortstop hurled the softball at you*, followed 750 ms later by a task requiring a same-different judgment of two visual objects presented in sequence. On critical trials, pictures of the same object occurred with the first version being slightly larger or smaller than the second image. A bigger-sized object occurring immediately after a smaller version would suggest movement towards the observer while a smaller object following a bigger image of the same object would imply movement away from the observer. In the example sentence, a small-to-big transition of the softball would yield a match with the description of a softball thrown at the observer, but the same sequence would mismatch a sentence like *You hurled the softball at the shortstop*. The prediction – confirmed by Zwaan et al. – was that if motor resonance occurs as part of sentence processing, participants should be faster at judging pictures in the match condition than in the mismatch condition, even though the perceptual task has ostensibly nothing to do with the task of understanding the auditory sentence. Clearly, this evidence indicates that under certain conditions, motor resonance effects can be measured well beyond the end of the sentence. Additional results demonstrating persistent effects of motor resonance after sentence processing have been obtained by Kaschak and Borreggine (2008; also see Zwaan et al., 2002, for perceptual resonance effects induced by a sentence on subsequent picture processing).

How do we reconcile these contradictory results, some experiments showing a brief influence of motor resonance that dissipates quickly over the course of sentence processing while other experiments yield resonance effects that endure well beyond the last word in the sentence? A reasonable conjecture is that motor representations evoked during early stages of sentence comprehension are limited in duration because they do not reflect the full meaning of the sentence, which evolves gradually as words are processed in sequence. According to one influential account, the initial construction of the meaning of a word can be indifferent to context (see Kintsch, 1998, for a very clear account), although selective access to meaning may occur immediately when context is sufficiently constraining (e.g., Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tabossi, 1988). The word *cup* for example, in the sentence *After ending the match, the champion lifted the cup* will briefly activate both the conventional meaning of *cup* (a ceramic drinking vessel) and the sense in which it is used here (a trophy made of metal). A short time later, the contextually inappropriate meaning of *cup* is no longer active and the contextually relevant word sense has become established. Inferences guided by the actual thematic content of the sentence, however, require still more time to occur. Activation of globally relevant ideas based on a thematic inference (in the present example, that spectators were likely in attendance, that photographs were taken, that the cup was awarded in a ceremony, etc.) cannot be observed until a considerable time (greater than 500 ms) after the end of the sentence (Long, Oppy, & Seely, 1994; Till, Mross, & Kintsch, 1988).

If motor representations are dynamically evoked as part of sentence meaning, they must stand in some plausible relationship to the timing of processing events that take place during comprehension. The theme of the example sentence implies that the cup is probably raised above the head with both arms, not lifted towards the mouth with one hand (we do not expect the champion to drink wine from the cup). But according to the evidence, the details of the action required by the main verb (*lift*) will become available only late in sentence processing. Motor representations evoked before that point in time are unlikely to reflect the full thematic content of the sentence (see Borreggine and Kaschak (2006), for a very similar proposal).

In support of this argument, we note that motor resonance reported by Zwaan and Taylor (2006), evoked in response to the verb of a sentence, was not only short-lived but also rested on a rather superficial correspondence between the implied action and the motor task carried out by subjects. To view successive segments of a written sentence, subjects rotated a small horizontally mounted knob, an action that required a precision grip using the forefinger and thumb. Although all of the sentences implied rotation that matched or mismatched the direction of the torque applied in the motor task (clockwise or counter-clockwise), there was often little overlap between the details of the action conveyed by the verb and the movement involved in turning the knob. Sentences included actions that required a power rather than a precision grasp, (e.g., *Craving a juicy pickle he took the jar off the shelf and opened the jar*), and other examples in which the turning action was applied to a steering wheel rather than a knob or the lid of an object (e.g., *While driving to work he approached the intersection and turned left onto the street*).

We infer that motor resonance effects obtained by Zwaan and Taylor (2006) reflect only the evocation of rather generic action features like direction of motion. A similar constraint applies to the action-compatibility effect demonstrated by Glenberg and Kaschak (2002; Borreggine & Kaschak, 2006; Kaschak & Borreggine, 2008). The possibility that more specific action representations are evoked during sentence comprehension is raised by studies showing that a single word can exert contextual effects based on the type of action shared with a target (Bub, Masson, & Cree, 2008; Myung, Blumstein, & Sedivy, 2006; Myung et al., 2010; Yee, Drucker, & Thompson-Schill, 2010). In the next section, we describe a methodology that allows us to measure the evocation of particular hand-action representations conforming to actions implied by a sentence context.

4. Measurement of action representations

We have introduced two questions that serve to motivate the present enquiry into the nature of motor resonance during sentence comprehension. One question we raised concerns the dynamics of sentence processing. We have argued that if motor resonance reflects the processing of meaning in real time, then the nature of the motor representations evoked after a verb-noun combination in a simple declarative sentence is of particular interest. For example, the word *lift* in isolation is consistent with a large number of possible actions depending on the kind of object that is being lifted, but the action referred to by the sentence *John lifted the thumbtack* usually implies a precision grip with the forefinger and thumb held in opposition.

A further question has to do with the selection of a single motor representation from the competing affordances dictated by our motor experience of an object. We pick up a thumbtack using a precision grip, but we use it by pressing the ball of the thumb against the top. We argued, given recent neurophysiological work, that the motor representations evoked during sentence comprehension must emerge dynamically. Multiple action representations may initially be elicited by a word denoting a manipulable object, followed by selection of the contextually-relevant affordance as the meaning of the sentence works to select from competing alternatives.

In this article, we employ a novel procedure that allowed us to measure the dynamic representation of hand actions activated by written sentences describing manual interactions with objects. We constructed a multi-element response apparatus consisting of abstract shapes associated through training with eight distinct hand actions (see Fig. 1). Subjects learned to produce each action when cued by a photograph depicting the relevant hand posture. Speeded responses were made to the cue by lifting the preferred hand from a button box and manually carrying out the action on the correct element of the apparatus.

Immediately before carrying out a hand action, subjects read aloud a sentence referring to a manual interaction with an object. The sentence context always described unambiguously either a functional (using) or a volumetric (lifting) interaction with the object. The objects referred to in the sentences were chosen so that they were associated with distinct functional and volumetric hand actions (e.g., calculator, thumbtack; see Fig. 2). Thus, the sentence context served to select among a number of action representations potentially evoked by the object.

We assume that mental representation of actions triggered by the meaning of the sentence will prime the execution of similar actions on the response apparatus, even though the parameters of the cued action may not precisely match those evoked by the object referred to in the sentence. For example, we assume that calculator will prime a volumetric action consisting of an open grasp with the palm facing down and that carrying out this action on the designated response element (in this case, a horizontal cylinder affording a power grasp) shares parameters with actions commonly used to pick up calculators. In support of this assumption, many cells in primate (and, presumably, human) anterior intraparietal (AIP) cortex code generic aspects of hand shape (e.g., pinch) without specifying full details of the object that is being grasped (Fagg & Arbib, 1998; Sakata, Taira, Kusunoki, Murata, & Tanaka, 1997). Thus, the parametric specification for the representation for grasping a small cylinder will overlap substantially with that for grasping a small block. Grasping an element of the apparatus will be primed to the extent that there is sufficient overlap between this grasp and the representation of the grasp evoked by the object concept referred to in the sentence. Motor tasks generally show influences of semantic knowledge on early stages of preparation and response execution (Glover, Rosenbaum, Graham, & Dixon, 2004; Lindemann, Stenneken, Van Schie, & Bekkering, 2006), whereas later stages of movement are affected more by the actual physical parameters of the object being grasped (Goodale, Milner, Jakobson, & Carey, 1991). Response latency was therefore measured from the onset of the hand cue to lift off from the button box.

We can obtain evidence for the activation and selection of specific motor representations by measuring the latency with which a cued hand action is initiated at varying delays after reading the sentence. We varied the relationship between the cued action and the motor representations that we hypothesize are evoked by the sentence. In the unrelated condition, the cued action bore no relationship to the object. In the related condition, the features of the cued actions were similar to either the functional or volu-

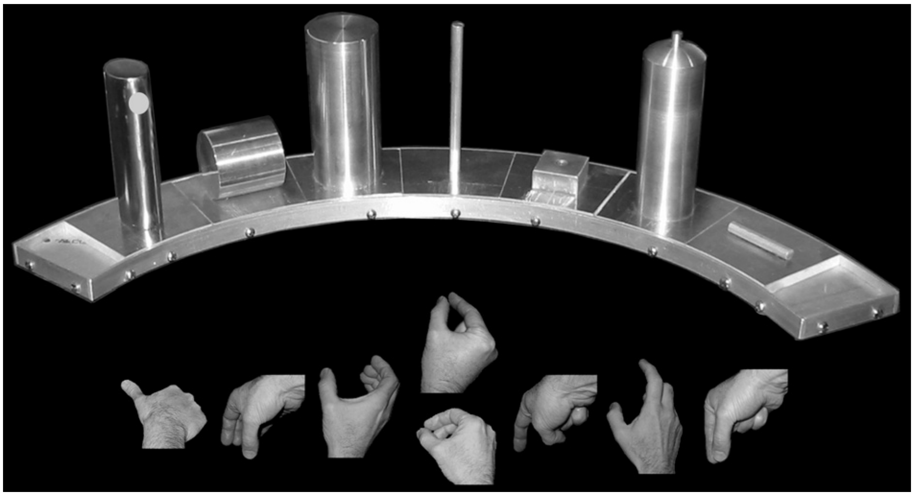

Fig. 1. The response apparatus used by subjects to make hand actions in the reported experiments. The apparatus is configured with seven elements that accommodated the eight different actions that were used. The actions are shown in the bottom of the figure, ordered so as to match the position of the corresponding response element used for that action. The thin, vertical rod in the middle of the element array shown here was used for two different actions.

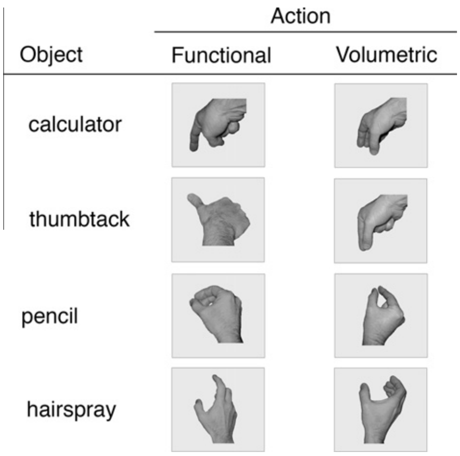

Fig. 2. The four critical object names and their typical functional and volumetric actions used in the experiments. The objects are calculator, thumbtack, pencil, and hairspray (aerosol can).

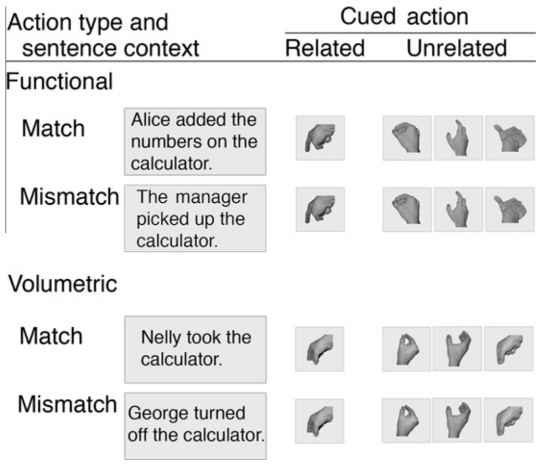

Fig. 3. Examples of context sentences and actions used in each condition of the design of Experiment 1.

metric action associated with the object noun referred to by the sentence. The sentence context as a whole modulated the relationship between the cued action and the object. On match trials, the context implied a functional or volumetric action consistent with the relationship between the cued action and the object. On mismatch trials, the context referred to the complementary action. Thus, for instance, when the cued action was related to the functional action associated with the object, the sentence implied a volumetric action. The reverse was true when the cued action was related to the volumetric action (see Fig. 3 for examples).

We have argued that words referring to manipulable objects evoke multiple action representations, a subset of which are selected by context. If sentence context serves to select a functional or volumetric action and this selection process takes time, we should initially see both action representations evoked by the noun. Thus, after a short delay following sentence reading, *John gave the calculator to Mary* should prime both functional and volumetric actions (poke and inverted open grasp), relative to unrelated actions (e.g., thumb press or pinch). At later stages of processing, one type of action representation should be selected by context. If the motor resonance evoked by the sentence ultimately reflects a literal interpretation of the action conveyed by the sentence then at later stages of comprehension *John gave the calculator to Mary* should prime only the volumetric action representation, just as *Quinton added the numbers on the calculator* should prime only the functional action (see Fig. 4).

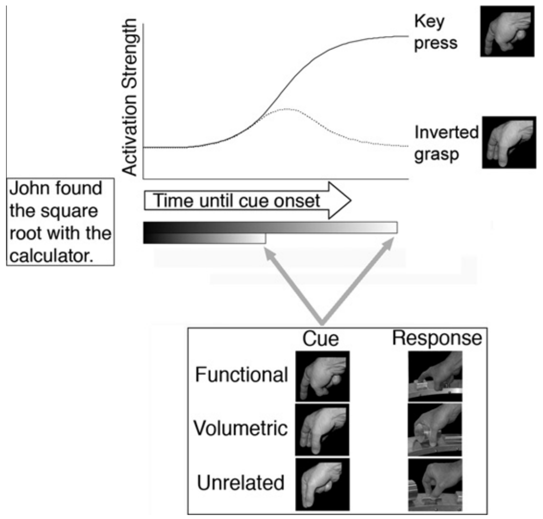

Fig. 4. Hypothetical activation functions for the functional and the volumetric action associated with a calculator. Activation of both action representations is assumed to occur immediately after reading a sentence. Sentence context is assumed to select and enhance the activation of the relevant action whereas activation of the alternative action diminishes. Cuing a hand action that resembles either one of these representations at varying points after object onset provides a measure of the relative activation strength of action representations. Cuing an unrelated hand action provides a baseline for measuring activation. Action responses are illustrated by a hand interacting with an element of the response apparatus.

5. Experiment 1

In Experiment 1 we wished to examine the effect of two kinds of sentence context on the evocation of hand-action representations. One type of context emphasized manual interactions based on the shape of the object by describing activities such as picking up, moving, or giving an object to another person. On the assumption that a literal interpretation of these contexts determines motor resonance, such sentences will ultimately evoke volumetric action representations of object concepts. The other type of sentence context described interactions based on the conventional use of objects. These contexts should elicit hand actions that correspond to functional representations. Subjects read each sentence aloud prior to presentation of the cue that prompted a specific hand action on the response apparatus. The task of reading aloud imposes modest requirements for comprehension, given that subjects were not tested for understanding after reading. We assume that the evocation of hand-action representations will vary considerably given different task demands, ranging from the simple requirement of articulating the sentence to more enactive situations in which subjects are encouraged to imagine carrying out the action implied by a sentence. In the experiments we report here, we wish to examine the nature of hand-action representations evoked by a written sentence when subjects are not required to engage in elaborative processing. This approach allows us to establish a lower bound on the evocation of hand-action representations in language comprehension.

A further objective of Experiment 1 was to examine the recruitment of multiple affordances and selection among these over time. Accordingly, we varied the duration after the sentence before presenting the hand action cue. At a short duration, the possibility exists that several affordances are active regardless of whether the sentence context implied a volumetric or a functional action. At a longer duration, contextual influences would be expected to select the relevant action whereas other, irrelevant actions are suppressed or otherwise dissipate.

5.1. Method

5.1.1. Subjects

Twenty-nine introductory psychology students at the University of Victoria participated in the experiment for extra credit in their course.

5.1.2. Materials

Four different objects, each associated with a unique typical functional and volumetric grasp, were selected. Generic versions of these grasps were used as cues to indicate which response was to be produced by the subject. A grayscale photograph was produced for each hand cue, as shown in Fig. 2. The hand cues subtended maximum horizontal and vertical visual angles of 10.9° and 13.3°, respectively. Two variants of each hand cue photograph were produced, one for right-handed subjects and another (a mirror inversion of the first) for left-handed subjects.

A set of 288 critical and 12 practice sentences were constructed. Among the critical sentences, 36 described a functional interaction (e.g., *David wrote with the pencil*) and 36 described a volumetric interaction (e.g., *Bert picked up the pencil*) with each of the four critical objects. Sentences were displayed in courier font. A four-letter word subtended a horizontal visual angle of 1.1°.

A response apparatus was constructed so that a variety of hand actions could be made by grasping a target element of the apparatus. The apparatus consisted of a curved platform and a set of aluminum response elements, each mounted on its own base (see Fig. 1). The platform spanned 55 cm from left-most to right-most edge. The elements could be arrayed in any order on the platform. Seven elements were used for the experiments reported here. Each of the elements was used for one of the eight critical actions, except for the thin vertical rod, which was used for two actions, as shown in Fig. 1. The four tallest elements were approximately 14 cm in height and the thick vertical cylinder was 6 cm in diameter.

5.1.3. Design

Four factors were varied within subjects and an equal number of critical trials was presented at each level of each factor: action type (functional or volumetric), sentence context (match or mismatch), cued action (the cued action was related or unrelated to the object mentioned in the context sentence), and duration of the delay between reading the sentence and the appearance of the action cue (300 or 750 ms). Examples depicting the first three of these factors are shown in Fig. 3. The comparison between related and unrelated action conditions was the basis for our measurement of priming effects. We made the related–unrelated comparison within action type, anticipating the possibility that functional and volumetric actions might differ in mean latency. If that were the case, then comparing, say, functional actions in the related condition to volumetric actions in the unrelated condition would contaminate estimates of priming with the influence of differences between action types. Assignment of critical sentences to conditions was randomized across subjects. The left-to-right order of the elements in the response apparatus was also varied across subjects.

5.1.4. Procedure

Subjects were tested individually in a quiet room in the presence of an experimenter. Stimulus display was controlled by a Macintosh G3 computer equipped with two monitors, one viewed by the subject and the other viewed by the experimenter. The response apparatus was positioned in front of the base of the subject's monitor and did not obscure the view of the monitor. A button box, placed in front of the response apparatus, consisted of a horizontal array of six buttons. Subjects were first given a training phase in which they learned to make a specific action using the response apparatus in response to each hand cue. Each trial began with the subject placing his or her index fingers on the left- and right-most buttons of the button box. When a hand cue appeared, the subject lifted the preferred hand and grasped the appropriate element of the response apparatus. Subjects were instructed to respond as quickly and as accurately as possible. They were trained not to lift the response hand until they knew what action they were about to make so that their movement from liftoff to contact with the response apparatus was very brief. Subjects were given 72 training trials of this form (nine per action). They were then shown and asked to name the four critical objects to ensure that they were familiar with the objects mentioned in the sentences.

In the final phase of the experiment, subjects were shown 12 practice followed by 288 critical trials that consisted of the following sequence of events. Each trial began with the subject resting his or her index fingers on two buttons of the button box as in the training phase. A context sentence then appeared on the monitor and the subject read the sentence aloud. The sentence was also displayed on the experimenter's monitor. As soon as the last word was articulated, the experimenter pressed a key which caused the sentence to be erased and a blank screen interval was then presented for either 300 ms or 750 ms. This interval was then immediately followed by a hand cue that prompted the subject to make one of the eight possible actions. Although the timing of the interval between the completion of sentence reading and onset of the hand cue included a variable delay that we did not measure (the time taken by the experimenter to signal the end of reading), we note that this delay would be short (in the same range as reaction time to the onset of a simple stimulus). By viewing the sentence as the subject read aloud, the experimenter could prepare in advance for the moment when the final word of the sentence would be articulated. This preparation for the completion of a particular word ensured that the experimenter was able to initiate rapid onset of the delay interval following sentence reading. Our primary goal was to compare the evocation of hand-action representations at a relatively short and longer interval after sentence reading. Although there is imprecision in the timing of the presentation of the hand cue, this limitation does not seriously compromise the general inferences about the dynamic events that we make from our results.

Subjects were instructed to respond to the hand cue as quickly and as accurately as possible by lifting the preferred hand from the button box and grasping the appropriate element of the response apparatus using the action indicated in the cue. Raising a hand from the button box triggered the computer clock which was used to measure response time from onset of the hand cue to release of the button. Once the subject's hand made contact with the response apparatus, the experimenter pressed a key to classify the response as correct or incorrect. The next trial began automatically when the subject returned the response hand to the button box. Order of trials within the practice and critical sets was randomized for each subject. Subjects were given occasional breaks over the course of the critical trials.

5.2. Results

Data from three subjects were excluded because of 10% or more errors in at least one condition. For the remaining 26 subjects, four observations were classified as false starts (latency less than 200 ms) and 0.4% were excluded as outliers due to a response latency greater than 2200 ms. This upper cutoff was set such that no more than 0.5% of the correct response latencies were omitted (Ulrich & Miller, 1994). Mean correct response latency was computed for each condition for each subject and the condition means are shown in Table 1.

An analysis of variance (ANOVA) of the response latency data with action type, sentence context, cued action, and cue delay as repeated measures factors was computed with type I error rate set at .05. This analysis revealed three significant effects. First, there was a main effect of action type, $F(1, 25) = 12.36$, $MSE = 5890$, $p < .005$, indicating that functional actions were initiated more quickly than volumetric actions (647 ms versus 673 ms). There was also a main effect of cued action, $F(1, 25) = 11.81$, $MSE = 1834$, $p < .005$, with shorter latencies when the action was related to the

Table 1
Mean response latencies in Experiment 1.

Sentence context and cue delay (ms)	Action set and cued action			
	Functional		Volumetric	
	Related	Unrelated	Related	Unrelated
Match				
300	644	659	666	676
750	620	658	672	674
Mismatch				
300	647	654	666	685
750	635	657	672	674

Average standard error of the mean = 24.3 ms.

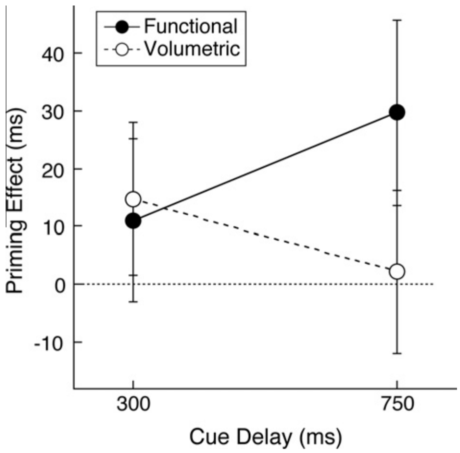

Fig. 5. Mean priming effect in Experiment 1 as a function of action type and cue delay. The means are averaged across the two types of sentence context. Error bars represent 95% confidence intervals suitable for determining whether a priming effect is reliably different from zero.

object indicated in the context sentence (653 ms versus 667 ms). Finally, there was a three-way interaction between action type, cued action, and cue delay, $F(1, 25) = 8.03$, $MSE = 787$, $p < .01$. To examine this interaction, we computed priming effects for each condition by subtracting latencies in the related action condition from latencies in the unrelated action condition. Mean priming effects are shown in Fig. 5 averaging across sentence context as this factor did not contribute to the three-way interaction. At a cue delay of 300 ms, there was a reliable 13-ms priming effect, averaging across both action types, $F(1, 25) = 5.64$, $MSE = 771$, $p < .05$. At the long delay, however, the priming effect for functional actions significantly increased relative to the effect seen at the short delay, $F(1, 25) = 9.01$, $MSE = 1575$, $p < .01$, whereas for volumetric actions, there was no significant change in priming across cue delay, $F(1, 25) = 1.77$, $MSE = 1164$, $p > .10$. The small change in priming for volumetric actions across cue delay was actually a decrease, resulting in no significant priming for volumetric actions at the long delay ($F < 1$). This difference in priming effects across the cue delays for functional and volumetric actions accounts for the three-way interaction.

Error rates were very low (0.4% of all trials), so no analysis is reported.

5.3. Discussion

In Experiment 1, cued hand actions were faster when they matched the affordance of the object referred to in the sentence. Thus, sentences that describe generic hand actions applied to manipulable objects evoke motor representations that prime a related grasp response, consistent with much published evidence on motor resonance. But what is the nature of the relationship between motor resonance and the actual meaning of the sentence? Previous studies have shown that a sentence like *John turned down the radio* exerts an effect on a motor response incorporating similar directional features as those implied by the sentence (e.g., Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006). We do not know, however, whether motor resonance captures the details of the action, or reflects only a more general correspondence between some feature of the motor task defined by instruction (e.g., the direction of the movement needed to produce a sentence fragment) and the spatial representations constructed as part of sentence meaning.

Experiment 1 adds to this literature by examining in more detail the kind of motor representations evoked by sentences describing volumetric or functional hand actions on manipulable objects. The evidence indicates that shortly after the end of a sentence both of these representations are evoked, regardless of whether the meaning implies using or lifting the object. This result is strikingly consistent with recent neurophysiological data from single-cell recording work in primates indicating that target objects initially evoke multiple potential action representations that compete with each other until context biases the competition in favor of a single outcome (Cisek, 2007).

A possibility raised earlier is that motor resonance during sentence comprehension captures the literal meaning of the sentence. On this view, the competing affordances initially evoked by an object should resolve into a contextually relevant action representation. The meaning of a sentence like *John picked up the calculator* should induce a representation of the volumetric action applied to the object while *John used the calculator* would entail a functional action. The results of Experiment 1, however, provide no support for this idea. Motor affordances evoked shortly after the reading of a sentence, resolve over time to favor a functional hand-action representation regardless of whether context favors a volumetric or functional interpretation. This outcome confirms previous results from our laboratory: sentences describing mental acts in regard to objects (e.g., thinking about, remembering) prime functional but not volumetric actions (Masson, Bub, & Newton-Taylor, 2008; Masson, Bub, & Warren, 2008).

These results imply that the relationship between motor resonance and sentence meaning is not simply determined by a literal interpretation of the action carried out on the designated object. We considered two other possibilities, both of which are compatible with the outcome of Experiment 1. Hand-action representations evoked after the end of a sentence may reflect an anticipated goal that typically involves the use of an object even in sentences that refer to moving rather than using. Alternatively, given that the function of an object is considered to be a more dominant component of its meaning than is the shape of the object (Greif, Kemler Nelson, Keil, & Gutierrez, 2006; Keil, 1986; Kemler Nelson, Chan Egan, & Holt, 2004), the consistent evocation of functional hand-action representations emerging at the end of a sentence, even when the context describes a volumetric action, may reflect a generic interpretation of the noun rather than a literal interpretation of the action carried out on the object.

We find no evidence, then, to support claims like the following: "... the neural activation observed during action word processing most likely codes for action-specific activation that matches the semantic context of text with some degree of precision ..." (Taylor & Zwaan, 2009; p. 48). Though perhaps disappointing, we should point out that our results are in fact remarkably compatible with recent single-cell recording work in the monkey. We previously referred to a study by Baumann et al. (2009) who trained monkeys to use either a power or precision grip in a delayed grasping task on a handle presented at various orientations. Grip type was cued by means of a colored light (green for a power grip, white for a precision grip). In relation to Experiment 1 of the present article, we can think of the color cue as the equivalent of a sentence that conveys the use of a particular hand action. A cue (or in our case, a sentence) denoting a specific grasp type should trigger a representation of the corresponding hand action in AIP, according to the widely held intuition that motor planning involves a mental simulation of an intended action (e.g., Gallese & Lakoff, 2005). This preconception, however, does not gain support from the neurophysiological data. The activity of cells in the anterior intraparietal cortex (AIP) – a region strongly involved in the sensorimotor transformation of grasping movements – was recorded after the onset of the color cue but before the handle was presented. When the color cue was displayed in the absence of the handle as object, grasp type was not well represented in the AIP, despite the fact that the cue unambiguously indicated a power or precision grip. As soon as the object was displayed, however, activation of cells in the AIP reflected the correctly cued grip. The AIP encodes context-specific hand grasping movements to perceived objects but not to an abstract visual cue, although neural activity in the parietal lobe can also represent abstract behavioral rules independently of the final motor goal (Gail & Andersen, 2006). Our results are entirely consistent with this evidence.

The hand-action representations revealed in Experiment 1 emerged rather late, given that the strongest form in which we saw them occurred at least 750 ms after the sentence had been read. The lateness of this effect stands in contrast to the more rapid activation of word meaning in sentence contexts (e.g., Seidenberg et al., 1982; Till et al., 1988). Earlier effects of sentence context on hand-action representations were seen in the Masson, Bub, and Warren (2008) study, where priming effects were observed when the cue to respond was presented either in the middle or at the end of the auditory presentation of an object name. The speed with which action representations are evoked will likely vary depending on the depth of processing allocated to the sentence. We note, however, that despite the very modest task requirements we imposed, motor representations were potent some time after the end of the sentence, and showed a pattern of initial activation of both functional and volumetric action representations that fits pleasingly with neurophysiological evidence and theory. The fact that context specificity was not observed even at later stages merits additional consideration. We further examine this interesting outcome in the next experiment.

6. Experiment 2

Written sentences that describe hand actions involving manipulable objects ultimately refer to physical events taking place in a world external to the reader. What kind of relationship exists between the mental world, in which the senses of individual words are combined, and the world where the actions referred to are actually taking place? The question prompts us to look beyond tasks that require only the processing of written sentences, to conditions that would require a linkage between the meaning of words and an externally displayed physical object. We wish to determine whether a combination of language and perceptual events will evoke motor representations that capture the meaning implied by a particular action (e.g., *lift*) applied to a specific object (e.g., *calculator*) in a sentence.

As a first step, we draw upon an important principle originally formulated by Simon (1975). Complex mental processing (in particular, complex problem solving) occurring in the context of an external visual display ("display-based problem solving") capitalizes on the fact that many elements of the data structure are visible attributes of external objects. Accordingly, these attributes are not maintained in working memory but are directly observed as they occur in the world. Other necessary processing elements may not be directly visible but are associated through experience with an object which therefore serves as a cue for attribute retrieval.

Because display-based problem solving incurs little load on working memory, the process seems introspectively easy. For example, Larkin (1989) points out that preparing a cup of coffee requires little mental effort, is resistant to interruption (we simply continue where we left off), can be performed in many different orders and is easy to modify. The reason for this efficacy is that various external objects (working as a "display") can be used to evoke or cue information about the actions they require when performing the task.

We argue that the principles underlying display-based problem solving, applicable to quotidian tasks like preparing a cup of coffee, are equally applicable to the processing of sentences that describe actions and objects. Consider a sentence like *John carefully added the ground coffee to the percolator*. We assume that, independent of motor resonance effects accruing during comprehension of this sentence, contextually detailed action representations will be evoked directly if the sentence is accompanied by a visual display of the referenced objects. Perceptual objects, in other words, directly cue motor affordances that are not explicitly activated in working memory by the sentence context alone.

In this regard, our claim tacitly invokes a distinction widely maintained in linguistic theory, between the sense of a word and its referent. The sense of a word is determined by relating its meaning to the meaning of other words. A sentence about a manipulable object also refers to external perceptual events. This is the referential aspect of a sentence and it depends on representations accessed for recognizing perceptual instances of actions and objects (Miller & Johnson-Laird, 1976). To propose that context-specific motor representations are evoked by a display of the referenced object in combination with a sentence (but not by the sentence on its own) is tantamount to asserting that referential aspects of meaning are triggered when a sentence is combined with an external object, and not necessarily when the meaning of a sentence is considered in isolation (see Ohlsson (1999), for a very similar argument).

We can again point to evidence from single-cell recording in the monkey that appears consistent with the above suggestion. This evidence does not entail language, to be sure, but the use of an abstract sensory cue to signal an action. The nature of the motor representations evoked by such a symbolic cue, however, is of considerable interest given the analogous question on the nature of motor resonance associated with a sentence. Cisek (2005) argues convincingly from neural data that during an instructed delay period (such as occurs after a color cue signalling a particular response before the appearance of the target object), stages of action planning do not capture much information on the details of the upcoming movement. Rather, during preparation or rehearsal of an action the premotor cortex represents abstract elements of the action "... such as the specification of possible targets and the selection of one as the correct target for the action ..." (p. 18). Furthermore, we have already referred to a study by Baumann et al. (2009) showing that a color cue instructing the monkey on a grasp action yields only a weak representation of grasp-type activity in the AIP if the abstract cue is presented without the object as the visible target of the action. In contrast, the color cue produced strong contextually relevant activity after the target object becomes visible. The authors conclude that "... AIP encodes context-specific hand grasping movements to perceived objects, but in the absence of a grasp target, the encoding of context information is weak" (p. 6436).

We infer then, given current neurophysiological evidence, that an abstract color cue (and we suggest, also, a sentence of the form tested in the present experiments) does not elicit a detailed motor representation of the impending action. A context-specific action representation should be evoked, however, by a visual object when presented in combination with an abstract cue (or sentence) denoting a particular motor affordance. We suggest, accordingly, that sentences referring to a functional or volumetric interaction with a manipulable object should evoke a contextually relevant hand-action representation if a visual image of the object is displayed shortly after the sentence. Activation of the motor affordance described in the sentence should take place via the same process of competition followed by selection of the relevant action representation that we documented in Experiment 1. During early stages of processing, the visual object should evoke multiple affordances, until the contextually-relevant affordance referred to in the sentence gains in strength and activation of the competing set of affordances diminishes. We test this conjecture in the next experiment.

6.1. Method

6.1.1. Subjects

Twenty-four subjects from the same population as in Experiment 1 were tested.

6.1.2. Materials and procedure

The same materials and procedure were used as in Experiment 1, except that during the 300-ms or 750-ms delay between reading a sentence context and the appearance of the hand action cue, a grey scale image of the referenced object was displayed.

6.2. Results

Data from four subjects were excluded from analysis because they made 10% or more errors in at least one condition of the design. For the remaining subjects, response latencies less than 200 ms were excluded from analysis and treated as false starts. A total of five trials were eliminated for this reason. Latencies over 1800 ms were excluded as outliers (0.4%). The mean correct response latency was computed for each subject, and the condition means taken across subjects are shown in Table 2.

Response latencies were analyzed in a four-factor analysis of variance (ANOVA) with action set (functional, volumetric), sentence context (match, mistmatch), cued action (related, unrelated), and cue delay (300 ms, 750 ms) as repeated measures factors. Two significant effects were found. There was a main effect of cued action, $F(1, 19) = 22.23$, $MSE = 1141$, $p < .001$, constituting a priming effect whereby responses were faster when the action was related to the object. The other significant effect was a three-way interaction involving sentence context, cued action, and cue delay, $F(1, 19) = 4.46$, $MSE = 816$, $p < .05$. To help interpret this interaction, priming effects were computed for each subject by subtracting response latency in the related condition from response latency in the unrelated condition, within each condition defined by the combination of action set, sentence context, and cue delay. The mean priming effect for each of these conditions is shown in Fig. 6. It is apparent from this figure that a small priming effect is present for both functional and volumetric actions at the short cue delay, regardless of the type of sentence context. Averaging across the two action sets and the two contexts, the priming effect of 16 ms is significant, $F(1, 19) = 9.90$, $MSE = 251$, $p < .01$. At the long cue delay, again averaging over the two action sets, a priming effect of 34 ms is apparent only with matching sentence contexts, $F(1, 19) = 24.10$, $MSE = 476$, $p < .001$, and this effect was larger than the overall priming effect seen at the short delay, $F(1, 19) = 5.61$, $MSE = 584$, $p < .05$. There was no significant priming effect at the long delay for the mismatch context condition ($F < 1$).

Subjects made very few errors (a total of 20 errors or 0.3% of all trials), so no analysis of error rates is reported.

Table 2
Mean response latencies in Experiment 2.

Sentence context and cue delay (ms)	Action set and cued action			
	Functional		Volumetric	
	Related	Unrelated	Related	Unrelated
Match				
300	654	673	659	672
750	637	667	640	678
Mismatch				
300	657	669	667	685
750	648	653	664	670

Average standard error of the mean = 22.6 ms.

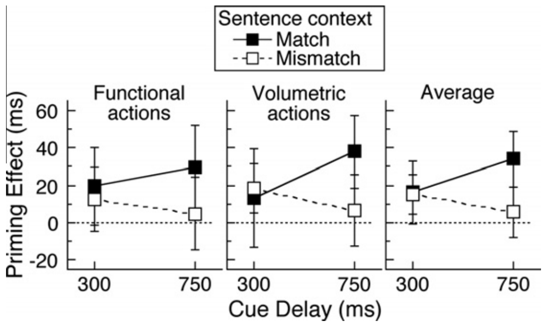

Fig. 6. Mean priming effect in Experiment 2 as a function of sentence context, action set, and cue delay. The right panel shows the priming effect averaged across the two types of action. Error bars represent 95% confidence intervals suitable for determining whether a priming effect is reliably different from zero.

6.3. Discussion

As in Experiment 1, both functional and volumetric hand-action representations are evoked when a picture of the referenced object is presented for 300 ms after the end of the sentence. Our results provide consistent behavioral evidence in support of the idea that an action representation emerges dynamically over time as competition is resolved between multiple affordances (Cisek, 2007). In Experiment 1 we observed that content-specific effects do not occur to written sentences denoting functional or volumetric ways of interacting with a manipulable object. Instead, multiple affordances that are initially evoked ultimately resolve into a functional hand-action representation, independent of whether the meaning implies a functional or volumetric interaction. By contrast, in Experiment 2 a visual image of the object displayed immediately after the sentence shows a very clear resolution of multiple hand-action representations in favor of a contextually determined outcome. Volumetric hand-action representations are evoked by sentences like *Bert picked up the pencil* when an image of the object is displayed for 750 ms after the end of the sentence, and the activation of the incipient functional hand-action representation is no longer observed. The reverse is true for a sentence describing a functional interaction (*Bert wrote with the pencil*). Motor resonance is therefore closely linked to content-specific hand-action representations if the sentence is directly followed by a visual object.

7. Experiment 3

We turn now to the question of which action representations are evoked by the sight of an object alone, passively viewed in the absence of a sentence context or other intentional cue. We do not know whether activation of action representations during passive viewing of an object is sufficiently precise to enable priming of volumetric and/or functional representations. Furthermore, we need to establish whether an object on its own differentially evokes functional versus volumetric action representations. Conceivably, the results of Experiment 2 may reflect combined asymmetrical influences of sentence context and object viewing. For example, the picture on its own may evoke volumetric representations that are blocked in a functional sentence context, but not in a volumetric context. The outcome would emulate the context-specific effects seen in Experiment 2, but simply because sentence and picture contribute differentially to the evocation of functional and volumetric action representations. In Experiment 3, then, we rule out the possibility that passive viewing of an object is biased in favor of one type of action representation over another. Subjects viewed the photograph of an object for either 300 or 750 ms, and then were presented with a hand cue that was either related or unrelated to the object.

7.1. Method

7.1.1. Subjects

Forty subjects from the same population as the earlier experiments were tested.

7.1.2. Materials and procedure

The same objects and actions were used as in Experiment 2, but no sentences were presented. Subjects were trained to make cued hand actions using the elements of the response apparatus, just as in the first two experiments. The training phase was followed by 12 practice and 288 critical trials. Each of these trials began with a 300-ms or 750-ms display of one of the four critical objects, followed by a action cue. No response to the object was required. Equal numbers of critical trials were tested at each level of the independent variables: action set (functional, volumetric), cued action (related, unrelated), and object duration (300 ms, 750 ms).

7.2. Results

Data from four subjects were excluded from the analyses because of error rates of 10% or higher in at least one condition, leaving a sample of 36 subjects. Across these subjects, 19 observations were excluded because response latency was less than 200 ms (false starts) and 0.4% were excluded as outliers (latency greater than 2400 ms). Mean correct response latency for each condition is shown in Table 3. An ANOVA of these data with action set, cued action, and object duration revealed a main effect of object duration, with shorter latencies in the long duration condition, $F(1, 35) = 62.98$, $MSE = 2964$, $p < .001$. More important, there was also a significant interaction between cued action and object duration, $F(1, 35) = 7.81$, $MSE = 351$, $p < .01$. The interaction was examined by testing for a priming effect at each of the object durations. With the 300-ms duration, there was no priming effect ($F < 1$), but when the object duration was 750 ms, a small (9 ms) but significant priming effect was found, $F(1, 35) = 7.47$, $MSE = 193$, $p < .01$. This priming effect was very similar for both action sets.

The mean error rate was 0.9% and most subjects made at least one error. An ANOVA of the error rates found only one significant effect: a lower error rate for related (0.6%)

Table 3
Mean response latencies in Experiment 3.

Object prime duration (ms)	Action set and cued action			
	Functional		Volumetric	
	Related	Unrelated	Related	Unrelated
300	645	637	643	644
750	586	595	588	597

Average standard error of the mean = 16.0 ms.

than for unrelated (1.2%) actions, $F(1, 35) = 11.72$, $MSE = 1.81$, $p < .005$.

7.3. Discussion

Passive viewing of objects yields weak and relatively late-occurring activation of action representations corresponding to both volumetric and functional grasps. The combined effect of a sentence and a visual object seen in Experiment 2, then, is due to the integration of sentence meaning with object recognition. This integration evokes hand-action representations in response to the picture and that are consistent with the interpretation of the sentence. This conclusion fits well with neurophysiological evidence. Action representations in the AIP evoked by an object are specific to the grasp type indicated by a color cue. These action representations emerge only once the target object appears and can be integrated with the cue to clearly specify not only the grasp type but also the grasp orientation (Baumann et al., 2009).

8. General discussion

Motor resonance is an undeniable aspect of sentence comprehension. Reach and grasp responses are primed by sentences that reference actions describing lifting or using manipulable objects. What though, is the relationship between the evocation of hand-action representations and the processing of sentence meaning? We introduced, as background, the idea that a word referring to a manipulable object is associated by past experience with multiple action representations (e.g., we pick up a thumbtack with one type of hand action but use the object with another). According to neurophysiological evidence, multiple affordances compete with each other until the selection of a contextually determined response (Cisek, 2007). Motor resonance should reflect the dynamics of this process of activation-then-selection.

Does sentence context work to select among a set of evoked action representations during comprehension? Note that it is the verb in combination with the noun that denotes the particular action; to *pick up a thumbtack* implies one hand action but to *use a thumbtack* implies another. It is obviously the case that carrying out the action itself requires implementing one particular action representation instead of another. Our interest, however, lies in whether merely understanding a declarative sentence involves the same type of action selectivity.

Using a methodology that allows us to track the evocation of functional and volumetric hand-action representations, we found replicable evidence that multiple affordances are initially activated by a written sentence referring to a manipulable object. These behavioral data are consistent with single-cell recording work indicating the same competition between affordances at a neurophysiological level.

We find no evidence, however, that the motor resonance evoked by a sentence on its own exclusively captures a representation of the hand action on the object implied by the verb. Whether the sentence denotes grasping or using the object, multiple affordances that were initially evoked soon gave way to functional hand-action representations. This bias toward functional actions replicates previous results (Bub et al., 2008; Masson, Bub, & Newton-Taylor, 2008; Masson, Bub, & Warren, 2008). The bias may be driven in two possible ways. First, a sentence describing a volumetric interaction (e.g., *John picked up the calculator*), though not entailing a functional action (i.e., poking keys), may conceivably evoke functional-grasp representations simply because the reader assumes that John intended to use the calculator after picking it up. The alternative possibility is that processing the meaning of a word denoting a manipulable object inevitably elicits a representation of a functional action because an object's function is intrinsically part of its meaning (e.g., Greif et al., 2006).

Experiment 1 shows that reading a sentence about an action does not necessarily evoke the same representations as those involved in carrying out the described action. Our findings, however, can be linked to recent work examining neural activity in the monkey cortex during an instructed delay period, where an abstract cue denotes a particular action before the occurrence of a target object. We suggest that the planning of an action driven by an abstract intentional cue in the monkey invokes the same or very similar motor processes to those that would be required for planning or simulating an action denoted by a sentence. What kind of representation underlies motor preparation in response to an abstract cue (e.g., a color cue in the monkey, a sentence for a human) prior to response execution? The answer from single-cell recording in primates is surprising, and is quite inconsistent with the view that planning a hand action involves a mental simulation of the grasp. Neurons in the AIP respond weakly to an abstract cue denoting a power or precision grasp but react strongly as soon as the grasp target is revealed (Baumann et al., 2009). Gail and Andersen (2006) note that the parietal region can represent abstract rule information that is independent of a spatial cue or motor goal representations. Similarly, Cisek (2005) argues: "The neural representation of a plan does not consist of a desired trajectory or motor program optimized prior to movement onset ... Most of the details of kinematics and kinetics do not emerge until after movement begins, even in very familiar tasks and even when full information about the upcoming movement is provided well ahead of time" (p. 19).

The evidence we have obtained that context-specific motor representations are not available directly after reading a written sentence is consistent, therefore, with substantial neurophysiological data that a detailed motor representation of an intended action does not occur in response to an abstract sensory cue. Single-cell recording in the monkey indicates, though, that context-specific motor representations are indeed evoked if a cue denoting a particular grasp action is coupled with the target object (Baumann et al., 2009). Motivated in part by this recent observation, as well as by two additional theoretical considerations (discussed below), we provided an argument that context-specific motor resonance, although absent after the reading of a sentence, will indeed be observed if the sentence is coupled with a visual image of the designated object.

The first consideration is derived from models of display-based problem solving in which a visual object can be used to directly cue a set of actions required for a task such as preparing a cup of coffee (Larkin, 1989). We argued that context-specific motor representations may likewise be evoked if the visual image of a manipulable object is consulted shortly after a written sentence that describes a functional or volumetric interaction with the object. The image will directly cue a specific motor representation implied by the meaning of the sentence and that becomes part of the conceptual representation held in working memory.

A second consideration further elucidates how central aspects of word meaning can be differentiated from motor and sensory representations (i.e., the traditional distinction between the sense of a word and its reference). Comprehension may be based either on the general sense of the words in combination, or may be supplemented by perceptual and motor representations when conditions so demand, as when one must carry out the described action or when sentence meaning is integrated with a pictured object.

On the basis of these arguments, we further evaluated the possibility that a visual image of the referenced object, presented immediately after a sentence, may provide the perceptual information needed to yield contextual effects on the evocation of action representations. In addition, selection of the contextually determined representation should conform to the pattern we have already observed: initial competition between multiple affordances followed by resolution in favor of a single outcome.

We found striking evidence confirming this conjecture in Experiment 2. An image of a manipulable object presented for 350 ms after the end of a sentence evoked both functional and volumetric action representations, regardless of whether the sentence described lifting or using the object. This outcome replicates Experiment 1, showing that multiple affordances are evoked before sufficient time has elapsed for selection of a unique response. An image presented for the longer duration of 750 ms, however, showed context-specific resolution of the initial competition between multiple action representations. At this exposure duration, an image of a manipulable object displayed immediately after a sentence describing a functional interaction, evoked a hand-action representation conforming to the use of the referenced object. No corresponding activation was found for a volumetric action representation. Conversely, an image of the referenced object displayed after a sentence describing a volumetric interaction, evoked a hand-action representation commensurate with picking up or grasping the object whereas no activation was observed for the functional hand-action representation. It is particularly notable that the hand-action representations evoked by a pictured object are modulated by functional or volumetric sentence context given that the picture on its own yields both kinds of hand-action representations in equal degree, albeit weakly (Experiments 3; see Bub et al. (2008), for stronger but equivalent effects with functional and volumetric actions).

Although sentence context exerts a powerful and specific modulatory effect on the affordances evoked by a picture, it is equally clear that motor resonance does not straightforwardly reflect a language-induced simulation of the actions implied by a sentence. In the absence of a visual referent, hand-action representations activated shortly after sentence processing show no context specificity. What then is the relationship between motor resonance and sentence comprehension?

We noted in the introduction to Experiment 1 that comprehension, at least from operational standpoint, may be assessed in a variety of ways requiring different degrees of elaboration with respect to sentence meaning. It seems very reasonable to assume that the evocation of hand-action representations will vary in informative ways as a function of different task requirements. For example, the instruction to pantomime the action described in a sentence like *John lifted the calculator* presumably will have a different impact on the motor system than the request to read the sentence aloud. In addition, the dynamics of evocation are likely to vary substantially as well. How quickly are different motor representations evoked given particular comprehension demands and how long do they endure? The answer is likely to depend on factors such as richness of the sentence context. For example, Seidenberg et al. (1982) showed that selective access to ambiguous nouns occurred when the sentence context included a word that was strongly associated with the critical ambiguous word (e.g., *The farmer bought the straw*). Our sentences were less strongly constrained and it is reasonable to assume that the effects of multiple affordances we obtained may be modulated by richer contexts. In addition to the important issue of the nature of sentence contexts, a further question for future research concerns the relationship between evoked motor representations and the kind of comprehension task assigned to the subject. For example, what kind of affordances are dynamically evoked by a sentence when the task is to pantomime the action described in it versus a task that simply requires answering a question at the end of the sentence? The methodology we have described can readily be applied to these issues. In the present experiments, we cannot assess the on-line evocation of hand-action representations as sentence meaning unfolds. Our methodology, however, readily opens the possibility of tracking the detailed dynamics of hand-action representations as they unfold in real time in different task and sentence contexts using auditory sentence presentation and visual hand cues.

Resonance effects, as we have noted earlier, may occur at various points in time as the meaning of individual words (or words in combination) unfolds. A verb (e.g., *raise*) considered in isolation may imply nothing about the manner in which the action is carried out, and even in combination with a noun (*raise the newspaper*) may not reveal the details of the action until the full thematic content of the sentence is determined. Consider for example, the difference between (1) *John raised the newspaper and angrily scanned the headlines* versus (2) *John raised the newspaper and angrily swatted the fly*. Our interest in the link between sentence comprehension and motor resonance ultimately concerns the complex (but empirically tractable) issue of how motor representations vary in relation to local and global aspects of sentence comprehension.

Motor resonance effects can in principle reflect a variety of relationships between action representations evoked by the meaning of words and the motor task that subjects are required to carry out during sentence processing. One form of resonance occurs when there is a general correspondence between the spatial direction implied by the meaning of words in the sentence and the directional component of the motor task (Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006). Such priming effects may occur even when other parameters of the motor task show little or no overlap with the details of the action implied by the sentence. Thus, *John raised the newspaper* may prime the feature ⟨*up*⟩ for any response that requires an upward movement, despite the fact that the motor task bears no other resemblance to the action of raising a newspaper.

Our methodology is designed to reveal a more detailed correspondence between the cued action and the action described in a sentence: motor resonance effects include the evocation of a hand shape corresponding to functional or volumetric interactions with a manipulable object. The evidence from the current experiments indicates that the motor representation active at the end of a sentence concerns the use of the object, regardless of whether the sentence entails a functional or volumetric interaction. This result implies, say, that even for a sentence like *John raised the newspaper and angrily swatted the fly*, a representation of the thematic content held in working memory does not include details of the action required to swat a fly with a rolled up newspaper. Instead, motor resonance present at the end of a sentence observed in Experiment 1 reflects functional action representations associated with the meaning of the noun rather than the contextually determined action implied by the sentence. This outcome may appear surprising at first glance, but note there is good evidence that the function of a manipulable object is central to its identity. An object's function rather than its shape is crucial to determining the name, even for very young children (Malt & Johnson, 1992; Smith, Jones, & Landau, 1996; Miller & Johnson-Laird, 1976). In addition, young children classify novel artifacts on the basis of their intended use (e.g., Greif et al., 2006) rather than their appearance. Further, in experiments similar to Experiment 1 reported here, using written or auditory sentences with verb constructions denoting non-physical interactions with manipulable objects (e.g., *think about*, *remember*, *look at*), we found evidence for priming of functional but not volumetric actions when the manipulable object is mentioned at the end of the sentence (Masson, Bub, & Newton-Taylor, 2008; Masson, Bub, & Warren, 2008). Thus, motor resonance offers clues on the nature of the semantic representations available at the end of a sentence; resonance indicates activation of core aspects of word meaning, not the literal simulation of the described action.

We have so far described two kinds of relationship between motor resonance and word meaning: (1) general effects attributable to spatial feature overlap between the described and cued action (e.g., a sentence denoting the direction away from or toward the body will prime a hand movement in the corresponding direction; Glenberg & Kaschak, 2002), and (2) more specific (but context independent) effects that occur at the end of a sentence, attributable to the functional action representation associated with the identity of the referenced object (e.g., our Experiment 1; Masson, Bub, & Newton-Taylor, 2008). We consider now a third possibility, namely, the existence of context-specific motor representations that emerge dynamically when an action verb (e.g., *lift*) is locally combined with a noun (e.g., *spray can*) to further specify a movement that is undetermined by the verb in isolation.

Bailey (1997; also see Feldman & Narayanan, 2004) describes a computational model of action verb learning (in particular, hand actions like *lift*, *grasp*, *clench*, etc.) based on motor representations or schemas linked via intermediate representations to words. The intermediate layer (or linking feature–structure) contains parameter values for aspects of hand actions that include details of posture, direction, acceleration, and so on. Since a verb like *lift*, according to the model, is ambiguous with respect to hand posture (we may use a power or precision grip depending on the object being lifted, and the wrist may be horizontally or vertically oriented), relevant linking features maintain all possible parameter values, unless additional information is available to select from amongst them. In the case of a sentence fragment like *John lifted the spray can in order to* . . ., multiple parameter values for *lift* should quickly resolve in favor of a vertical closed-grasp, just as the sentence *John lifted the thumbtack in order to* . . . should select a horizontal pinch as the hand posture now assigned to the verb. We assume that such context specificity is short-lived and reflects the dynamics of verb processing in combination with a disambiguating noun rather than a more complete representation of the sentence. To see that this is a reasonable conjecture, we need only consider a sentence like *John lifted the spray can in order to flatten the dough*. The final action implied by the full sentence has nothing to do with the action that is locally evoked by the verb-noun combination.

To conclude, we raise an interesting hypothesis based on the arguments we have just presented. A sentence describing a particular hand action (e.g., *John lifted the calculator in order to clear the shelf*; *John used the calculator to add the numbers*) will show context specificity very shortly after the noun is encountered and the action implied by the verb is disambiguated, but this effect will rapidly diminish as the full meaning of the sentence unfolds.

Acknowledgements

This research was supported by discovery grants from the Natural Sciences and Engineering Research Council of Canada to Daniel Bub and to Michael Masson, and by a grant from the J.S. McDonnell Foundation, and National Science Foundation Grant #SBE-0542013 to the Temporal Dynamics of Learning Center, an NSF Science of Learning Center. We are grateful to Marnie Jedynak and Lisa Worth for assistance in conducting the experiments.

References

Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*, 583–609.

Bailey, D. (1997). *When push comes to shove: A computational model of the role of motor control in the acquisition of action verbs*. Ph.D. Dissertation, Computer Science Division, University of California, Berkeley.

Baumann, M. A., Fluet, M.-C., & Scherberger, H. (2009). Context-specific grasp movement representation in the macaque anterior intraparietal area. *Journal of Neuroscience, 29*, 6436–6448.

Borreggine, K. L., & Kaschak, M. P. (2006). The action–sentence compatibility effect: It's all in the timing. *Cognitive Science, 30*, 1097–1112.

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

Cisek, P. (2005). Neural representations of motor plans, desired trajectories, and controlled objects. *Cognitive Processing, 6*, 15–24.

Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society B, 362*, 1585–1599.

Cisek, P., & Kalaska, J. F. (2002). Simultaneous encoding of multiple potential reach directions in dorsal premotor cortex. *Journal of Neurophysiology, 87*, 1149–1154.

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron, 45*, 801–814.

Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal–premotor interactions in primate control of grasping. *Neural Networks, 11*, 1277–1303.

Feldman, J., & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language, 89*, 385–392.

Gail, A., & Andersen, R. A. (2006). Neural dynamics in monkey parietal reach region reflect context-specific sensorimotor transformations. *Journal of Neuroscience, 13*, 9376–9384.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*, 455–479.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558–565.

Glover, S., Rosenbaum, D. A., Graham, J., & Dixon, P. (2004). Grasping the meaning of words. *Experimental Brain Research, 154*, 103–108.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature, 349*, 154–156.

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., & Gutierrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science, 17*, 455–459.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA.: MIT Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Kaschak, M. P., & Borreggine, K. L. (2008). Temporal dynamics of the action–sentence compatibility effect. *Quarterly Journal of Experimental Psychology, 61*, 883–895.

Keil, F. C. (1986). The acquisition of natural kind and artifact terms. In W. Demopoulos & A. Marras (Eds.), *Language learning and concept acquisition* (pp. 133–153). Norwood, NJ: Ablex.

Kemler Nelson, D. G., Chan Egan, L., & Holt, M. B. (2004). When children ask, "What is it?" what do they want to know about artifacts? *Psychological Science, 15*, 384–389.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.

Larkin, J. H. (1989). Display-based problem solving. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing* (pp. 319–432). Hillsale, NJ: Erlbaum.

Lindemann, O., Stenneken, P., Van Schie, H. T., & Bekkering, H. (2006). Semantic activation in action planning. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 633–643.

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1456–1470.

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of Memory and Language, 31*, 195–217.

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, 71*, 869–882.

Masson, M. E. J., Bub, D. N., & Warren, C. M. (2008). Kicking calculators: Contribution of embodied representations to sentence comprehension. *Journal of Memory and Language, 59*, 256–265.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, England: Cambridge University Press.

Myung, J.-Y., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition, 98*, 223–243.

Myung, J.-Y., Blumstein, S. E., Yee, E., Sedivy, J. C., Thompson-Schill, S. L., & Buxbaum, L. J. (2010). Impaired access to manipulation feature in apraxia: Evidence from eyetracking and semantic judgment tasks. *Brain & Language, 112*, 101–112.

Ohlsson, S. (1999). Anchoring language in reality: Observations on reference and representation. *Discourse Processes, 28*, 93–105.

Sakata, H., Taira, M., Kusunoki, M., Murata, A., & Tanaka, Y. (1997). The TINS lecture. The parietal association cortex in depth perception and visual control of hand action. *Trends in Neurosciences, 20*, 350–357.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*, 489–537.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7*, 266–268.

Smith, L., Jones, S., & Landau, B. (1996). Naming in young children: A dumb mechanism? *Cognition, 60*, 143–171.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential context. *Journal of Memory and Language, 27*, 324–340.

Taylor, L. J., & Zwaan, R. A. (2009). Action in cognition: The case of language. *Language and Cognition, 1*, 45–58.

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition, 16*, 283–298.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123*, 34–80.

Yee, E., Drucker, D. M., & Thompson-Schill, S. L. (2010). fMRI-adaptation evidence of overlapping neural representations for objects related in function or manipulation. *NeuroImage, 50*, 753–763.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic mental representations in language comprehension. *Cognitive Science, 28*, 611–619.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*, 168–171.

Zwaan, R. A., & Taylor, L. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135*, 1–11.