Experimental methods for simulation semantics*

Benjamin Bergen

1. Simulation semantics and language understanding

How do people understand language? Though vital to the study of language and the mind, a disproportionately small body of empirical work has historically addressed this question. Research on language understanding presumably falls under the purview of cognitive linguistics, the study of the mind and language, but cognitive linguistics has predominantly produced static, verbal models of linguistic and other conceptual representations, rather than the dynamic models of psychological processing required to explain the processes of language understanding. At the same time, empirical psycholinguistics, whose experimental methods could in principle yield profound insights into the question, has shied away from deep language understanding, preferring to stick to more easily measurable and manipulable aspects of language – principally aspects of linguistic form like syntax and phonology. However, a number of lines of research have recently emerged, which wed psycholinguistic techniques with a cognitive linguistic perspective on language knowledge and use. This work introduces a new, integrated field, which focuses on the idea that language understanding is contingent upon the understander mentally simulating, or imagining, the content of utterances.

This simulation-based view of meaning grows out of theories of language and the mind in which "embodiment" plays a central role. The idea of embodiment in cognitive science (Johnson 1987; Lakoff 1987; Varela et al. 1991; Clark 1997; Lakoff & Johnson 1999; Gibbs 2005) is quite straightforward. It's the notion that aspects of cognition cannot be understood without referring to aspects of the systems they are embedded in – in the biology of the organism, including its brain and the rest of its body, and in its physical and social context. When it comes to understanding language, the embodied perspective suggests that meaning centrally involves the activation of perceptual, motor, social, and

* I owe a great debt of gratitude to Art Glenberg and Michael Spivey for invaluable, formative consultation on experimental methods in cognitive linguistics, as well as to Teenie Matlock, Shweta Narayan, Meylysa Tseng, and Kathryn Wheeler for their useful comments and suggestions on this paper. All remaining errors or omissions are exclusively mine.

Gonzalez-Marquez, M., Mittleberg, I., Coulson, S., & Spivey, M. (Eds.). (2007) *Methods in cognitive linguistics.* Amsterdam: John Benjamins.

affective knowledge that characterizes the content of utterances. The way this works is as follows. Through exposure to language in context, language users learn to pair chunks of language like *kick, Mary,* or *John* with perceptual, motor, social, and affective experiences. In subsequent instances of language use, when the original perceptual, motor, social, and affective stimuli are not contextually present, the experience of them is re-created through the activation of neural structures responsible for experiencing them in the first place. This view of meaning is embodied in that meaning depends on an individual having had experiences in their body in the actual world, where they recreate those experiences in response to linguistic input, and use them to produce meaningful linguistic output.

As for the actual mechanisms underlying language processing on this embodied view, understanding a piece of language is hypothesized to entail performing mental perceptual and motor simulations of its content (Narayanan 1997; Barsalou 1999; Glenberg & Robertson 2000; Bergen et al. 2004; Bergen & Chang 2005). This implies that the meanings of words and of their grammatical configurations are precisely the contributions those linguistic elements make to the construction of mental simulations. The study of how different aspects of language contribute to the construction of mental imagery, and the corresponding theory of linguistic meaning as linguistic specifications of what and how to simulate in response to language, is known as *simulation semantics* (Bergen et al. 2003; Bergen & Chang 2005; Bergen et al. 2004; Feldman & Narayanan 2004; see also foundational work on simulation and language by Bailey 1997; and Narayanan 1997). Beyond language, mental simulation or imagery has long been suggested as a fundamental tool for accessing concepts and their properties (Barsalou 1999; Kosslyn et al. 2001) and recalling events (Wheeler et al. 2000).

This embodied view contrasts with an alternative, disembodied perspective, whereupon understanding language can be entirely characterized as the manipulation of abstract symbols (May 1985). The core proposal of this alternative view is that meanings of words and sentences are like a formal language, composed of abstract, amodal symbols, which stand for aspects of the world (e.g., things, relations, properties – see the discussions in Lakoff 1987 and Barsalou 1999). The substance of mental experience, including meaning, is thus entirely unaffected by its biological, physical, and social context. To understand an utterance, the language user maps words onto the semantic symbols that represent their meanings, and these are then aligned as dictated by the sentence (the symbols for the things take their appropriate positions as arguments for the symbols representing relations). For example, a sentence like *Mary kicked John* would be interpreted by determining which symbolic representations are appropriate for each of the words in the sentence, where *Mary* and *John* are represented by unary symbols, perhaps JOHN and MARY, which contrast with the more complex two-place predicate symbol representing *kicked*, perhaps KICKED(X,Y). These symbols are combined appropriately, so that the meaning of the sentence is something like KICKED(MARY,JOHN). The conceptual system, which as it happens is believed itself to be made up of such abstract, amodal symbols, is consequently updated on the basis of the new information that has just been entered into the system. The content of the utterance is thus understood.

There is very limited empirical evidence for the symbol-manipulation view of language understanding (as argued by Glenberg & Robertson 2000). By contrast, there is

Gonzalez-Marquez, M., Mittleberg, I., Coulson, S., & Spivey, M. (Eds.). (2007) *Methods in cognitive linguistics.* Amsterdam: John Benjamins.

substantial support from both behavioral and brain imaging research for the notion that language understanding is based on the unconscious and automatic internal recreation of previous, embodied experiences, using brain structures dedicated to perception and action. This research has begun to uncover the ways in which and the extent to which mental simulation plays a role in language understanding. The goal of the current chapter is to survey the various methods used, with an emphasis on the detailed procedures for performing them. The methods surveyed here address several questions pertaining to mental simulation and natural language understanding: Are mental simulations activated when understanding language? If so, what kind of language triggers them (literal or figurative, concrete or abstract)? What are mental simulations like, and how are they constrained? How are simulations performed, neurally?

Although these questions await conclusive and systematic answers, significant progress has been made, using four principle types of methods. Each is addressed in turn below, with case studies. *Compatibility* effects between the content of language and the performance of actions or perception of percepts tell us that a simulation is performed, and what its motor or perceptual properties are, as do *interference* effects. *Simulation time* effects tell us how the internal dynamics of simulation may be affected by details of the described scenario. Finally, *neural imaging* provides convergent evidence on the localization of simulation.

2. Compatibility effects

Simulation-based theories of language understanding make a straightforward behavioral prediction. If understanding an utterance does indeed involve the activation of perceptual and motor representations, then it should prime these specific, modal representations for subsequent use. For example, it could be that a sentence like *Give Andy the pizza* activates the understander's internal representation of how to move her arm forward (as if to transfer possession of a pizza). If so, then when she is subsequently asked to actually move her real arm forward, she should do so more quickly (the action should be facilitated) than she would without such priming. By contrast, a subsequent action in which the understander performs another, perhaps incompatible action, like moving her arm backwards, should be slowed down (inhibited) by the same sentence. The basic logic of the method is that in order to perform a motor action, one must activate neural motor structures responsible for that particular type of action, and if understanding a sentence leads to increased or decreased excitation of those same neural motor structures, then this should result in quicker compatible (and slower incompatible) actions.

The simulation hypothesis produces the same prediction for perceiving images that it does for performing actions. Just as performing an action should be facilitated or inhibited by preceding language describing compatible or incompatible actions respectively, so perceiving an image (visual, auditory, etc.) should be facilitated or inhibited by language including compatible or incompatible images. For example, a sentence like *The man hammered the nail into the floor* implies a particular direction for the nail, presumably point-down. A very similar sentence, *The man hammered the nail into the wall*, implies a different direction – horizontal – for the nail. We would thus expect to find that tasks involving visual stimuli, where the understander is asked to perceive and perform a categorization task on objects mentioned in a preceding sentence, should show facilitation or inhibition, depending on how well the picture matched the visual characteristics the sentence implied it to have. So after processing a sentence like *The man hammered the nail into the floor*, it should take less time for an understander to perform a task in which she has to perceive a picture of a nail pointing down, and more time for a picture of a nail pointing horizontally.

The main idea that compatibility-type experiments can test is that understanding language activates brain structures responsible for acting and perceiving. They do this through the observation of ways in which language understanding facilitates the performance of described actions or perception of described percepts. Two main lines of research, already hinted at in the above discussion, have addressed this hypothesis, for perception and for action respectively, and each of these is discussed in turn below. We begin with studies demonstrating that after reading a sentence implying an orientation or shape for an object, subjects can perform tasks with an image of that object more quickly if it is displayed with the implied orientation or shape. We then turn to work showing that responses to a sentence are faster when the motion the subject has to perform to respond is compatible with the direction of (hand) motion implied by the sentence.

2.1 Implied object orientation and shape

Does processing sentences automatically and unconsciously activate visual imagery, using visual processing systems? A method for testing this, presented by Stanfield and Zwaan (2001) and Zwaan et al. (2002), had the following setup. The experimenters created pairs of sentences which manipulated the implied orientation (or shape) of objects, like *The man hammered the nail into the floor* versus *The man hammered the nail into the wall.* They then presented an image of the manipulated object, which was either compatible in terms of the manipulated variable, i.e. orientation or shape, or was incompatible. For example, a nail oriented downwards would be compatible with the first sentence above, but incompatible with the second.

Subjects performed one of two tasks. In the first study, subjects were instructed to say as quickly as possible whether the object had been mentioned in the previous sentence. In the subsequent study, they had to simply name the object. This change was motivated by the desire to ensure that any difference in response time resulted from the prior activation of a visual model and was not caused by properties of the task itself. The two tasks differed in that the first, recalling whether the object was mentioned in the previous sentence, explicitly drew the subject's attention to the relation between the sentence and object, while the second, simply naming the object, did not do so at all. Thus, the second method improved over the first since the task it used did not prompt subjects to recall visual memories or even potentially retroactively construct a mental model of the scene described by the sentence.

The research hypothesis in these studies was that the orientation or shape of the object (whichever characteristic was being manipulated) would affect how long it took subjects to respond to images of those objects. This is precisely what was found – when the image of an object was shown with the same orientation or shape it was implied to have in the scenario described by the sentence (e.g. when the nail was described as having been hammered into the floor and was depicted as pointing downwards), it took subjects less time to perform the task than when it was in a different orientation (e.g. horizontal). Zwaan and colleagues also found that when sentences implied that an object would have different shapes (e.g. an eagle in flight versus at rest), subjects once again responded more quickly to images of that object that were coherent with the sentence – having the same shape as they had in the sentence.

In designing a visual compatibility experiment, a number of considerations arise (for more details, see Stanfield & Zwaan 2001; and Zwaan et al. 2002). The first is the nature of the visual property that might be represented in mental simulations of utterance content. While shape and orientation have been studied, others might include color, texture, brightness, size, etc. What is critical in the selection of such a visual feature is that it be possible to construct a large number (perhaps 20–30 pairs) of sentences that, through simple modifications, yield different, incompatible values of the visual property. For example, the sentence *The ranger saw the eagle in the sky* contrasts with *The ranger saw the eagle in the nest* only in the final noun, but the two sentences yield different implied shapes for the eagle (flying versus resting). To ensure that the different sentences can only be reasonably interpreted as implying the predicted visual property, a norming study can be performed, in which subjects (who are different from those participating in the main experiment) are shown the sentence and the two corresponding pictures, and are asked to decide whether one picture, the other, both, or neither matches the sentence.

In selecting or building stimuli, the images that depict the target objects should differ (to the extent possible) only in the relevant visual property. For example, the flying and resting eagle images should be the same size, color, and so on, differing only in the posture the eagle is assuming. A norming study as described in Stanfield and Zwaan (2001) can eliminate the possibility that responses to the pictures are influenced by one version being more canonical or frequent than the other.

One question of potential interest is whether any reaction time differences in the compatible versus incompatible conditions are the result of facilitation, inhibition, or both. In order to assess this possibility, one can use a set of control sentences that mention the target object but do not imply any particular configuration for the visual property in question. For example, *The ranger saw the eagle in the park* does not imply whether the eagle is resting or flying.

Most empirical studies of language processing include filler stimuli, stimuli that are not related to the intent of the experiment, but which are presented to decrease the likelihood that subjects will notice the critical stimuli (the stimuli of interest to the data collection). Experiments like the ones described here often have at least as many fillers as critical sentences. Filler stimuli should be indistinguishable from those in the critical trials. Thus, filler sentences and critical sentences should be the same length, concreteness, etc., and filler images and critical images should be equally large, bright, colored, etc. In half of the trials overall, the image should depict an object mentioned in the sentence (and all of the critical sentence-image pairs should be in this condition). In the other half of the trials (all of which will be fillers), the depicted object should not be mentioned in the sentence. Therefore, depending on the number of fillers included, none to some of them will include an image mentioned in the sentence.

Stimuli should be presented to subjects in four separate lists of stimuli (each subject sees only one of the lists). (If control sentences as mentioned above are included, there is a total of 6 separate lists.) For each critical sentence pair (Sentences 1a and 1b) and its two associated images (Images 1a and 1b), there are four possible presentation combinations: Sentence 1a + Image 1a; Sentence 1a + Image 1b; Sentence 1b + Image 1a; Sentence 1b + Image 1b. Each list should include one of these four versions, for each sentence pair. Thus, if there are 24 critical sentence pairs (resulting in 96 total possible presentation combinations), each list should include a total of 6 of each type of Sentence + Image combination, for a total of 24 critical trials.

Results should be analyzed using Repeated Measures ANOVA (Gravetter & Wallnau 2006; also see Nuñez this volume, for further details on statistical analysis.). ANOVA is a standard statistical test used when independent variables (conditioning factors) are categorical, and when the dependent variable (the thing measured) is continuous. In this particular case, condition (matching versus non-matching (vs. control)) and picture version (which of the two pictures was shown) are within-subjects factors and list (which of the four or six lists was presented) is a between-subjects factor. The hypothesized effect is a main effect of condition, where the matching condition is faster than the non-matching condition.

2.2 The action-sentence compatibility effect

A second compatibility-based method tests the extent to which motor representations are activated for language understanding. The Action-sentence Compatibility Effect (ACE – Glenberg and Kashak 2002) is based on the idea that if language understanders perform motor imagery, using neural structures dedicated to motor control, then understanding sentences about actions should facilitate actually performing compatible motor actions. In ACE experiments, subjects read sentences, of which the critical ones are all meaningful and encode one of two actions – usually motion of the hand away from or towards the body. Subjects indicate whether or not the sentences make sense by pushing a button that requires them to actually perform one of those two actions. For example, in Glenberg and Kashak (2002), sentences encoded movements towards or away from the reader of the sentence, like *Andy handed you the pizza* versus *You handed Andy the pizza.* Subjects started with their hands at an intermediate distance from their body, and then indicated their meaningfulness judgments by pushing a button that was closer to them or farther away from that central location.

The hypothesized effect was an interaction between the direction of motion implied by the sentence and the direction of motion performed by the subject in response to the sentence. Glenberg and Kaschak (2002) report exactly this – a significant interaction between response direction and sentence direction, where responses were faster when the action the subject performed was compatible with the action encoded in the sentence – the Action-sentence Compatibility Effect. For example, responses to sentences that encoded motion toward the subject were faster when the subject had to move her hand towards herself to indicate that it was meaningful than when she had to move her hand away from her body.

A number of considerations are critical to using the ACE to test whether understanding language describing actions makes use of the same cognitive machinery responsible for enacting the same actions. First, the motor actions in question must be both simple to perform and incompatible with each other. Ideally, they use mutually antagonistic muscle groups, like moving one's arm *away* from the body versus moving it *towards* the body, or making a *fist* versus an open *palm* handshape.

Second, these actions should be describable using a broad range of language, and language should exist that can encode one action or the other, depending on a single, simple modification. For example, the verb *catch* implies different handshapes when paired with different direct objects – *catching a marble* involves making a fist while *catching a watermelon* involves more of a palm handshape. Using pairs of sentences that differ only on the basis of a simple modification and strongly imply one type of action or the other decreases the possibility that any ACE effects are produced on the basis of the individual words appearing in the sentences. One way to tell whether sentences imply one type of action or another is to run a norming study, in which subjects are presented with each sentence and asked which of the two actions (or both, or neither) it describes. Moreover, there must be a way for the subjects to respond to the linguistic stimuli (making sensibility or grammaticality judgments) by performing the described action, either by pressing a button placed at a particular location requiring that the subject perform the action to reach it, or simply by performing the action such that it can be videorecorded. The former solution (button press) is preferred to the latter since responses can be automatically recorded, without need for transcription of videotape.

When it comes to filler sentences, there should be at least as many of these as there are critical sentences, and they should be randomly distributed among the presented stimuli, so as to minimize possible effects from trial to trial. As discussed above, filler stimuli should be indistinguishable from those in the critical trials, being of the same length, concreteness, and so on. In half of the trials overall, the sentence should be meaningful (or grammatical, depending on the question subjects are responding to), and all of the critical sentence-image pairs should be in this condition. In the other half of the trials (all of which are fillers), the sentence should not be meaningful (or, in the case of grammaticality judgments, grammatical). Therefore, depending on the number of fillers included, none to some of them will be meaningful. An equal number of the meaningful (or grammatical) and non-meaningful (or ungrammatical) sentences should refer to the same types of action that the critical sentences do, in order to ensure that subjects cannot simply rely on superficial properties of the sentences (thus criteria other than meaningfulness or grammaticality) in order to make their judgments.

Finally, halfway through the experiment, subjects switch which response indicates a 'yes' response and which means 'no'. In order to eliminate order effects, subjects should be randomly assigned to one of two groups, distinguished by the assignment of responses to 'yes' and 'no'.

Results should once again be analyzed using Repeated Measures ANOVA, where sentence-action and response-action (*towards* versus *away*, or *palm* versus *fist*, for example) are within-subjects factors and list (which of the orderings the subject was exposed to) is a between-subjects factor. The hypothesized effect, as mentioned above, is an interaction between sentence-action and response-action, where identical sentence and response actions yield faster reaction times.

2.3 Design issues for compatibility methods

There are several properties of designs for experiments like these that are worth mentioning. First, notice that in designing stimuli for experiments like these, one important consideration is what linguistic sources could potentially yield effects. In the experiments described here, the stimuli never explicitly mentioned the direction of motion, or shape or orientation of the object. Any differences among the sentences, then, must be the product of implied direction, shape, or orientation, which can have resulted only from the construction of a mental simulation of the motor and perceptual content of the utterance. In other words, by making sure that the linguistic stimuli lexically underspecify the independent variable, we can be certain that effects arise not from straightforward lexical meaning, which might itself be a quite interesting effect, but rather from the process of sentence understanding.

Second, these experiments use pairs of stimuli that differ only in a single dimension – the dimension of variation in the implied properties of simulation. Thus, experimenters create sentence pairs like *The man hammered the nail into the floor/wall*, where the only difference is the final word, a word which is not associated with the predicted vertical and horizontal orientation. (In fact, this example is particularly instructive, since wall and floor might in fact in isolation prime the opposite orientation to the one the nail is oriented in – a nail in the floor is vertical, while a floor is predominantly horizontal.) If the main interest of the study is to determine what linguistic properties, e.g. words, phrasal constructions, etc., yield what different effects in visual simulation, then the only differences within pairs of sentences should be the linguistic properties in question.

Third, in their original work on the ACE, Glenberg and Kaschak included only sentences that included the experimental subject in their content, like *Andy handed you the pizza*. However, the simulation semantics hypothesis claims that even understanding language that does not involve the understander, like *Andy handed Sheila the pizza* should engage mental simulation. Indeed, the results from the visual compatibility studies described above, using sentences like *The ranger saw the eagle in the sky* do not involve the experimental subject at all. Subsequent research on the ACE has demonstrated that motor actions described as being performed by someone other than the experimental subject also yield significant effects (Bergen & Wheeler 2005; Tseng & Bergen 2005).

Finally, to position the research described here in terms of the literature, the visual compatibility experiments are a form of priming experiment, in which the presentation of a given stimulus is assumed to yield brain activation that makes faster (or facilitatorily primes) a response governed by brain structures that are identical or connected to those activated by the first stimulus. In particular, in the studies by Zwaan et al. discussed above, the primed response in question is a response to a compatible or incompatible percept. The Action-Sentence-Compatibility effect is a particular sort of priming effect, where the response may itself be compatible or incompatible with the original stimulus. Studies of this form go under the rubric of Stimulus-Response experiments. These effects are analyzed as resulting from "common coding": the neural substrate of the stimulus and the response are overlapping. Other Stimulus-Response Compatibility effects have been shown for spatial location (the Simon Effect – Simon 1969) as well as cross-modal associations like intensity of sound with physical force (Romaiguère et al. 1993; Mattes et al. 2002), among others.

3. Interference effects

Closely related to compatibility effects are *interference* effects. Compatibility effects, as described in the previous section, ostensibly arise from the fact that understanding language about an action activates neural machinery responsible for performing the described action or perceiving the described percept. As a result, identical actions or percepts are facilitated subsequent to language processing. In each of the studies described above, the subjects executed a response after having interpreted the stimulus. As such, these compatibility effects can be seen as a type of priming – a set of neural structures is activated by one activity and thus subsequent use of the same structures is facilitated. Something different happens, however, when the same neural structures are recruited by multiple tasks at the same time. For example, if a particular utterance activates motor or perceptual structures, and if the subject is asked to simultaneously perform another motor or perceptual task, which presumably makes use of the same neural structures, then we will observe not facilitation but interference between the two tasks. To reiterate, interference effects, like compatibility effects, result from the use of the same neural structures to understand language and perform a perception or motion task, but differ from compatibility effects in that understanding the language and performing the perceptual or motor task require the same neural structures to perform different tasks at the same time. The causes for compatibility versus interference effects are somewhat more complex than just temporal overlap, and are discussed in more detail at the end of this section.

Existing interference studies are all based on the use of visual stimuli that are either compatible or incompatible with the presumed simulation evoked by language that is produced at the same time, or immediately before or afterward. However, within this framework, two types of interference have been investigated, deriving from perceptual and motor effects. The first of these lines of research – investigating perceptual interference – is based on an effect known as the Perky effect (Perky 1910; Segal & Gordon 1969). The second is based on recent neuroscientific work on the use of motor systems to perceive and understand actions. Each is addressed in turn below.

Gonzalez-Marquez, M., Mittleberg, I., Coulson, S., & Spivey, M. (Eds.). (2007) *Methods in cognitive linguistics.* Amsterdam: John Benjamins.

3.1 Visual interference effects

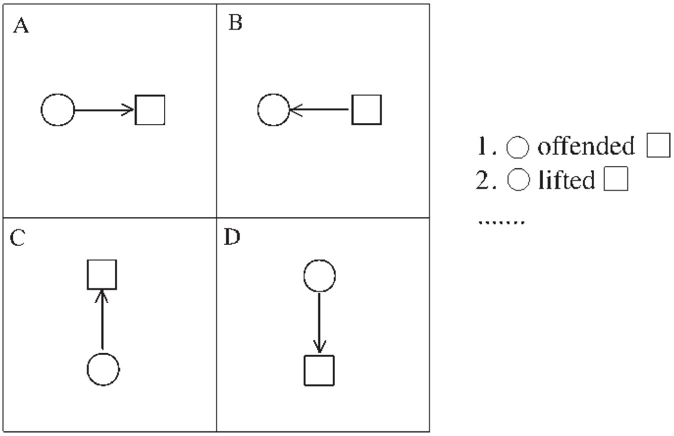
Researchers interested in testing whether understanding language with visual content makes use of the visual system look for ways in which processing such language interferes with the simultaneous processing of visual percepts. It has been known for a century that visual imagery can selectively interfere with visual perception. Early work by Perky (1910) had subjects imagine seeing an object (such as a banana or a leaf) while they were looking at a blank screen. At the same time, unbeknownst to them, an actual image of the same object was projected on the screen, starting below the threshold for conscious perception, but with progressively greater and greater definiteness. Perky found that subjects continued to believe that they were still just imagining the stimulus, and failed to recognize that there was actually a real, projected image, even at levels where the projected image was perfectly perceptible to subjects not performing simultaneous imagery.

More recent work on the Perky effect has shown that inteference can arise not just from shared identity of a real and imagined object, but also from shared location. Craver-Lemley and Arterberry (2001) presented subjects with visual stimuli in the upper or lower half of their visual field, while they were performing imagery in the same region where the visual stimulus was, or in a different region, or were performing no imagery at all. They were asked to say whether they saw the visual image or not, and were significantly less accurate at doing so when they were imagining an object (of whatever sort) in the same region than when they were performing no imagery. Performing imagery in a different part of the visual field did not interfere with the visual discrimination task at all.

If Perky effects like these are indeed indicative of visual imagery making use of the same neural resources recruited for actual vision, then they can naturally be extended to language processing. Rather than asking subjects to imagine visual objects, experimenters can ask subjects to process language hypothesized to evoke visual imagery of a particular type – of particular objects with particular properties, or of objects in particular locations. If visual language selectively activates visual imagery, then we should expect a Perky-type effect that results in interference between the displayed visual image and the visual properties implied by the language.

This is precisely the tack taken by Richardson et al. (2003) and subsequently by Bergen et al. (2007). In the work by Richardson and colleagues, subjects heard sentences whose content had implied visual characteristics and then very quickly thereafter performed a visual categorization task where the image either overlapped with the sentence's meaning or did not. They hypothesized that if sentence understanding entailed visual imagery, then there should be Perky-like interference on the object categorization task – it should take longer to categorize an object when it had visual properties similar to the image evoked by the sentence.

More specifically, Richardson et al. suggested that processing language about concrete or abstract motion along different trajectories in the visual field (like vertical versus horizontal) leads language understanders to activate the parts of their visual system used to perceive trajectories with those same orientations. For example, a sentence like *The poacher hunts the deer* implies horizontal motion, while *The ship sinks in the ocean* implies vertical motion. If understanders selectively perform vertical or horizontal visual imagery



1. ○ offended □
2. ○ lifted □
....

Figure 1. The four possible images from which subjects selected in Richardson et al. (2001)

in processing these sentences, then when they are immediately afterward asked to visually perceive an object that appears in their actual visual field, they should take longer to do so when it appears on the same axis as the motion implied by the sentence. Thus after *The poacher hunts the deer*, subjects should take longer to categorize an object, say as a circle versus a square, when it appears to the right or left of the middle of the screen, but their categorization rate should not be affected when the visual object appears above or below the middle of the screen.

In order to construct stimuli for this experiment, Richardson and colleagues performed two off-line norming studies (described in detail in Richardson et al. 2001), in which subjects provided their intuitions about whether sentences containing particular verbs have vertical or horizontal meanings. The two norming studies used different methods. In the first, subjects simply picked one of four visual images depicting horizontal or vertical motion (depicted in Figure 1) that they decided best captured the meaning of the sentence. In the second, subjects themselves used a graphical interface to produce representations of the scene described by the sentence using the same primitives presented in the images in Figure 1 – circles, squares, and arrows. The interest in both of these tasks was to determine whether language users uniformly assign a vertical or horizontal interpretation to sentences, such that those sentences could be used as stimuli in the Perky experiment.

One additional point of interest here regards the nature of the sentences used. The experimenters were interested in the spatial orientation not just of concrete verbs, like *hunt* and *sink*, but also of abstract verbs, like *respect* and *tempt*. They wanted to determine whether abstract events, just like concrete events, were selectively associated with particular spatial orientations. How abstract concepts are represented and understood is a critical question for all theories of meaning and understanding, but is particularly vital to simulation-based models, because of their reliance on perceptual and motor knowledge. It may not be obvious at first blush what the nature of mental simulations that capture the embodied understandings of abstract notions like *respect* and *tempt* might be. There are insightful discussions of how abstract concepts can be grounded in embodied systems in various places (Lakoff 1987; Barsalou 1999; and Kaschak & Glenberg 2000), so the topic will not be explored in depth here. For current purposes, it will have to suffice that it is an interesting question whether abstract events contain a spatial component. Abstract verbs were thus included in their norming studies, as well as in the Perky experiment described below.

Richardson et al. took verbs, with their associated horizontality/verticality ratings, and presented them to subjects in the interest of ascertaining whether they would induce Perky-like effects on the categorization of visual objects (shapes) that were presented on the screen in locations that overlapped with the sentences' implied orientation. After seeing a fixation cross for 1 second, subjects heard a sentence, then, after a brief pause (randomly selected for each trial from among 50, 100, 150, or 200ms), they saw a visual object that was either a circle or a square, positioned in one of the four quadrants of the screen (right, left, top, or bottom). Their task was to press a button indicating the identity of the object (one button each for 'circle' and 'square') as quickly as possible.

The categorization task the subjects were performing was not transparently related to the sentences that preceded it. This type of design has advantages and disadvantages. One major advantage to having subjects perform a categorization task (rather than a more directly related one, like an up-down or left-right task) is that they are less likely to become aware of the relation between the independent and dependent variables, and thus other high-level cognitive processes involved in reflection are less likely to be activated. In other words, the dependent measure is less likely to be influenced by confounding factors arising from subjects' guessing the purpose of the experiment or simply recognizing the potential relation between the sentence understanding task and the object categorization task.

A disadvantage of this method, compared for example with one in which subjects are asked to imagine the content of the sentences that are presented, is that subjects may be prone to ignoring the content of the utterances, and paying particular attention to the visual objects. In order to eliminate this temptation, it is standard in studies of this sort to include comprehension questions following an unpredictable (at least, to the subjects) subset of the sentences. This ensures that, since they do not know which sentences will be followed by a comprehension question, subjects attend to the meaning of the sentences, even though they don't know that they're really being forced to do so for the purpose of the object categorization task.

Richardson et al. (2003) constructed filler sentences that could be followed by comprehension questions, and added these to the critical sentences, of which fifteen had verbs that had been designated as the most vertical and fifteen the most horizontal, based on the norming studies. It should be noted that in studies in which verbs are more similar to one another (e.g. Bergen et al. 2007), there is a need for a greater number of filler sentences, to obscure the intent of the experiment.

The results were indicative of a clear interference effect – subjects took longer to categorize objects in the vertical axis when they followed vertical sentences than horizontal sentences, and vice versa for objects in the horizontal axis.

A number of subsequent studies have demonstrated Perky effects in language comprehension, and have explored different variations of the effect. The original study did not indicate what parts of the sentences yielded the interference effects that were found. To investigate this question, Bergen et al. (2007) independently varied the verbs and subject nouns in intransitive sentences to determine whether either could single-handedly yield interference effects, and found a strong interference effect when just subject nouns or verbs were associated with a particular region of space. Another methodological modification in that same follow-up study was to split up the vertical axis into the upper and lower regions by using more specifically upwards- or downwards- oriented sentences, rather than conflating them together into a single, vertical condition (a strategy also independently adopted by Lindsay 2003; and Kaschak et al. 2005). Bergen et al. found significant interference within two separate vertical conditions, such that upwards-oriented sentences (like *The mule climbed*) made subjects take longer to categorize objects in the upper quadrant of the screen, but not in the lower quadrant. Finally, the Richardson et al. study did not yield any indication of the role of spatial processing in the comprehension of concrete versus abstract language. The Bergen et al. study separately studied literal (*The mule climbed*), metaphorical (*The rates climbed*), and abstract (*The rates increased*) language, and found that while there were strong Perky effects for literal sentences, there were none in response to abstract or metaphorical utterances.

Another recent follow-up (Lindsay 2003) switched the order of the visual perception and sentence comprehension tasks. In this work, subjects first saw a rectangle move on the screen in one of four directions (upward, downward, rightward, or leftward), and then performed a language comprehension task, in which they read a sentence, and pressed a button as soon as they had understood it. This experiment produced a significant interference effect on reading time of compatible motion – just the same as in the Perky experiments described above, but with the tasks in the reverse order.

Finally, Kaschak et al. (2005) have used a quite similar methodology with slightly different spatial dimensions. In their work, subjects heard a sentence that indicated motion upwards, downwards, toward the subject, or away from the subject, while they simultaneously observed a visual illusion of motion in one of those directions. The subjects' task was to respond as quickly as possible whether the sentence was meaningful or not. The results demonstrated a clear interference effect – it took subjects longer to respond that the sentence was meaningful when it was presented simultaneously with a visual illusion depicting motion in the same direction.

As has been shown above, visual interference effects are reliable and replicable, in a number of methodological permutations. These findings as a whole can be taken as evidence that perceptual systems – in particular the visual system – are unconsciously and automatically engaged in the process of natural language understanding.

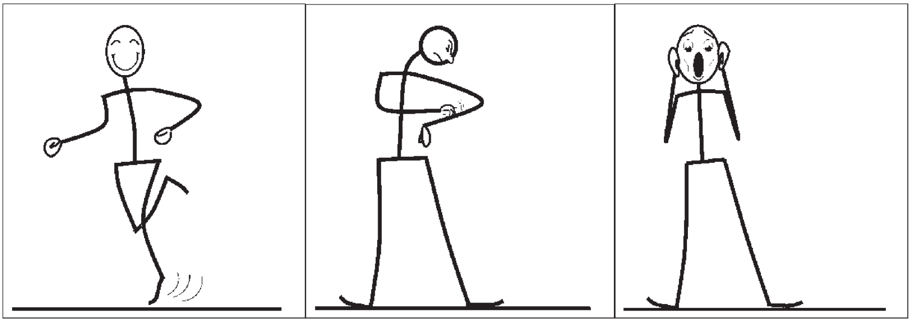
3.2 Motor interference effects

As we have seen, the tasks of understanding language about perceptual content and engaging in visual perception can interfere with each other, when performed simultaneously or in rapid succession. Extending this effect to the domain of motion, we can ask whether understanding language about motor actions can similarly interfere with activation of the motor machinery responsible for performing the described actions. To the author's knowledge, no studies currently exist that have tested for interference effects between performing actual actions and understanding language describing the same actions (that is, an interference version of the Action-sentence Compatibility effect, described above). However, a method does exist for testing the activation of motor structures in response to motor language, albeit indirectly, using a type of cross-modal matching.

The use of this cross-modal matching methodology (discussed in Bergen et al. 2003; Bergen et al. 2004; Narayan et al. 2004; and Chan ms.) is based on the relatively recent discovery that perceiving actions activates certain neurons in motor areas that are responsible for enacting those same actions – so-called "mirror neurons" (Gallese et al. 1996; Rizzolatti et al. 1996). Mirror neurons are cells in the motor cortex of monkeys, and presumably also humans, that are selectively activated during the performance of specific motor functions, but which also become active when the individual perceives another person or monkey performing the same function. There are few single neuron studies in humans, but comparable "mirror activity" patterns in humans have been demonstrated through brain imaging (Tettamanti et al. 2005). It has also been established that this mirror system extends to the somatotopic organization of the pre-motor and parietal cortex (Buccino et al. 2001). In particular, the execution or observation of actions produced by the mouth, leg, and hand activate distinct parts of pre-motor cortex, found in ventral sites, dorsal sites, and intermediate foci, respectively. When appropriate target objects are present, there is also activation in a somatotopic activity map in parietal cortex. Mirror neurons and the circuits they participate in have thus been shown to serve dual roles in producing actions and recognizing those actions when performed by others.

Thus, the reasoning goes, when subjects are asked to perceive an action, they activate parts of motor cortex responsible for performing the same action. If they are also asked to simultaneously understand language pertaining to an action, then we may see interference effects when the two actions overlap – just as perceiving an image and simultaneously understanding language that overlaps with that image interfere with each other in the visual domain. Since we know that mirror circuitry is organized by effector, e.g. hand, leg, or mouth, it might be the case that perceiving actions used by a particular effector may selectively interfere with processing language describing actions performed using the same effector. This should contrast with perceiving an action and processing language that indicates actions performed by different effectors, which should not be subject to the same selective interference effect.

In a series of studies (Bergen et al. 2003; Chan, ms.), subjects were shown a stick-figure image of some type of action (performed primarily with the mouth, hand, or leg) and a verb that also described some such action. Subjects were instructed to decide as quickly as possible whether the verb and the image depicted the same action or different actions. The cases of interest were those where the verb and image did not depict the same action. It was hypothesized that subjects should take longer to decide that the verb and image did not match if their actions were both primarily executed using the same effector, compared to the case where they used different effectors. The simulation-based explanation for this hypothesis was straightforward – if perceiving actions and understanding language about actions makes use of motor structures, then very similar actions presented in the two



Match	"hop"	"scratch"	"scream"
Same	"punt"	"juggle"	"lick"
Diff	"juggle"	"punt"	"kick"

Figure 2. Sample stimuli for the image-verb matching task

modalities should yield interference, much like the interference seen in the Perky-type studies described above.

Sixty image stimuli were drawn using a graphics program for these experiments by one of the experimenters, and those images were then paired with verbs through a norming study (described in Bergen et al. 2003). In that study, subjects were asked to provide the word that they thought best captured the action depicted by the image. From the original set of images, 16 images depicting hand, mouth, and leg actions (for a total of 48 images) were selected, on the basis of uniformity of subjects' responses. The matching verb for each image was the one most frequently identified by subjects, and the non-matching verbs using the same and different effectors were randomly selected from among those verbs that no subjects said were the best description of the image. Some of the non-matching pairings were subsequently changed in cases where the experimenters believed the randomly selected verb might plausibly be interpreted as describing the image. In a subsequent study (Chan, ms.), verbs were selected by having subjects choose the best verb for an image from among a small set of options. Some examples of stimuli for the three effectors are seen in Figure 2, along with examples of matching verbs, as well as non-matching verbs in both the same-effector and different-effector conditions:

In the original study (Bergen et al. 2003), the image was presented before the verb. Each trial consisted of a visual stimulus like the images shown above, which was presented for one second, followed immediately by a 500 millisecond interstimulus interval, the first 450 milliseconds of which included a visual mask covering the whole screen. The visual mask was meant to reduce any priming effects that resulted from visual imagery. An English verb in written form was then presented until the subject pressed a button indicating that the verb was or was not a good description of the action depicted in the image. The verb fell into one of the three conditions described above: (1) matching, (2) non-matching same effector, and (3) non-matching different effector.

The results of this experiment were as predicted – subjects took significantly longer to respond that an image and verb did not match when the two actions were produced by the same effector – the mean size of the effect was around 50ms.

One drawback of this method is that it does not eliminate several possible confounds. A first potential confound involves the role of memory in the observed interference effect. It has been shown through neural imaging that recalling motor actions results in the selective activation of parts of motor cortex specialized for performing those same actions (Wheeler et al. 2000; Nyberg et al. 2001). It could be that activation of motor cortex, which produces the interference effect observed in the first study, arises due to demands of the experiment, namely that subjects are required to recall a motor action in order to subsequently decide whether it matches a verb or not. In order to eliminate this possibility, Chan (ms.) reversed the order of the image and verb stimuli, such that subjects were now not recalling an image such that it could be compared to a verb, but were recalling the meaning of the verb, such that it could be compared to the image. In her experiment, Chan found the very same, significant interference effect. She ran the experiment in English and Chinese, and English-speaking subjects took on average 35ms longer to respond in the same effector condition than in the different effector condition, while Chinese-speaking subjects took 85ms longer in the same effector condition.

A second possible confound stems from the possible ambiguity of the images. If the images look in some way like subjects' mental representations of the actions described by the non-matching same effector verbs (for example, if the image of *hop* above looks like it could be an instance of *punt*), then we would anticipate subjects should take longer to reject same-effector verbs than different-effector verbs, simply because they look more like the actions described by the verbs and it thus takes them longer to determine that the verb in fact is not a good description of the action. Notice that this was still a possible confound, despite the experimenters' attempts to pair together same-effector verbs and images that were unambiguous. In order to assess the viability of this explanation, we constructed another version of the task, in which trials paired together not a verb and an image but two verbs (Narayan et al. 2004). In this version of the method, subjects were asked to decide as quickly as possible whether two verbs that were presented in sequence (with just the same procedure used in the previous experiment, except that the image was now replaced by a verb) meant nearly the same thing or not. The same three conditions were possible – the verbs could describe very similar actions (like *run* and *jog*), or could describe different actions, using the same (*run* and *dance*) or different (*run* and *sneeze*) effectors. If the effects we found with image-verb pairs disappeared in this condition, it could be concluded that the original interference effect derived from ambiguity of the original images. If, however, the interference remained in the verb-verb matching study, then properties of the image used would not be a viable explanation for the effect, since no images were used. Narayan et al. once again found an interference effect – subjects took on average 100ms longer to reject the non-matching word pairs when the actions they described used the same effector than when they used different effectors. Thus, the interference effect cannot be due simply to ambiguity of the images used in the original experiment.

The final potential confound requires slightly more explanation. It could be that the difference in response time results from greater overall similarity between the actions in the same-effector condition than between those in the different effector condition. This would mean that effectors might have nothing to do with the effect, which arises strictly on the basis of similarity – namely that it takes longer to reject concepts as identical concepts the more similar they are. Conceptual similarity of actions, regardless of effector identity, is difficult to assess objectively, but a related and quite accessible tool, Latent Semantic Analysis (LSA – Landauer et al. 1998, http://lsa.colorado.edu/) affords a useful substitute.

LSA, among other things, is a statistical method for extracting and representing the similarity between words or texts on the basis of the contexts they do and do not appear in. Two words or texts will be rated as more similar the more alike their distributions are. LSA has been shown to perform quite like humans in a range of behaviors, including synonym and multiple-choice tasks. Of relevance to the current discussion is the pairwise comparison function, which produces a similarity rating from –1 to 1 for any pair of texts. Identical texts have a rating of 1, while completely dissimilar ones would have a rating of –1.

LSA was used to determined the semantic similarity between the presented verb and the verb that was most commonly associated with the particular image in the pretest described above. This similarity rating was then used as a substitute for the conceptual similarity of the actions they denoted. In other words, for the three examples in Figure 1, there was a semantic similarity score between run and run (matching), between run and kick (non-matching, same effector) and between run and drink (non-matching, different effector). This is an indirect way of evaluating the similarity between an image and a verb, since it is mediated by a verb describing the image, but for the time being it has to do in the absence of more direct methodologies.

With LSA ratings assigned to each trial, the average response time per trial (that is, per image-verb pair) was entered into a regression analysis along with the LSA rating for the trial, as described above. This regression included only the non-matching conditions, as including the matching condition (with LSA ratings of 1) produces an abnormal distribution (all matching cases by definition have an LSA value of 1, which is not particularly interesting). Considering only the two non-matching conditions, there was a very weak correlation between LSA rating and reaction time ($R = 0.094$). The trend for subjects to take longer to reject more similar pairs of words and pictures than less similar ones was insignificant ($p = .378$). So while the similarity between a non-matching verb and image as measured by LSA qualitatively seems to account for a small amount of the variance in reaction time, it does not do so significantly. Of course, this does not prove that sharing an effector and not other sorts of similarity is responsible for the reaction time effects we've seen. The LSA rating might be a flawed measure of similarity in general or with respect to verb-image similarity. For this reason, further studies like the ones described below will be required to test whether the effects are actually based on effector identity. The absence of a significant relation between LSA rating and reaction time shown by the regression above does, however, suggest that overall similarity does not transparently account for the interference behavior that was found.

To conclude, cross-modal matching provides a way to test for the activation of motor circuitry during the processing of motor action language. The interference effects shown using this methodology indicate that determining the meaning of an action verb uses overlapping neural resources with the systems responsible for perceiving actions, which themselves are partially constituted by motor structures.

3.3 Interference or compatibility?

As we have seen, interference effects, like compatibility effects, are used to assess access to detailed perceptual and motor representations during various sorts of language processing. But when should one anticipate interference between similar actions or percepts, and when is compatibility most likely? The current state of knowledge is that interference arises when the two matching tasks are performed at the same time, as in the interference studies described above. Kaschak et al. (2005) argue that temporal overlap by itself does not suffice to account for interference effects, but can only do so in combination with integratabilty. On their model, interference effects arise only when the content of a simulation evoked by language is performed simultaneously with a response, and also cannot be integrated with it (to take an example from their study, an image of a spiral apparently moving towards the subject gives the appearance of forward motion but cannot be readily integrated into a simulation of walking towards a building).

Whether or not it is a sufficient condition, it seems that temporal overlap is not a necessary condition for interference. In the image-verb matching task, the stimuli were presented with some delay (500ms), but we nevertheless observe an interference effect. We should not be misled by this case, however – even though the verb and image or two verbs were not presented simultaneously, the interference presumably results from the co-activation of the two motor images required by the task – in order to perform a comparison, both representations need to be activated simultaneously. More critically, though, Lindsay (2003) reports a significant interference effect when the prime and target stimuli were separated by 1500ms. If this finding is reliable, as it appears to be, then it poses problems for the idea that incompatibility effects arise from temporal overlap. One interesting way in which this study differs from the others discussed above that display compatibility effects is in the order of presentation of the sentence and perceptual stimulus. In Lindsay's study, a visual prime (a rectangle moving along the vertical or horizontal axis on a computer screen), preceded a sentence target to be read and understood. By contrast, other such studies with little or no temporal overlap (e.g. Stanfield & Zwaan 2001; Zwaan et al. 2002; Glenberg & Kaschak 2002) all presented the sentence stimulus first. To generalize at this juncture might be premature, but it appears that matching motor or perceptual processes will yield interference when (1) they overlap temporally, or (2) the sentence is presented after the image or action. They will result in compatibility effects when (1) they do not overlap temporally, and (2) the sentence is presented before the image or action.

4. Simulation time effects

Mental simulations are clearly not identical to the real-world experiences they recreate along a number of dimensions, including geometry, physics, and, importantly, time. But it remains an open question precisely how alike real world experiences and simulations are. Since mental simulations appear to include some fine-grained perceptual and motor detail, it might also be the case that they encode some degree of temporal detail – they might unravel over a course of time that correlates positively with the amount of time it would take to perform or perceive the same event in the real world. Events that take longer in the world might take longer to simulate.

In fact, it seems that when people consult a mental image, they scan through it in a way that mirrors actual visual scanning, such that the time it takes to mentally scan from one point in their mental image to another reflects increased time that it would take to actually visually scan between the same two points. Evidence for this observation comes from map tasks. It has been shown that people who study a map and memorize locations on it take longer to mentally scan between landmarks the farther apart they are located on the original map (Kosslyn et al. 1978). More recent work has shown the same effect when a map is simply described, rather than visually inspected (Denis & Cocude 1989).

If imagery time correlates with real time, then simulations evoked by language should take longer, the longer the events they describe take. A straightforward way to investigate the relation between real and simulated time was devised by Matlock (2004). The basic setup of the methodology is to have subjects read a paragraph describing a scenario, which in one way or another evokes relatively slow or relatively fast motion. This paragraph is followed by a final sentence that might be coherent with the paragraph or not – subjects are asked to read this final sentence and decide if it fits with the paragraph. Matlock proposed that language describing slower motion should result in slower, thus longer, mental simulations, and should thus yield longer response times.

As it turns out, Matlock was primarily interested in the processing not of literal motion language, but of fictive motion language. Fictive motion language (Talmy 1996; Langacker 1986) describes static events and scenes using motion language. For example, in *The road meanders through the valley* the road itself does not actually move any more than the fence does in *The fence runs from the house down to the road.* The question Matlock asked was whether the processing of fictive motion sentences displayed time effects, resulting from simulations taking more or less time.

For example, subjects read one of the following paragraphs, intended to evoke motion along a short or a long path:

> *Short Distance Scenario*
> Imagine a desert. From above, the desert looks round. The desert is small. It is only 30 miles in diameter. There is a road in the desert. It is called Road 49. It starts at the north end of the desert. It ends at the south end of the desert. Maria lives in a town on the north end of the desert. Her aunt lives in a town on the south end. Road 49 connects the two towns. Today Maria is driving to her aunt's house. She is driving on Road 49. It takes her only 20 minutes to get to her aunt's house. After she arrives, Maria says, "What a quick drive!"

> *Long Distance Scenario*
> Imagine a desert. From above, the desert looks round. The desert is large. It is 400 miles in diameter. There is a road in the desert. It is called Road 49. Road 49 starts at the north end of the desert. Road 49 ends at the south end of the desert. Maria lives in a town on the north end of the desert. Her aunt lives in a town on the south end. Road 49 connects the two towns. Today Maria is driving to her aunt's house. She is driving on Road 49. It takes her over 7 hours to get to her aunt's house. After she arrives, Maria says, "What a long drive!"

Subjects subsequently saw a fictive motion sentence like the following, and were asked to decide whether it related to the story or not:

> *Target sentence:*
> Road 49 crosses the desert.

If subjects took significantly longer to respond 'yes' to the sentence when the paragraph described slow motion than when it described fast motion, then this would imply that subjects were performing longer simulations when the language described motion that would take longer to actually perform or observe.

In order to ensure that any difference in response time did not arise from differences in how well the target sentences fit with the preceding paragraphs, the paragraph-sentence pairs were subjected to a norming procedure. A different set of subjects was asked to rate the paragraph-sentence pairs for how well they went together, on a 1–7 scale. The results showed that the short-distance and long-distance motion paragraphs were indistinguishable in terms of their fit with the final sentence.

The experimental results were exactly in line with the research hypothesis – subjects took 391ms longer to read and make a decision about a fictive motion sentence when it followed a paragraph describing long movement than when it followed a description of short movement. The same significant differences were subsequently replicated for paragraphs that differed in the speed of travel (fast versus slow), and difficulty of the terrain (difficult to navigate versus easy to navigate).

The proposed explanation for these results, that subjects construct a mental model while understanding language and subsequently perform a mental simulation using it to interpret the target fictive motion sentences, is only one of two possibilities. The other is that something about processing language about slow versus fast motion results in slower or faster processing of subsequent language in general. In other words, perhaps the subjects, through processing language about speedy motion or motion over a short path, found themselves in a fast-processing mindset, which resulted in faster responses, regardless of properties of the final sentence. In order to eliminate this possibility, a control study was conducted, in which each of the presented paragraphs was followed by a sentence not encoding motion, fictive or otherwise. Each of these sentences was determined (through another norming procedure) to be comparable in meaning to the fictive motion sentences originally used. For instance, *The road is next to the coast* was a control sentence that was equivalent to the fictive motion sentence *The highway runs along the coast*. If the differences in response time following the short versus long or fast versus slow motion paragraphs had arisen with the non-motion target sentences, then the second explanation, evoking differences in global processing, would be viable. However, no such difference was found

5. Neural imaging

The behavioral evidence from compatibility, interference, and simulation time studies provides strong indications that shared cognitive mechanisms effect the processing of both percepts and actions on the one hand and mental simulations of those same percepts and actions in response to language on the other. However, without convergent evidence from neural imaging studies, it is impossible to draw the strong conclusion that it is those brain areas principally responsible for acting or perceiving that are engaged for language understanding.

The main techniques used for imaging the living human brain are are PET (Positron Emission Tomography) and fMRI (functional Magnetic Resonance Imaging). Both methods are non-invasive and function through the detection of metabolic changes in particular regions of the brain, since increased blood flow correlates positively with neural activity. The two methods detect blood flow to particular regions of the brain in different ways. In PET studies, a radioactive substance emitting positrons is introduced into the subject's bloodstream and blood flow to particular regions is measured by the intensity of positron emissions in those regions. In fMRI studies, nothing is injected into subjects – rather, changes in the magnetic resonance of regions of the brain, resulting from changes in blood flow, are measured using magnetic fields and radio waves. Crucially, though their temporal and spatial acuity differ, both methods allow a snapshot of the brain at a given time, including indications of where neural activity is taking place.

While the phrenologists of the 19th century were mistaken about the possibility of inferring mental properties of individuals from the superficial structure of their skull, they were right that many cognitive functions appear to be at least somewhat localized to particular brain regions, though these may differ more or less among individuals. And neural imaging, along with studies of brain damaged patients, invasive single-cell studies of brain surgery patients, and work with other animals, has provided a great deal of insight into the neural substrates of action and perception. Of particular relevance to the study of simulation, the major responsibility for detailed motor control is shared by a set of motor areas, including primary, supplementary, and secondary motor cortices, as well as regions of the cerebellum. Each of these areas is structured somatotopically – distinct body regions map onto distinct regions of the given brain area, as shown in Figure 3 for primary motor cortex. Similarly, visual cortical areas are arranged retinotopically, such that parts of the retina are mapped spatially onto parts of the visual cortex. During any cognitive behavior, neural activity is not strictly restricted to those brain areas primarily associated with the particular function being performed. Nevertheless, certain brain regions are reliably associated with perceiving objects in particular places in the visual field, and for performing actions with particular effectors.

As it turns out, imagery appears to be selectively executed, at least in part, by the very motor and perceptual areas responsible for the real-world correlates of the particular imagery performed. For example, motor imagery activates the same parts of motor cortex responsible for performing actions using the very same effectors (Porro et al. 1996; Lotze et al. 1999; Ehrsson et al. 2003). Visual imagery selectively activates brain regions responsible for perceiving similar images (Kosslyn et al. 2001), including primary and secondary visual areas. Recalling motor control events also reactivates modal brain structures responsible for performing the very same actions (Wheeler et al. 2000; Nyberg et al. 2001).

Motor and perceptual imagery and memory seem to make use of the specific brain regions used to perceive or perform the given experiences, and we have seen from the behavioral studies above that language understanding makes use of motor and perceptual

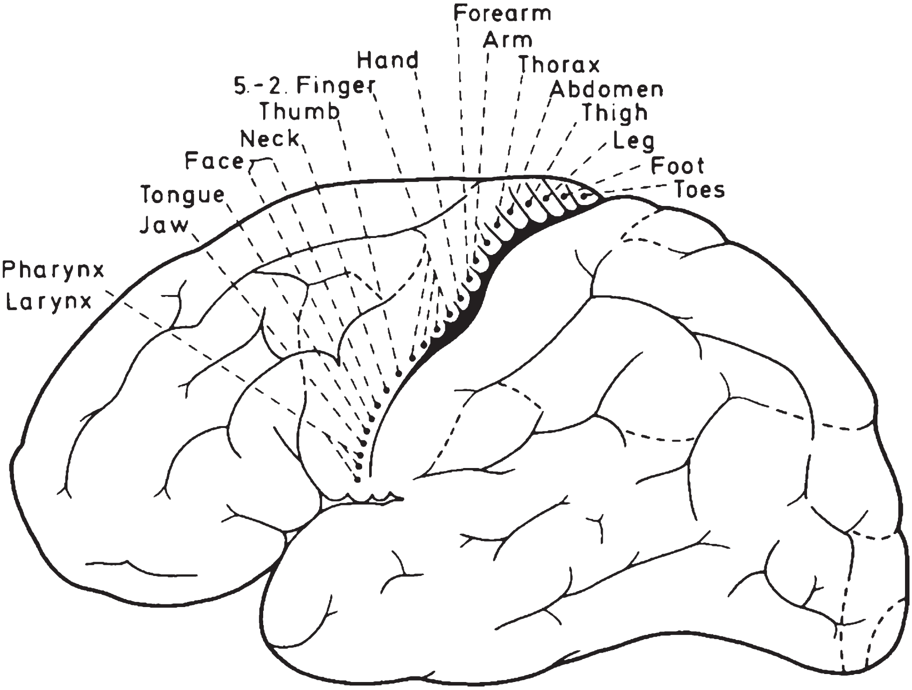


Figure 3. Somatotopic organization of the primary motor strip (from Brodal 1998)

imagery. It would thus be entirely unsurprising for language processing to make use of the same brain regions. Indeed, this is precisely what would be predicted by the simulation semantics hypothesis. Several recent studies show that motor and pre-motor areas associated with specific body parts (i.e. the hand, leg, and mouth) become active in response to motor language referring to those body parts. For example, Pulvermüller et al. (2001) and Hauk et al. (2004) found that verbs associated with different effectors were processed at different rates and in different regions of motor cortex. In particular, when subjects read verbs referring to actions involving the mouth (*chew*), leg (*kick*), or hand (*grab*), the motor areas responsible for mouth versus leg versus hand motion received more activation, respectively. In sentence processing work, Tettamanti et al. (2005) have also shown through imaging that passive listening to sentences describing mouth versus leg versus hand motions activates different parts of pre-motor cortex (as well as other areas).

6. Conclusions

This chapter began by outlining a concrete elaboration of the notion that the content of the mental processes underlying language use are inherently embodied. On this particular view, understanding a piece of language entails performing perceptual or motor simulations, or both. In performing these simulations, the language understander effectively creates or recreates perceptual or motor experiences, using a set of brain struc-

Gonzalez-Marquez, M., Mittleberg, I., Coulson, S., & Spivey, M. (Eds.). (2007) *Methods in cognitive linguistics.* Amsterdam: John Benjamins.

tures that overlap with those used to perceive the described percepts or perform the described actions.

The methods we have surveyed here – compatibility effects, interference effects, simulation time effects, and neural imaging, have all incrementally contributed to the body of convergent evidence that now supports the simulation semantics view of language understanding. It appears that language understanders naturally perform both motor and visual simulations, and that the motor and visual systems participate in these processes. Further, they do so in a selective manner – language about events that would by default take place in the upper quadrant of the visual field, for example, specifically makes use of parts of the visual system responsible for perceiving the upper quadrant of the visual field.

Through the application of methods like those described above, and their successors, to a set of progressively more accutely refined questions, the coming years are poised to yield enormous insights into the role of simulation in language processing. Among the major questions that will surely be addressed in detail are the following. What do different types of linguistic units (like different parts of speech, but also different types of phrasal patterns) contribute to the content or form of mental simulations? How do motor and perceptual simulations relate to each other – are they mutually exclusive or can they co-occur (and if the latter, how?) How are different perceptual perspectives taken during the enactment of mental simulations, and how does language trigger these? Are simulations different for speakers of different languages? And finally, how closely do simulations adhere to properties of the real world, like time, space, and physics? With the increasing availability of a broad range of empirical methods, the coming years will bring developments in simulation research that we can hardly imagine.

References

Bailey, D. (1997). *When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs*. Unpublished UC Berkeley Ph.D. Thesis.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.

Bergen, B., Narayan, S., & Feldman, J. (2003). Embodied verbal semantics: Evidence from an image-verb matching task. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*.

Bergen, B., Chang, N., & Narayan, S. (2004). Simulated Action in an Embodied Construction Grammar. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*.

Bergen, B. & Chang, N. (2005). Embodied Construction Grammar in Simulation-Based Language Understanding. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive grounding and theoretical extensions*. John Benjamins.

Bergen, B., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science, 31*.

Bergen, B. & Wheeler, K. (2005). Sentence Understanding Engages Motor Processes. In *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*.

Brodal, P. (1998). *The central nervous system: structure and function*. New York: Oxford University Press.

Buccino, G, Binkofski, F., Fink, G., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R., Zilles, K., Rizzolatti, G., & Freund, H. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience, 13*(2), 400–404.

Chan, A. (ms). Embodied verbal semantics in English and Chinese: Evidence from a verb-image matching task.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge: MIT Press.

Craver-Lemley, C. & Arterberry, M. (2001). Imagery-induced interference on a visual detection task. *Spatial Vision, 14,* 101–119.

Denis, M. & Cocude, M. (1989). Scanning visual images generated from verbal descriptions. *European Journal of Cognitive Psychology, 1,* 293–307.

Ehrsson, H. H., Geyer, S., & Naito, E. (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part specific motor representations. *Journal of Neurophysiology, 90*, 3304–3316.

Feldman, J. & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language, 89*(2), 385–392.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119,* 593–609.

Gibbs, R. (2005). *Embodiment in Cognitive Science*. Cambridge University Press.

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

Glenberg, A. & Robertson, D. (2000). Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language, 43,* 379–401.

Gravetter, F. & Wallnau. L. (2006). *Statistics for the Behavioral Sciences* (7th ed.). Wadsworth Publishing.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*(2), 301–307.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press.

Kaschak, M. & Glenberg, A. (2000). Constructing Meaning: The Role of Affordances and Grammatical. Constructions in Sentence Comprehension. *Journal of Memory and Language, 43,* 508–529.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M. E., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, *94,* B79–B89.

Kosslyn, S., Ball, T., & Reiser, B. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 47–60.

Kosslyn, S., Ganis, G., & Thompson, W. (2001). Neural foundations of imagery. Nature Reviews, *Neuroscience, 2,* 635–642.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: Chicago University Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Landauer, T., Foltz, P., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25*, 259–284.

Langacker, R. W. (1986). Abstract motion. *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, 455–471.

Lindsay, S. (2003). *Visual priming of language comprehension.* Unpublished University of Sussex Master's Thesis.

Lotze, M., Montoya, P., Erb, M., Hülsmann, E., Flor, H., Klose, U., Birbaumer, N., & Grodd, W. (1999). Activation of cortical and cerebellar motor areas during executed and imagined hand movements: An fMRI study. *Journal of Cognitive Neuroscience, 11*(5), 491–501

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition, 32*, 1389–1400.

Mattes, S., Leuthold, H., & Ulrich, R. (2002). Stimulus-response compatibility in intensity-force relations. *Quarterly Journal of Experimental Psychology, 55A,* 1175–1191.

May, R. (1985). *Logical Form*. Cambridge: MIT Press.

Narayan, S., Bergen, B., & Weinberg, Z. (2004). Embodied Verbal Semantics: Evidence from a Lexical Matching Task. *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*.

Narayanan, S. (1997). *KARMA: Knowledge-based Action Representations for Metaphor and Aspect*. Unpublished UC Berkeley Ph.D. Thesis.

Nyberg, L., Petersson, K.-M., Nilsson, L.-G., Sandblom, J., Åberg, C., & Ingvar, M. (2001). Reactivation of motor brain areas during explicit memory for actions. *NeuroImage, 14*, 521–528.

Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology, 21,* 422–452.

Porro C., Francescato, M., Cettolo, V., Diamond, M., Baraldi, P., Zuian, C., Bazzocchi, M., & di Prampero, P. (1996). Primary motor andsensory cortex activation during motor performance and motor imagery: A functional magnetic resonance imaging study. *Journal of Neuroscience, 16,* 7688–7698.

Pulvermüller, F., Haerle, M., & Hummel, F. (2001). Walking or Talking? Behavioral and Neurophysiological Correlates of Action Verb Processing. *Brain and Language, 78,* 143–168.

Richardson, D., Spivey, M., Edelman, S., & Naples, A. (2001). "Language is spatial": Experimental evidence for image schemas of concrete and abstract spatial representations of verbs. *Proceedings of the Twenty-third Annual Meeting of the Cognitive Science Society* (pp. 873–878). Mawhah, NJ: Erlbaum.

Richardson, D., Spivey, M., McRae, K., & Barsalou, L. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27*, 767–780.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.

Romaiguère, P., Hasbroucq, T. Possamai, C., & Seal, J. (1993). Intensity to force translation: A new effect of stimulus-response compatibility revealed by analysis of response time and electromyographic activity of a prime mover. *Cognitive Brain Research, 1,* 197–201.

Segal, S. & Gordon, P. (1969). The Perky Effect revisited: Blocking of visual signals by imagery. *Perceptual and Motor Skills, 28,* 791–797.

Simon, J. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology, 81,* 174–176.

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

Talmy, L. (1996). Fictive motion in language and "ception". In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 211–276). Cambridge, MA: MIT Press.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273–281.

Tseng, M. & Bergen, B. (2005). Lexical Processing Drives Motor Simulation. In *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. Cambridge, Mass: MIT Press.

Wheeler, M., Petersen, S., & Buckner, R. (2000). Memory's echo: Vivid remembering reactivates sensory specific cortex. *Proceedings of the National Academy of Science of the U.S.A., 97,* 11125–11129.

Zwaan, R., Stanfield, R., & Yaxley, R. (2002). Do language comprehenders routinely represent the shapes of objects? *Psychological Science, 13,* 168–171.