Grounding Language in Bodily States: The Case for Emotion

Arthur M. Glenberg, David Havas, Raymond Becker, and Mike Rinck

It has become increasingly clear over the past few years that the symbols used by language become meaningful through grounding. For example, Glenberg and Kaschak (2002) demonstrated that some linguistic constructions are grounded in literal action, whereas Pecher, Zeelenberg, and Barsalou (2003) and Stanfield and Zwaan (2001) show how language is grounded in perceptual and imaginal states, respectively. In this chapter we report initial results demonstrating how language can also be grounded in the bodily states that comprise emotions. We begin by discussing the logical and theoretical reasons for supposing that language is grounded in bodily states, and then we move to a brief review of the recent work demonstrating grounding in action and perceptual states. This introductory material is followed by the report of two experiments consistent with the claim that language about emotional states is only fully understood when those states are literally embodied during comprehension.

Why Language Must be Grounded Outside the Linguistic System

To ask how a symbol is grounded is to ask how it becomes meaningful. One hypothesis is that linguistic symbols (e.g., words and syntactic constructions) become meaningful only when they are mapped to non-linguistic experiences such as actions and perceptions. A second hypothesis is that linguistic symbols can be grounded in other linguistic symbols. For example, words are often defined in terms of other words: That is exactly how a dictionary works. In contrast to what is intuitive about the dictionary view of grounding, Searle's (1980) Chinese Room Argument, and Harnad's (1990) symbolic merry-go-round argument are meant to provide the contrary intuition that word meaning cannot result from defining words solely in terms of other words. In Harnad's argument, he asks us to imagine traveling to a country where we do not speak the language (e.g., China), and our only resource is a dictionary written solely in that language. When we arrive, we are confronted with a sign comprised of linguistic characters. Attempting to comprehend the sign, we look up the first symbol in the dictionary to be confronted with a definition (definition 1) comprised of more symbols, none of which we understand. Undaunted, we look up the definition (definition 2) of the first symbol contained in definition 1, only to find that the symbols in definition 2 are also meaningless to us. We continue to look up the definitions of symbols used in other definitions, but none of definitions provide any help. Apparently, no matter how many symbols we look up, if they are only defined in terms of other symbols, the process will not generate any meaning. That is, if we are to learn the meaning of the Chinese symbols, those symbols must be grounded in something other than additional Chinese symbols.

Although Harnad's argument seems compelling, traditional and contemporary theories of meaning proposed by cognitive psychologists suggest otherwise - that meaning of undefined and arbitrary symbols arises from definitions that are themselves comprised of more undefined and arbitrary symbols.[1] For example, the Collins and Loftus theory (1975) proposes that conceptual information arises from the pattern of relations among nodes in a network. Here, every node corresponds to an un-defined word, and the set of nodes to which a particular node is connected corresponds to the words in the dictionary definition. Similarly, recently proposed theories based on the mathematics of high-dimensional spaces (e.g., Burgess & Lund, 1997; Landauer and Dumais, 1997) suggest that linguistic meaning can arise from defining linguistic symbols in terms of other linguistic symbols. For example, HAL (Burgess & Lund, 1997) is a computer program that combs the internet for linguistic stimuli. It creates a high-dimensional matrix with both rows and columns corresponding to the words encountered. The entries into the cell defined by the intersection of a particular row and column denote the frequency with which the words appear together in pieces of text (word pairs, triplets, quadruplets, and so on). According to Burgess and Lund, the meaning of a word is given by a vector created by the (approximately) 70,000 numbers in the row corresponding to a particular word combined with the (approximately) 70,000 numbers in the column corresponding to the same word. That is, they claim that finding enough other words with which a particular word co-occurs is sufficient to ground meaning.

In contrast to the dictionary-like theories of Collins and Loftus (1975), Burgess and Lund (1997) and many other (e.g., Anderson, Matessa, & Lebiere, 1997; Kintsch, 1988), there is recent theoretical (e.g., Barsalou, 1999; Glenberg, 1997) and empirical (Glenberg & Kaschak, 2002; Pecher, Zeelenberg, & Barsalou, 2003; Stanfield & Zwaan, 2001) work within the embodiment framework that has demonstrated how language is grounded in bodily states of action and perception. Using a property verification task, Pecher, Zeelenberg, and Barsalou (2003) observed a priming effect with a perceptual basis. Participants responded whether or not an object (e.g., a lawn mower) has a particular property (e.g., loud). Pecher et al. found that when the perceptual dimension probed on the next trial (e.g., leaves - rustle) was the same as the dimension probed on the previous trial, responding was faster than when the perceptual dimension probed on the previous trial was different (e.g., lawn mower - heavy). Apparently, understanding a concept presented linguistically calls on perceptual experience, not just arbitrary nodes.

Zwaan and his associates (e.g., Stanfield & Zwaan, 2001) asked participants to verify that a picture (e.g., of a pencil) depicted an object mentioned in a sentence (e.g., "The pencil is in a cup"). They found that pictures matching the implied orientation of the object (a pencil depicted vertically in this case) were responded to faster than pictures of the object in an orientation that mismatched orientation implied by the sentence. Thus, understanding the sentence appears to call on experience with real pencils and real cups and the orientations that they can take, rather that just the association of nodes representing pencils and cups.

As another example of grounding in bodily states, Glenberg and Kaschak (2002) asked each participant to judge the sensibility of sentences such as "You gave Andy the pizza" or "Andy gave you the pizza" by moving the hand from a start button to a Yes button. Location of the Yes button required a literal movement either toward the body or away from the body. Responding was faster when the hand movement was consistent with the action implied by the sentence (e.g., moving away from the body to indicate "Yes" for the sentence, "You gave Andy the pizza") than when the literal movement was inconsistent with that implied by the sentence. Apparently, understanding these action sentences called on the same neural and bodily states involved in real action. Thus, in contrast to theories that claim language symbols are grounded solely in other symbols, these results imply that understanding language calls on bodily states involved in perception, imagery, and action.

Grounding Language in Emotional States

There are strong connections between language and emotion. When reading or listening, we often find ourselves becoming sad, angry, afraid, happy, joyous, or aroused depending on the meaning of the language we are processing. In fact, it is likely that much of the pleasure we gain in reading and listening to narratives and poetry is directly related to an author's skill in producing those emotional states in us. Furthermore, there is formal evidence for the effects of emotions on language processing. For example, Haenggi, Gernsbacher, & Bolliger (1994) demonstrated that emotional inferences can be processed faster than spatial inferences. That is, participants read target sentences that matched the emotional state of a character faster than target sentences that matched spatial states. Participants were also faster to notice incongruities between emotional states than between spatial states.

Thus, common sense and research both demonstrate a connection between language and emotion. The embodied account of grounding makes a stronger claim, however: The full understanding of language about emotional states requires that those emotional states be simulated, or partially induced, using the same neural and bodily mechanisms recruited during emotional experiences. That is, language about emotions is grounded in emotional states of the body, and simulating those states is a prerequisite for full understanding of the language. The experiments presented next are our initial attempts to test this strong claim.

What might we expect, if the strong embodiment claim is correct? According to this claim, part of understanding emotional language is getting the body into the appropriate emotional state, because that is what gives the words their meaning. Consequently, if bodily systems are already in (or close to) those appropriate states, then understanding should be facilitated. And conversely, if the bodily systems are in inappropriate states, these states should interfere with language understanding. More concretely, if we are reading about pleasant events, we should be faster to understand those sentences if we are in a happy state than if we are in an unhappy state. Conversely, if we are reading about unpleasant events, we should be faster to understand those sentences if we are in an unhappy state than if we are in a happy state. Note that these predictions are based on two assumptions. The first is a dimensional assumption, namely that the bodily states corresponding to happiness are further away from those corresponding to unhappiness than to a neutral state. The second assumption is a type of inertia. Because the body is literally a physical and biological system that cannot change states instantaneously, it will take longer to shift from a happy state to an unhappy state (states that are far away) than to shift from a happy state to a neutral state.

To test these predictions, we need a means of reliably shifting the body into happy and unhappy states. Strack, Martin, and Stepper (1988) provide just such a methodology. They noted that holding a pen in one's mouth using only the teeth (and not the lips) forces a partial smile. In contrast, holding the pen using only the lips (and not the teeth) forces a partial frown. Note that having the face in a particular configuration is part of the bodily state corresponding to a particular emotion. Furthermore, Strack et al. demonstrated that these facial configurations differentially affected peoples felt emotions as well as their emotional assessment of stimuli. That is, participants rated cartoons as funnier when holding the pen in their teeth (and smiling) than when holding the pen in their lips (and frowning). This effect of facial configuration has been replicated numerous times (e.g., Berkowitz & Troccoli, 1990; Ohira & Kurono, 1993; Larsen, Kasimatis, & Frey 1992; Soussignan, 2002).

In our experiments, we use the Strack et. al. (1988) procedure to manipulate mood while participants read and understood sentences describing pleasant and unpleasant events. We measured how long it took to read and understand the sentences. On the basis of the strong embodiment claim, we predict an interaction between the pen condition (Teeth or Lips) and the valence of the sentence (Pleasant or Unpleasant). That is, participants should read and judge pleasant sentences more quickly while holding the pen in their teeth than when holding the pen in their lips, and the converse should be found for unpleasant sentences.

Examples of the sentences are given in Table 1. The 96 sentences (48 pleasant and 48 unpleasant) were based on those constructed and normed by Fischler et al., although we made changes in a small number of them.[2] Each of the original sentences was rated by approximately 60 participants on a scale from 1 (most unpleasant) to 9 (most pleasant). The mean (and standard deviation) for the unpleasant sentences was 2.92 (.66), and the mean for the pleasant sentences was 6.50 (.58).

Insert Table 1

In the first experiment, participants viewed a sentence on a computer screen by pressing the space bar on the computer keyboard (which also initiated a timer). The participant judged the valence of the sentence by pressing the "3" key on the keyboard, which was labeled with the letter "U," for unpleasant, or the "0" key, which was labeled with the letter "P," for pleasant. The left and right index fingers were used to make the Unpleasant and Pleasant responses, respectively. Half of the 96 participants began the experiment in the Teeth condition (holding the pen using only their teeth), and half began in the Lips condition (holding the pen using only their lips). During the experiment, participants switched between holding the pen in their teeth and lips every 12 sentences. In each of these blocks of 12 sentences, half of the sentences were Pleasant and half Unpleasant. Participants were told that the purpose of the pen manipulation was to examine the effects of interfering with the speech articulators during reading.

Overall, the participants were highly accurate in their judgments, agreeing with the normative classification 96% of the time (range of 87% to 100%). Because of two worries, we decided to analyze the data including a factor of experiment half (First half or Second half). One of these worries was that the pen manipulation would become onerous and unpleasant in both the Teeth and Lips conditions as the musculature fatigued. The other worry was that the task was so simple (judging the valence of the sentences) that people would learn to make the judgment on the basis of a quick scan of a few words rather than reading and understanding the whole sentence.

The reading (and judgment) times for the two halves of the experiment are presented in Table 2. The means were computed only for those sentences judged correctly. Also, reading times were eliminated if they were more than 2.5 standard deviations from a participant's mean in any particular condition.

Insert Table 2

The critical interaction between the Pen condition and Sentence valence was significant, $F(1,95) = 5.41$, $MSe = 80515$, $p = .02$. Although the triple interaction of Half of the experiment with Pen condition and Sentence condition was not significant ($p = .10$), it is clear from the data in Table 2 that the majority of the interaction comes from the first half of the experiment. Note that the form of this interaction is just what was predicted on the basis of the strong embodiment claim. Namely, reading of the pleasant sentences is 122 msec. faster when holding the pen in the teeth (and smiling) than when holding the pen in the lips (and frowning), whereas the reading of the unpleasant sentences is 45 msec. slower when holding the pen in the teeth than when holding the pen in the lips.

Although the data are consistent with the strong embodiment claim, there is a possible problem with the experimental procedure, namely the judgment draws the participant's attention to the valence of the sentences. Thus, a skeptic might argue that the bodily effects of the pen manipulation emerge only when people must explicitly judge emotional valence, and thus the effects are not reflective of basic understanding processes. To address this problem, we conducted a second experiment in which the task was to judge whether or not the sentences were easy or hard to understand. We told participants that we had written the sentences to be easy to understand, but that some difficult ones might have snuck through and that we needed their help in finding them. Thus, the great majority of their responses would be "Easy," but that there might be a few "Hard" responses. Note that participants were never informed that the sentences differed in valence. Furthermore, taking a cue from the social psychology literature, we asked participants after the experimental session if they had ever heard of the pen procedure or had suspected that we were attempting to manipulate their moods. We eliminated the data from four of the 42 participants who answered the latter question in the affirmative.

The 38 participants judged most of the sentences as easy to understand (94%, with a range of 60% to 100%). The reading time data are presented in Table 3. These data are based solely on sentences classified as "Easy," and reading times more that 2.5 standard deviations from the participant's mean for a particular condition have been eliminated.

Insert Table 3

Overall there was a significant interaction between the Pen condition and the Sentence valence, $F(1,37) = 6.63$, $MSe = 103382$, $p = .01$. However, the same interaction was significant for the judgments, $F(1,37) = 4.48$, $MSe = .003$, $p = .04$, indicating the possibility of some sort of judgment-by-reading speed tradeoff. Because the judgments seemed to have stabilized in the second half of the experiment, we also examined the data by halves of the experiment. Considering only the first half of the experiment, the critical interaction was significant for the judgments, $F(1,37) = 8.56$, $MSe = .002$, $p = .01$, but the interaction was not significant for the reading times, $p > .3$. The opposite pattern was obtained for the second half. Namely, the critical interaction was not significant for the judgments, $p > .3$, but was significant for the reading times, $F(1,37) = 4.89$, $MSe = 134316$, $p = .03$.[3] Thus, if we consider the data from both halves of the experiment or only the data from the second half, there is ample evidence for the critical interaction in the time needed to fully understand the sentence.

The data from the two experiments are clear and consistent: When reading and understanding sentences, judgments are facilitated when the suggested mood of the sentence is congruent with the mood induced by the pen manipulation. This finding is consistent with the predictions derived from the strong embodiment claim. As such, the data lend support to the general claim that language is grounded in bodily states, and the data lend support to the specific claim that language about emotions is grounded in emotional states literally produced by the body. Nonetheless, two issues give us pause.

The first issue concerns the prediction derived from the strong embodiment claim that relative to a neutral condition, both facilitation and interference should be observed. Our data might simply reflect facilitation (e.g., smiling speeds processing of pleasant sentences) but no interference for incongruent conditions (e.g., smiling may not interfere with processing of unpleasant sentences). Unfortunately, two pilot experiments designed to induce a neutral condition were failures. In the first pilot experiment, we implemented Strack et al.'s neutral condition of holding the pen in one hand rather than in the mouth. This necessitated a change in the response method from using the left and right index fingers to respond "Easy" and "Hard" respectively (as in Experiments 1 and 2), to using the right index finger to indicate "Hard" and the right middle finger to indicate "Easy." With this change, however, the critical interaction was no longer significant. We suspect that the interaction disappeared because responding with the index finger was actually easier than responding with the middle finger, leading to a conflict between the labels "Hard" and "Easy" and the difficulty of actually making the response. In the second pilot experiment, we used as a neutral condition holding the pen between the knees so that the two index fingers could once again be used to make the "Easy" and "Hard" responses. We obtained the critical interaction, but to our surprise, the pen-between-the-knees condition seemed to be a super-pleasant condition and not a neutral condition. That is, for the pleasant sentences, responding was fastest in the pen-between-the-knees condition, and for the unpleasant sentences, responding was slowest in this condition!

A second issue calling for caution is that the results portrayed in Tables 2 and 3 are consistent with other accounts in addition to the strong embodiment claim. Consider for example, Bower's (1981) theory of the relation between cognition and emotion. In Bower's theory, bodily states influence cognition by activating nodes (e.g., a happy node) that can spread activation to associated nodes representing happy and pleasant words such as, from Table 1, "proudly," "lover," and "embrace." Because these nodes are already activated, the corresponding words are read faster, leading to the critical interaction between bodily state and reading time. Note that on this account, bodily states have an effect on cognition (by activating the presumed "happy" node), but the language understanding itself results from the manipulation of arbitrary symbols. In contrast, the strong embodiment claim is that language understanding is directly grounded in bodily states, that is, that language understanding requires bodily states to derive meaning from the words. We are currently searching for experimental procedures to differentiate between these accounts.

In conclusion, our results add an important new finding consistent with the claim that language is grounded in bodily states. Previous work has demonstrated how language may be grounded in action, in perceptual states, and in images. These new experiments demonstrate how language about emotion-producing situations may well require a simulation or partial induction of those emotional states to be fully understood.

Author Note

This work was supported by a NSF grant to Arthur Glenberg. We are very grateful to Brianna Buntje, Sheila Simhan, Terina Yip, and Bryan Webster for fantastically efficient data collection and interesting discussion of these results. Requests for reprints may be directed to Arthur Glenberg at glenberg@wisc.edu.

Footnotes

[1]The symbols are arbitrary in the sense that there is no natural connection between the symbol and what it represents. For example, the word "chair" does not look, taste, feel, or act like a chair, and the word "eight" is not larger in any sense then the word "seven."

[2]We are very grateful to Ira Fischler who provided us with the stimulus sentences.

[3]Why are the effects most apparent in the first half of Experiment 1 and the second half of Experiment 2? We think that the task used in Experiment 1 (judging valence of the sentence) was very easy, and eventually participants learned to make those judgments by quickly scanning the sentences for key words, thus obviating(obfuscating, obscuring) the effect in the second half. The task used in Experiment 2 (judging if the sentence was easy or hard to understand) is apparently more difficult (note the slower times compared to Experiment 1). Importantly, there were no surface cues as to whether a sentence should be judged easy or hard, and hence the participant was forced to carefully consider each sentence throughout the experiment.

References

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction. Special Issue: Cognitive Architectures and Human-Computer Interaction*. 12(4), 439-462.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, 22, 577-660.

Berkowitz, L. & Troccoli, B. T. (1990). Feelings, direction of attention, and expressed evaluations of others. *Cognition and Emotion*, 4, 305-325.

Bower, G. H. (1981). Mood and memory. *American Psychologist*, 36, 129-148.

Burgess, C., & Lund, K. (1997). Modeling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, 177-210.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.

Glenberg, A. M. (1997). What memory is for. *Behavioral & Brain Sciences*, 20, 1-55.

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

Haenggi, D., Gernsbacher, M. A., & Bolliger, C. M. (1993). Individual differences in situation-based inferencing during narrative text comprehension. In H. van Oostendorp & R. A. Zwaan (Eds.), *Naturalistic text comprehension : Vol. LIII. Advances in discourse processing* (pp. 79-96). Norwood, NJ: Ablex.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 335-346.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, 95, 163-182.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representations of knowledge. *Psychological Review*, 104, 211-240.

Larsen, R. J., Kasimatis, M., & Frey, K. (1992). Facilitating the furrowed brow: An unobtrusive test of the facial feedback hypothesis applied to unpleasant affect. *Cognition & Emotion*, 6(5), 321-338.

Ohira, H., & Kurono, K. (1993). Facial feedback effects on impression formation. *Perceptual & Motor Skills*, 77(3, Pt 2), 1251-1258.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, 14, 119-124.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral & Brain Sciences*, 3, 417-457.

Soussignan, R. (2002). Duchenne smile, emotional experience, and autonomic reactivity: A test of the facial feedback hypothesis. *Emotion*, 2(1), 52-74.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 121, 153-156.

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating condition of facial expressions: A non-obtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54, 768-777.

Table 1

Examples of Sentences Used in Experiments 1 and 2

Pleasant sentences

The college president announces your name, and you proudly step onto the stage.

You and your lover embrace after a long separation.

Unpleasant sentences

The police car rapidly pulls up behind you, siren blaring.

Your supervisor frowns as he hands you the sealed envelope.

Table 2

Reading Times in milliseconds and Proportion Consistent Valence Judgments (in parentheses) from Experiment 1

First Half of the Experiment

	Sentence Valence	
Pen Condition	Pleasant	Unpleasant
Teeth (smiling)	2706 (.96)	2735 (.97)
Lips (frowning)	2828 (.97)	2690 (.97)

Second Half of the Experiment

	Sentence Valence	
Pen Condition	Pleasant	Unpleasant
Teeth (smiling)	2678 (.95)	2663 (.96)
Lips (frowning)	2661 (.96)	2623 (.97)

Table 3

Reading Times in milliseconds and Proportion Consistent Valence Judgments (in parentheses) from Experiment 3

First Half of the Experiment

	Sentence Valence	
Pen Condition	Pleasant	Unpleasant
Teeth (smiling)	3442 (.93)	3496 (.94)
Lips (frowning)	3435 (.94)	3372 (.91)

Second Half of the Experiment

	Sentence Valence	
Pen Condition	Pleasant	Unpleasant
Teeth (smiling)	3128 (.95)	3261 (.94)
Lips (frowning)	3256 (.95)	3127 (.97)