



The embodied statistician

Elizabeth R. Marsh^{1*} and Arthur M. Glenberg^{1,2}

¹ Laboratory for Embodied Cognition, Department of Psychology, Arizona State University, Tempe, AZ, USA

² Department of Psychology, University of Wisconsin – Madison, Madison, WI, USA

Edited by:

Anna M. Borghi, University of Bologna, Italy

Reviewed by:

Sian Beilock, University of Chicago, USA

Michael Spivey, University of California, USA

*Correspondence:

Elizabeth R. Marsh, Department of Psychology, Arizona State University, 950 S. McAllister Ave, Tempe, AZ 85287, USA.
e-mail: ermarsh@asu.edu

How do infants, children, and adults learn grammatical rules from the mere observation of grammatically structured sequences? We present an embodied hypothesis that (a) people covertly imitate stimuli; (b) imitation tunes the particular neuromuscular systems used in the imitation, facilitating transitions between the states corresponding to the successive grammatical stimuli; and (c) the discrimination between grammatical and ungrammatical stimuli is based on differential ease of imitation of the sequences. We report two experiments consistent with the embodied account of statistical learning. Experiment 1 demonstrates that sequences composed of stimuli imitated with different neuromuscular systems were more difficult to learn compared to sequences imitated within a single neuromuscular system. Experiment 2 provides further evidence by showing that selectively interfering with the tuned neuromuscular system while attempting to discriminate between grammatical and ungrammatical sequences disrupted performance only on sequences imitated by that particular neuromuscular system. Together these results are difficult for theories postulating that grammatical rule learning is based primarily on abstract statistics representing transition probabilities.

Keywords: statistical learning, implicit imitation, artificial grammar, embodied cognition, fluency

INTRODUCTION

We speak grammatical English without being able to say much about the rules of English grammar. How is that procedural knowledge learned, and where does it reside? One answer is given by the study of statistical learning, which is the ability to learn (often without intent) which sequences of stimuli are consistent with a set of rules. It has been observed in both infants and adults, in the visual, auditory, and tactile domains, and in a variety of stimulus displays from simple to complex to real-world scenes (Reber, 1967; Saffran et al., 1996, 1999; Fiser and Aslin, 2001, 2005; Saffran, 2001; Creel et al., 2004; Conway and Christiansen, 2005; Brady and Oliva, 2008). One explanation of the phenomenon is that it reflects a domain-general learning process: the brain is wired to pick up and represent statistics. Another (Conway and Christiansen, 2006) is that this learning is modality-constrained. We describe and test an embodied mechanism for statistical learning that is consistent with the notion of modality constraints but that also suggests why the phenomenon is robust enough to be classified as a general learning process.

Artificial grammar learning (AGL, e.g., Reber, 1967; Conway and Christiansen, 2005) is a paradigm case of statistical learning. Participants are exposed to sequences of stimuli such as tones, visual patterns, or tactile sensations. Attention to the stimuli is enforced by, for example, asking the participant to judge if two successive sequences are identical. After exposure to the sequences, participants are asked to discriminate between (a) sequences consistent with the rules used to generate the training sequences and (b) sequences generated by other rules. Although participants often profess no knowledge of the rules, they are successful in making the discrimination. How do they do this?

Our theory has three tenets. First, when people attend to stimuli, they concurrently imitate successive stimuli, often without awareness. Imitation of this sort is documented both behaviorally (e.g., Chartrand

and Bargh, 1999) and using neurophysiological measures (e.g., Rizzolatti and Craighero, 2004). Wilson (2001) presents a review of the literature supporting the claim that people reliably imitate, and she suggests a wide variety of functions of imitation.

Second, imitation is a neuromuscular phenomenon, that is, it requires cortical mechanisms that control particular effectors (e.g., the speech articulators, the hands, etc.) to generate the motor commands that would produce an imitation of the stimulus, even if the actual production is inhibited (e.g., Grush, 2004; Iacoboni, 2008). As elements of the sequence are imitated, the particular neuromuscular system used to produce the imitations is forced to make transitions between neuromuscular states. In this manner, the neuromuscular system becomes tuned to the environment (the grammar) so that some sequences, the grammatical ones, can be imitated with alacrity. This tuning is consistent with the literature on use-induced plasticity (e.g., Classen et al., 1998).

Third, when participants are asked to determine if a sequence is grammatical, they base their judgments on the implicit fluency with which they can imitate the sequence. Because of previous tuning, grammatical sequences are imitated more fluently than non-grammatical sequences. This account of statistical learning is similar in kind to the account of the mere exposure effect provided by Topolinski and Strack (2009), as well as the differences observed in the recognition memory of skilled versus unskilled typists for fluent letter dyads (Yang et al., 2009).

We tested this embodied account of statistical learning in two experiments. In the first experiment, we show that people have little difficulty discriminating between grammatical and ungrammatical sequences when the successive stimuli in a sequence can be imitated within the same neuromuscular system. However, when the same grammatical rules are instantiated by stimuli imitated using different neuromuscular systems, performance drops to chance levels.

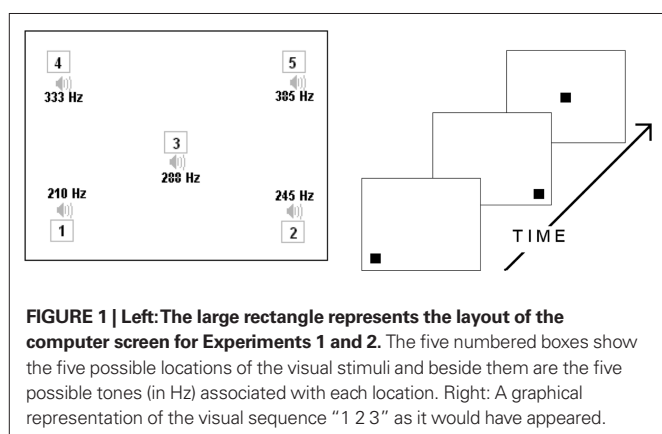
These results are consistent with the embodied account of statistical learning presented here, but they leave open the possibility of an alternative, attention-based account of domain-general statistical learning. In the second experiment, we directly test the importance of the domain-specific fluency gained from imitation of the training sequences by demonstrating that occupying neuromuscular system A disrupts the ability to discriminate between grammatical and ungrammatical sequences that are imitated using system A. Importantly, however, occupying system A has little effect on discrimination when the sequence is imitated using system B. The results suggest that the ability to access information in the form of domain-specific fluency is a large contributor to the subsequent recognition of grammatical sequences.

EXPERIMENT 1

We used a modified AGL paradigm similar to the one used by Conway and Christiansen (2005). During training, participants were exposed to pairs of grammatical sequences, and decided if the sequences within each pair were identical. During the test phase, participants were exposed to novel sequences. Half of the test sequences were generated by the grammar and half violated the grammar. Participants were asked to decide if each novel sequence was grammatical.

The key to the experiment is that participants received a single grammar expressed simultaneously in two modalities, auditory tones and visual boxes at different locations (Figure 1). During the test phase, however, they received only half of the information contained in the training sequences: auditory-only, visual-only, or alternating modalities. An alternating sequence begins with either a visual or auditory stimulus, and then it is followed by a stimulus in the other modality. This modality alternation continues until the end of the sequence.

Exposure to the dual modality sequences during training should engender imitation of the tones using the vocal folds (i.e., humming) and subsequent tuning of the laryngeal neuromuscular system. This tuning should allow easy discrimination between grammatical and ungrammatical single-modality auditory sequences at the test. Similarly, the dual modality sequences at training should engender tuning of the eye movement system to follow the sequence of spatial locations. This tuning should allow easy discrimination between grammatical and ungrammatical single-modality visual sequences at the test.



Consider, however, the alternating sequences. The alternating grammatical sequences instantiate exactly the same grammar used during training and used during the tests in the single-modality conditions. Thus, the alternating statistical regularities are exactly the same as the single-modality statistical regularities. If those regularities alone are the primary basis of performance, then performance should be the same as in the single-modality case.

On the embodied account, however, the imitation of sequences during training does not produce cross-modality tuning as readily as single-modality tuning. This is because the eye movement system tunes itself to produce sequences of eye movements, and the laryngeal system tunes itself to produce sequences of hummed tones. However, the eye movement system does not tune itself to produce a hummed tone, and vice versa. This is not to say that cross-modal tuning is impossible given the embodied account of statistical learning. Cross-modal tuning likely occurs on a regular basis in natural settings (e.g., the visual stimulus “D-O-G” may tune the speech articulators to produce the sound “dog”); this type of cross-domain mapping may only occur after considerable repetition relative to the sort of single-system tuning described here, which should arise out of biomechanical necessity.

To better understand why cross-modal tuning should not occur here, consider the neuromuscular systems used in imitating an alternating sequence. Starting with a stimulus 1 at a particular visual location, the eye movement system is tuned to move to other locations (for stimulus 2) that form a grammatical sequence. However, stimulus two is a tone. On imitating the tone, the laryngeal system is tuned to produce other tones that conform to stimulus 3, however stimulus 3 is a visual stimulus. Note that we left the eye movement system ready to move to the location of stimulus 2, but instead it must direct the eyes to the location of stimulus 3. This is an unpracticed transition that should be produced with less fluency than a practiced transition. Thus, when imitating the alternating test sequences, grammatical and ungrammatical sequences should be imitated with equal dysfluency, and discrimination performance should drop dramatically.

MATERIALS AND METHODS

This study was approved by the Arizona State University (ASU) IRB and was conducted in accordance with the approved standards. All participants gave informed consent. The participants were 20 Arizona State University undergraduates recruited from introductory psychology classes who received course credit. Four participants were excluded due to computer errors.

Visual stimuli were presented on 15” LCD monitors, and auditory stimuli were presented using headphones. Participants responded using the keyboard.

A modified version of the Gomez and Gerken (1999) grammar was used to create 12 grammatical sequences, ranging from 3 to 6 stimuli in length and composed of the numbers 1–5. Each of the numbers referred to one of five locations on the computer screen (each corner and the middle) at which a small black box was presented. In addition, each number referred to one of five tone frequencies (210, 245, 288, 333, and 385 Hz). The training sequences were constructed so that the box locations and tones were completely redundant with one another. Items within a single

sequence were displayed for 250 ms with a 150 ms inter-item interval; a sequence was separated from its mate within the training pair by a 1000 ms.

Half of the 24 novel test sequences conformed to the training grammar and half did not. The ungrammatical sequences contained the same beginning and ending stimuli as the grammatical sequences and differed only in internal transitions. Eight test sequences were used for each of the auditory, visual, and alternating tests. We counterbalanced both the order of the test-sequence modalities and the assignment of eight sequences to the modalities.

RESULTS

Proportions correct on the auditory and visual test sequences were above chance (0.50); Auditory-only sequences: $M = 0.625$, $SEM = 0.045$, $t(19) = 2.77$, $p\text{-rep} = 0.942$; visual-only sequences ($M = 0.625$, $SEM = 0.054$), $t(19) = 2.33$, $p\text{-rep} = 0.907$. However, performance on the alternating sequences was at chance ($M = 0.481$, $SEM = 0.032$), $t(19) = -0.590$, $p\text{-rep} = 0.456$. Tests of within-subjects contrasts demonstrated a significant difference between the auditory-only and the alternating sequences, $F(1,19) = 6.24$, $p\text{-rep} = 0.979$, and a significant difference between the visual-only and the alternating sequences, $F(1,19) = 7.33$, $p\text{-rep} = 0.940$.

These results are in accord with the predictions of the embodied account of statistical learning. There remains, however, an alternative explanation that is in keeping with a domain-general mechanism. It is possible that the learning of alternating sequences is constrained by attention to one modality at a time. Though it has been demonstrated that people can learn two separate grammars at one time when the stimuli comprising each grammar are perceptually distinct (Conway and Christiansen, 2006), the cost of switching attention between perceptually distinct stimuli within a single sequence may account for the apparent dysfluency for alternating sequences. Thus, the attention-switching account would claim that it is not necessarily the motor-fluency induced by modality-specific imitations that contributes to learning, but that attention is modality-specific. The resulting knowledge may still be domain-free statistics that can be applied to any single-modality sequences at the time of test (as this should produce no added demand on attentional resources). In this case, statistics abstracted from the auditory sequences alone should provide adequate recognition of grammatical sequences instantiated in the visual modality, and vice versa. However, the alternating sequences (and consequent attention-switching costs) disrupt the application of this domain-free statistical knowledge. Consequently, we tested another prediction of the embodied account.

EXPERIMENT 2

We used a selective interference task to test between the embodied account and the attention-switching account. The learning phase of this experiment was identical to that of Experiment 1. In the test phase, participants judged the grammaticality of both auditory and visual sequences while performing one of three secondary tasks: humming two notes requiring alternating changes in the laryngeal system (making the high-low sound of a siren), mouth sounds requiring alternating changes in the speech articulators (saying “da-da”), and feet sounds (alternating stomps with the left and

right feet). Sounds were demonstrated for the participants via the headphones to ensure that they understood the task. Furthermore, this demonstration ensured that participants produced the sounds with similar timing (i.e., one hummed note took approximately the same time to produce as one “da” and as one stomp).

The embodied stance predicts that humming should drastically interfere with imitation using the laryngeal system and thereby disrupt discrimination between grammatical and ungrammatical auditory sequences. The other two tasks should be far less disruptive for the auditory sequences (note that the da-da task requires transitions in the speech articulators but few transitions in the vocal folds). None of the tasks should interfere with discrimination of the visual sequences because they do not affect transitions of the eyes between successive stimulus locations.

The attention-switching hypothesis, however, suggests a number of possible outcomes. First, it might be the case that the presence of the secondary task will disrupt performance ubiquitously, since attention is being divided in all test conditions. Second, it might be the case that performance on visual sequences alone is affected, since attention is being divided between production of an auditory stimulus and perception of a visual sequence. Though it cannot be ruled out, this seems a slim possibility. More likely is the third possibility. That is, the production of auditory stimuli should interfere with the perception of auditory stimuli alone because attention is being divided within a single modality. Note, however, that this account provides no clear reason to assume any difference in the level of interference associated with the production of one sound versus another. The disruption of auditory sequences should be ubiquitous.

MATERIALS AND METHODS

This study was approved by the Arizona State University (ASU) IRB and was conducted in accordance with the approved standards. All participants provided informed consent. The participants were 69 ASU undergraduates and staff. Participants received \$10 in exchange for 1 h of participation. During the hour, participants completed this and an unrelated study. Three participants were excluded for not following instructions.

All materials were identical to those used in Experiment 1, as was the training phase procedure. In the testing phase, participants were exposed to 16 sequences in each modality while engaged in one of the three secondary tasks (manipulated between participants). The order of the modalities was counterbalanced.

RESULTS

Means are presented in **Table 1**.

The groups did not differ in regard to performance during training (all $ps > 0.25$). The major analysis was based on two, single df contrasts to test predictions derived from the embodied theory. The first contrast compared the difference between the auditory and visual test sequences during the siren task to the same difference during the da-da and stomping tasks. As predicted, the difference was larger for the siren task than for the others, $t(66) = 1.97$, $p\text{-rep} = 0.88$. The second contrast compared the visual-auditory difference for the da-da task to the difference for the stomping task, and as predicted this effect was not significant, $t(66) < 1$.

Table 1 | For each of the three secondary tasks, number correct during training (standard errors in parentheses), and proportion correct during the test phase.

Group	Training	Auditory test	Visual test
Siren (n=22)	69.32 (0.59)	0.59 (0.03)	0.70 (0.04)
Da-da (n=25)	67.76 (0.81)	0.66 (0.03)	0.66 (0.04)
Stomp (n=22)	68.77 (0.67)	0.66 (0.04)	0.66 (0.04)

These data challenge the attention-switching alternative in a number of ways. First, there is no clear reason why any of the interfering tasks should disrupt performance based on these abstract symbols unless attention was being divided at the time of test. On this account, however, there is no reason why the siren task should disrupt performance with the auditory sequences more than performance with the visual sequences. If the disruption was based on a masking effect of the sounds (unlikely given that headphones were used to present the stimuli) or the subdivision of attention within the auditory modality, then the da-da task should also have disrupted performance, but it did not.

In contrast, the results are in accord with the embodied account. Discriminating between grammatical and ungrammatical auditory sequences requires using the larynx to imitate the auditory stimuli. The siren task occupies that neuromuscular system and thereby reduces performance. Because the da-da task requires changes in the lips and tongue more than the larynx, it does not disrupt performance. Discriminating between grammatical and ungrammatical visual sequences requires using the eyes, possibly along with neck musculature, to imitate (or follow) the locations of the visual stimuli. None of the secondary tasks disrupt eye movements, and hence there is little disruption in discrimination performance.

Another reason for discounting the attention-switching account is that we have been able to demonstrate selective interference in the visual modality in the absence of any need for attention-switching. Marsh and Glenberg (2010) manipulated the participants' head orientation relative to the computer screen from training to test. In this experiment, participants either faced the screen directly or at a 45° angle during training. During test, this orientation was either preserved or switched to the alternative orientation. Participants who maintained the same head orientation (thus using the same eye movements during training and testing) performed normally on the visual and auditory test sequences. Participants who switched head orientation between training and test (thus using different, un-tuned eye movements during the test), however, performed poorly on the visual test sequences while performing well above chance on the auditory sequences. This finding provides further support for the generality of neuromuscular tuning account. Note that changing the orientation of the head to the screen does not require any attention-switching, so an attention-switching account would have difficulty predicting these results.

Thus, in Experiment 2, selectively interfering with the laryngeal system reduces discrimination between grammatical and ungrammatical auditory stimuli, but not visual stimuli. And, as found in Marsh and Glenberg (2010), selectively interfering with the eye movement control system reduces discrimination between grammatical and ungrammatical visual stimuli, but not auditory stimuli.

Although one might postulate special attentional demands within each of these systems (e.g., the siren task disrupts auditory attention but the da-da task does not), such an account would be highly unparsimonious to say the least.

DISCUSSION

Can our simple embodied account accommodate all of the facts of statistical learning? Probably not. For instance, the last parts of an auditory sequence are learned best, compared to the first parts of a tactile sequence (Conway and Christiansen, 2005). This is not predicted by our account; but then neither is it by the analytic process account.

Nonetheless, embodiment offers a compelling account of a good proportion of the literature as well as novel predictions. For example, it predicts the learning of non-adjacent dependencies (e.g., between the first and third stimulus), will be facilitated if the intervening stimulus is imitated with a different neuromuscular system. Suppose that the first stimulus is tonal, the second visual, and the third tonal. Imitation of first stimulus leaves the laryngeal system in a particular state that is not affected by imitation of the visual stimulus. Then, imitation of the third stimulus produces a transition from the state of the laryngeal system after humming the first stimulus to the state of the system used to hum the third. In fact, Gebhart et al. (2009) report that the learning of remote dependencies only occurs when intervening stimuli are dissimilar to their flankers.

The embodied account also explains how Conway and Christiansen (2006) participants could manage to learn two different grammars when sequences from the grammars were interleaved and presented in different modalities. In Conway and Christiansen, learning tuned two neuromuscular systems, and testing required tuned transitions within each of the systems. In contrast, in Experiment 1, learning tuned two systems, but testing required un-tuned transitions across systems.

Finally, Saffran et al. (2008) demonstrated that cotton-top tamarin monkeys can learn simple grammars instantiated by five spoken syllables, but not a more complex grammar instantiated by eight syllables. Our account predicts that if the complex grammar were instantiated in a neuromuscular system the tamarins can use more successfully for imitation (e.g., a reaching system), then the tamarins would have greater success in learning the complex grammar.

In the first paragraph of the introduction, we noted that the embodied mechanism for statistical learning is consistent with the notion of modality constraints, but that it also suggests why the phenomenon is robust enough to be classified as a general learning process. We think that neuromuscular tuning is likely to be found in all neuromuscular systems. Thus, one should be able to find statistical learning in for example, eye movement control, speech articulation, humming, finger movements, and so on. Thus, the learning mechanism (neuromuscular tuning) is found across many domains, and at the same time it is highly embodied.

Few native speakers can explain the rules with which grammarians describe their behavior. The present study suggests that such rules are embodied in differentially conditioned neuromuscular networks; the body has rules that its mind does not, and need not, know.

ACKNOWLEDGMENTS

We thank Andre Valdez for discussions about testing the embodied hypothesis, and Jenny Saffran and Peter Killeen for their insightful comments. We also thank Mathew Dexheimer and Frederick Raehl for assistance in data collection. Work on this article was supported

by the National Science Foundation under Grant No. BCS 0744105 to Arthur Glenberg, and an NSF IGERT award to Elizabeth Marsh. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- Brady, T. F., and Oliva, A. (2008). Statistical learning using real world scenes: extracting categorical regularities without conscious intent. *Psychol. Sci.* 19, 678–684.
- Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910.
- Classen, J., Liepert, J., Wise, S. P., Hallett, M., and Cohen, L. G. (1998). Rapid plasticity of human cortical movement representation induced by practice. *J. Neurophysiol.* 79, 1117–1123.
- Conway, C. M., and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 24–39.
- Conway, C. M., and Christiansen, M. H. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychol. Sci.* 17, 905–912.
- Creel, S. C., Newport, E. L., and Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1119–1130.
- Fiser, J., and Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.* 12, 499–504.
- Fiser, J., and Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *J. Exp. Psychol. Gen.* 4, 521–537.
- Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* 16, 486–490.
- Gomez, R. L., and Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70, 109–135.
- Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–442.
- Iacoboni, M. (2008). The role of premotor cortex in speech perception: evidence from fMRI and rTMS. *J. Physiol. (Paris)* 102, 31–34.
- Marsh, E. R., and Glenberg, A. M. (2010). “Head orientation affects visual statistical learning,” in *Poster Presented at the Annual Meeting of the Psychonomics Society*, St. Louis, MO.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* 6, 855–863.
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
- Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition* 81, 149–169.
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.
- Saffran, J. R., Hauser, M., Seibel, R., Kpfhamer, J., Tsao, F., and Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* 107, 479–500.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.
- Topolinski, S., and Strack, F. (2009). Motormouth: mere exposure effects depend on stimulus-specific motor-simulations. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 423–433.
- Wilson, M. (2001). Perceiving imitable stimuli: consequences of isomorphism between input and output. *Psychol. Bull.* 127, 543–553.
- Yang, S., Gallo, D. A., and Beilock, S. L. (2009). Embodied memory judgments: a case of motor fluency. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1359–1365.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 July 2010; paper pending published: 20 August 2010; accepted: 10 October 2010; published online: 16 November 2010.

Citation: Marsh ER and Glenberg AM (2010) The embodied statistician. *Front. Psychology* 1:184. doi: 10.3389/fpsyg.2010.00184

This article was submitted to *Frontiers in Cognition*, a specialty of *Frontiers in Psychology*.

Copyright © 2010 Marsh and Glenberg. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.