COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 39 (2015) 1987–1995

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12295

# Putting Everything in Context

Lauren A. M. Lebois,[a] Christine D. Wilson-Mendenhall,[b] Lawrence W. Barsalou[a]

[a]*Department of Psychology, Emory University*
[b]*Department of Psychology, Northeastern University*

Received 30 June 2015; accepted 3 July 2015

**Abstract**

In response to Casasanto, Brookshire, and Ivry (2015), we address four points: First, we engaged in conceptual replications of Brookshire, Casasanto, and Ivry (2010), not direct replications. Second, we did not question the validity of Brookshire et al.'s (2010) results, nor the similar findings of other researchers, but instead explained divergent findings within an integrated theoretical framework. Third, challenges to the construct of automaticity, including ours, were widespread, long before Brookshire et al.'s (2010) article. Fourth, the planned comparisons that we reported tested our theoretical claims and offered strong evidence for them.

## 1. The distinction between direct and conceptual replications

Replicating previous research is of much current interest (e.g., Schmidt, 2009; Zwaan & Pecher, 2012). Within the replication literature, a fundamental distinction exists between *direct replications* and *conceptual replications* (with direct replications sometimes referred to as "exact replications," although exact replications are never possible; Schmidt, 2009). In a direct replication, researchers perform an experiment that uses "the same stimuli, the same instructions, the same procedure and similar participants to approximate an exact replication" (Zwaan & Pecher, 2012, p. 2). In a conceptual replication, researchers attempt to "verify the underlying hypothesis of an earlier experiment" often by using "a different experimental set up that conveys the same primary information ... by a radically different material realization" (Schmidt, 2009, p. 94). The purpose of a conceptual replication in many articles is to establish common ground with previous articles, first establishing the presence of a robust effect under at least somewhat different conditions, before extending the relevant findings significantly further, both empirically and theoretically.

Correspondence should be sent to Lawrence W. Barsalou, Department of Psychology, Emory University, Atlanta, GA 30322. E-mail: barsalou@emory.edu

Casasanto et al. object to our use of "replication" to describe conditions in our experiments where we attempted to replicate the trial format in Brookshire, Casasanto, and Ivry's (2010) experiments. Specifically, Casasanto et al. state, "Although Lebois et al. adapted our method, they performed a different experiment, using a different kind of stimulus." We just saw, however, that this is exactly how Schmidt (2009) defined a conceptual replication. As Schmidt (2009, p. 91) further noted, "much of the confusion in the existing literature results from mixing up these two basic categories."

If we had been attempting to directly replicate one of Brookshire et al.'s experiments, there would have been much discussion about our attempts to reproduce the earlier experiment as exactly as possible. Instead, we made it clear that we were not attempting to perform a direct replication. When we first introduced our paradigm in the introduction (p. 1774), we referenced Casasanto (2008) and Brookshire et al. (2010) because they had demonstrated that, under some conditions, meaning becomes active without contextual support from instructions and orienting tasks to produce congruency effects. Specifically, we stated, "Across three experiments, we explored two manipulations designed to assess the automaticity of grounded information" using "a paradigm shown recently to produce grounded congruency effects" (i.e., the Brookshire et al. paradigm). When we first mentioned the Brookshire et al. replication condition in Experiment 2 (p. 1783), we stated that we wanted to see whether "replicating their trial format would again produce congruency effects . . . with no assistance from instructions or orienting tasks." In the same paragraph, we stated that we were using different materials than Brookshire et al. (concrete words instead of abstract words), given the theoretical assumption that concrete words would be more likely to show congruency effects. Explicitly mentioning our use of different materials at this point made it clear that we were not attempting a direct replication of Brookshire et al., but were simply performing a conceptual one. Indeed, in two review cycles at different journals, none of the reviewers or action editors questioned our use of "replication," with one of the signed reviews coming from an expert in the replication literature.

Finally, in the Discussion section (pp. 1791–1792), we reviewed significant differences in procedures and materials between Brookshire et al. (2010) and our experiments, offering explanations for why different results might have resulted. Again, it was clear that we had only attempted conceptual replications, not direct ones.

## 2. Integrating conflicting findings within a common theoretical framework

Contrary to Casasanto, Brookshire, and Ivry's (2015) claim, we did not question the validity of Brookshire et al.'s (2010) results, nor the similar findings of other researchers. Although we were not able to obtain congruency effects without contextual support, other laboratories have been able to do so. In a section entitled, "Reconciling Conflicting Results" (pp. 1791–1792), our central theme was that all of these results are valid within a unifying theoretical framework. If the construction of meaning is dynamical and context-dependent, then the specific conditions in an experiment determine how meaning is

constructed within it. As a consequence, the results obtained across experiments in a paradigm are likely to vary. The theoretical framework of our article not only explains conflicting results in the literature, but also predicts that they should be prevalent, especially when effects are small. In this spirit, we addressed methodological factors that might have contributed to differences between experiments on grounded congruency effects, including differences in procedures and materials (pp. 1791–1792). Rather than questioning the validity of conflicting results, we attempted to explain them within a common framework.

Similarly, in our introduction (pp. 1766–1770), we reviewed diverse phenomena in the automaticity literature, illustrating how these phenomena, too, are generally context-dependent (e.g., Stroop, SNARC, and Simon effects). Like grounded congruency effects, these phenomena also vary as a function of methodological factors. Rather than such variability being a problem (e.g., when one of these phenomena doesn't "replicate"), we and others have suggested that the dynamic nature of these phenomena makes them sensitive to contextual factors.

Within this framework, "failures to replicate" appear in a new light. Rather than reflecting the failure of one lab to directly replicate the results of another, different findings may reflect important systematic differences in the conditions that cause a phenomenon to wax and wane. Our treatment of variability across laboratories in obtaining grounded congruency effects clearly reflected this perspective. Rather than questioning the validity of previous findings, we viewed their variability as strong evidence for our theoretical position. To the extent that the access of meaning is context-dependent, variability should occur in congruency effects, and it does.

## 3. Challenges to the construct of automaticity were widespread and long coming

Casasanto et al. (2015) claim that we did not accurately represent their theoretical position in Brookshire et al. (2010). Specifically, Casasanto et al. stated, "Lebois et al. did not accurately represent our theoretical position on the context-dependence of spatial congruity effects, *which anticipated theirs*" (p. 1980, italics ours). Casasanto et al. further stated that Lebois et al. "used similar methods and arrived at *strikingly similar conclusions*" (p. 1980, italics ours). Finally, Casasanto et al. pointed out that Barsalou (1982) once claimed that some features are automatically activated independent of context (Barsalou, 1982), even though Lebois, Wilson-Mendenhall, and Barsalou (2015) acknowledged this in their opening paragraph. Given these points, we believe that some history on automaticity research, including our own, is in order.

Challenges to the construct of automaticity have been in the wind for some time, long before Brookshire et al.'s (2010) article. As Lebois et al. (2015, pp. 1766–1773) reviewed, many researchers in the 1990s, and especially in the 2000s, questioned the construct of automaticity and performed extensive research to assess this issue. Researchers studying the Stroop effect, SNARC effect, semantic priming, affective priming, attentional cuing, and the Simon task have been questioning whether these effects are

automatic for the past two decades. Thus, Brookshire et al. (2010) did not break new ground in challenging the construct of automaticity.

As Lebois et al.'s (2015) literature review further mentioned, several other major reviews have documented the widespread deconstruction of automaticity, including Santiago, Roman, and Ouellet (2011), Kiefer, Adams, and Zovko (2012), Tomasino and Rumiati (2013), and Gawronski and Cesario (2013). Like ours, each of these reviews drew on articles going back decades to demonstrate that diverse phenomena often assumed to be automatic are actually context-dependent. As these reviews illustrate, the case against automaticity has been developing for some time across literatures.

We first began to question the automatic activation of semantic features when Aron Barbey, a Ph.D. student in our laboratory, failed to obtain grounded congruency effects in his dissertation (Barbey, 2007). Across a series of experiments beginning in 2006, Barbey assessed whether conceptual combination utilizes simulations of central semantic features, such as verticality. Because he found little evidence for such simulations, we began to question whether simulations of central semantic features are activated automatically.

Notably, Barbey's experiments had a relatively high ratio of filler trials to target trials (75% to 25%, respectively). It occurred to us that if the activation of meaning simulations is not automatic, then high proportions of filler trials would generate a strategic set that worked against activating these simulations. Since Posner, Snyder, and Davidson's (1980) classic work, researchers have argued that an automatic process is one that is not affected by set (e.g., whether the proportion of filler trials is high or low). Conversely, when set does affect a process, the process is likely to be under strategic control and not be automatic.

To test this possibility, we designed a version of Barbey's experiments, using his materials that manipulated the proportion of filler trials (0% vs. 80%). This work was begun by Erica Sarti in 2007, immediately following Barbey's dissertation, and was completed in 2008 as her honors thesis (Sarti, 2008), under the supervision of Wilson-Mendenhall and Barsalou (Wilson-Mendenhall was a graduate student at the time, who also overlapped with Barbey). A congruency effect only occurred with no fillers in the first half of the experiment on the initial stimulus presentations. No congruency effect ever occurred when fillers were present, nor for subsequent stimulus presentations in the second half of the experiment without fillers. On the basis of these results, we became even more convinced that the activation of central semantic information is not automatic.

When Lauren (McDonough) Lebois joined the laboratory in 2008, her masters thesis immediately followed up on Sarti's work (McDonough, 2010). McDonough, like Sarti (2008), again manipulated the proportion of filler trials (20% vs. 80%), while also assessing task set. Again, congruency effects failed to occur consistently. Based on the results of all our experiments to that point, it appeared that drawing attention to relevant task information was required for congruency effects to occur. To explore this possibility, we increasingly assessed whether contextual manipulations such as instructions and orienting tasks produce these effects. The result was the theoretical perspective and experiments in Lebois et al. (2015), both highly consistent with our research just described, and also with much of our other previous and current research (e.g., Barsalou, 1987, 2003, in

press; Barsalou, Wilson, & Hasenkamp, 2010; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011; Yeh & Barsalou, 2006).

This history illustrates that our laboratory has been exploring the issue of automaticity in semantic processing since 2006, beginning to seriously question whether semantic activation is automatic in 2007. Our interest in this area did not originate in Brookshire et al.'s (2010) work but began years before. Indeed, Sarti's (2008) findings predated Brookshire et al.'s (2010) similar findings by 2 years, showing, as did Brookshire et al., that grounded congruency effects disappear after the first presentation of a critical stimulus and that they are affected by filler trials.

As this brief review illustrates, we like many other researchers, including Brookshire et al., were swept up in a larger movement that has been reassessing the construct of automaticity (also see Moors & De Houwer, 2006). Our primary interest in Brookshire et al.'s work was adopting a task developed earlier by Casasanto that appeared to produce robust congruency effects without supporting instructions or orienting tasks. Because our theoretical inspiration came from elsewhere, it did not occur to us that we needed to say anything about Brookshire et al.'s theoretical position, and we did not.

## 4. The distinction between planned comparisons and omnibus tests

In our article, we adopted an approach to inferential statistics that focuses on a priori theory-motivated tests of hypotheses (e.g., Abelson & Prentice, 1997; Keppel, 1991). Many researchers fail to distinguish between a priori and post hoc tests in their statistical analyses, opting to use conservative post hoc omnibus tests even when they have clear hypotheses. When, for example, researchers predict that one condition will have a higher mean than another, they often assess this prediction with an omnibus $F$ test or a two-tailed $t$ test, establishing whether a difference between the conditions exists in *either* direction, rather than in only the direction predicted. Similarly, when researchers predict that an interaction will take a particular form, they nevertheless use an omnibus test that assesses whether any one of many possible forms of the interaction occurred. Statisticians often argue, however, that such predictions should be tested with directional planned comparisons that only test the predicted patterns of results. Once a priori hypotheses have been assessed, omnibus tests can then be used for exploratory purposes.

We adopted the planned comparison perspective in the three experiments that we reported. Specifically we designed planned comparisons to test theoretically motivated hypotheses consistent with our previous research going back to 2006: Congruency effects should occur when contextual support from instructions or orienting tasks is present, but they should not occur when contextual support is absent. Specifically, to test whether a congruency effect occurred when contextual support was present, we used a directional linear contrast to determine whether congruent trials were faster than incongruent trials. To test whether a congruency effect did not occur when contextual support was absent, we first predicted that the above linear contrast would not be significant, and then further predicted that a Bayesian test would offer strong support for the null hypothesis. Because

our a priori predictions concerned the presence versus absence of congruency effects in specific conditions, this was how we assessed our data. In every condition where we predicted a directional congruency effect, one occurred. In five conditions where we predicted a null effect, strong evidence for a null effect was present. Thus, the statistical tests that we developed a priori to assess our hypotheses provided supporting evidence for them.

Casasanto et al. (2015) argue that we failed to provide evidence that context modulates congruency effects. Specifically, they note that the omnibus interactions between group and congruency in our Experiments 1 and 2 were not significant. Although we showed that predicted effects and lack of effects within particular groups did and did not occur as predicted, we did not actually show, via significant interactions, that context (i.e., groups) modulated congruency.

Essentially Casasanto et al. have reframed our hypotheses for us, arguing that we should have tested something a priori that we did not set out to test. Specifically, they suggest that we test the hypothesis that context modulates congruency effects. In our Experiment 2, for example, their hypothesis is that whether a relevant orienting task is present or not for a given condition (context) should modulate the size of congruency effects observed across conditions. Notably, however, this hypothesis differs considerably from the hypotheses that we set out to test. Specifically, we predicted that a significant congruency effect would occur in each condition that drew attention to verticality information with an orienting task. Conversely, we further predicted that no significant congruency effect would occur in any condition that did not. If we had simply shown that context modulates congruency using an omnibus interaction, as Casasanto et al. (2015) suggest, it would not have been clear that a significant congruency effect occurred in any of the predicted conditions. Nor would it have been clear that a congruency effect was absent in any of the other conditions. All that would have been required to support Casasanto et al.'s hypothesis was showing that context modulated congruency effects in an overall interaction, not that specific significant congruency effects did or did not occur. Because we specifically predicted the presence versus absence of congruency effects, this is what we tested, and this is what our analyses demonstrated. These predictions were stronger than simply showing that context modulated congruency.

Thus, in earlier versions of our manuscript prior to the published version, we *only* reported assessments of congruency within particular conditions. We did not perform or report analyses that assessed patterns across conditions, such as interactions. Across two independent review cycles that included statistical experts who signed their reviews, not a single one suggested that we use interactions to test our hypotheses.

Furthermore, the omnibus tests of the interactions that we reported in the published version were unduly conservative, given that they tested for any form that the respective interaction could have taken (as in a two-tailed test). Thus, it is not surprising that these interactions were not significant, even though our planned comparisons were. Because we did not plan to use interactions for testing our a priori hypotheses, we did not power our experiments so that interactions would be significant. If we had, say, increased the number of participants in each experiment by 50%, even the omnibus interactions might well

have been significant, given that the means were in the predicted direction, and the *p* values were approaching significance.

We only included results from omnibus interactions in the published version of our article because of a request late during the review process to perform mixed analyses that included random effects for both participants and words. Once we shifted to this way of analyzing our results, we continued to focus on congruency effects within conditions, as in earlier versions of our manuscript, given that they tested our a priori hypotheses. We later added the subsequent sections on fixed and random effects that reported omnibus effects to provide a thorough listing of all inferential tests resulting from our new analyses.[1]

## 5. Conclusions

Casasanto et al. (2015) ended their commentary with regret that we did not coordinate our experiments with them. We find this quite remarkable, given that such coordination often only occurs—and indeed is only expected—when one research group attempts a direct replication of another group's finding. When a group simply attempts a conceptual replication, coordination typically does not occur and is certainly not expected. As mentioned earlier, conceptual replication is ubiquitous, providing an important stepping stone from one project to the next, thereby making coordination relatively unusual. And again, we never questioned the validity of Brookshire et al.'s (2010) results, nor their theoretical position. We continue to believe, and hope you agree, that putting everything in context is generally a good idea.

## Acknowledgments

We are grateful to Christopher Hertzog, Teenie Matlock, and Esther Papies for discussion.

## Note

1. Implications of the issues addressed in this section become clear when reconsidering the drug treatment analogy that Casasanto et al. (2015) applied to our analyses. Imagine wanting to show that a drug has an effect on a disease across a pre-test and post-test relative to a control group. Obtaining a significant omnibus interaction between treatment versus control group and pre-test versus post-test simply indicates that one variable modulated the effect of the other. A significant omnibus interaction does not demonstrate, however, that administering the drug actually had a significant effect within the treatment group, nor that there was no change over time within the control group. The omnibus interaction could be significant without these more critical findings being present (e.g., there might be no significant treatment effect and/or significant change over time could occur in the control condition). For these reasons,

it is essential to assess whether individual change over time is significant within the treatment group and not significant within the control group. Notably, Brookshire et al. (2010) focused their analyses on omnibus interactions and did not probe their interactions to establish the simple effects significant within them. Thus, it is perhaps not surprising that Casasanto et al.'s (2015) statistical priorities differ from ours.

## References


Abelson, R. P., & Prentice, D. A. (1997). Contrast tests of interaction hypothesis. *Psychological Methods*, *2*, 315.

Barbey, A. (2007). Assessing modality-specific systems in conceptual combination. Doctoral dissertation, Emory University. Available at http://web.library.emory.edu/. Accessed September 10, 2015.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*, 82–93.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge, UK: Cambridge University Press.

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*, 513–562.

Barsalou, L. W. (in press). Situated conceptualization: Theory and application. In Y. Coello & M. H. Fischer (Eds.), *Foundations of embodied cognition*. East Sussex, UK: Psychology Press.

Barsalou, L. W., Wilson, C. D., & Hasenkamp, W. (2010). On the vices of nominalization and the virtues of contextualizing. In B. Mesquita, L. Feldman Barrett, & E. Smith (Eds.), *The mind in context* (pp. 334–360). New York: Guilford Press.

Brookshire, G., Casasanto, D., & Ivry, R. (2010). Modulation of motor-meaning congruity effects for valenced words. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1940–1945). Austin, TX: Cognitive Science Society.

Casasanto, D. (2008). Body-specific abstract concepts: mental metaphors without linguistic metaphors. Paper presented at the International Conference on Language, Communication, and Cognition, Brighton, UK.

Casasanto, D., Brookshire, G., & Ivry, R. (2015). Meaning is not a reflex: Context-dependence of spatial congruity effects. *Cognitive Science*, *39*, 1979–1986. doi: 10.1111/cogs.12294

Gawronski, B., & Cesario, J. (2013). Of mice and men: What animal research can tell us about context effects on automatic responses in humans. *Personality and Social Psychology Review*, *17*, 187–215.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Kiefer, M., Adams, S. C., & Zovko, M. (2012). Attentional sensitization of unconscious visual processing: Top-down influences on masked priming. *Advances in Cognitive Psychology*, *8*, 50–61.

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*, 1764–1801. doi: 10.1111/cogs.12174.

McDonough, L. A. (2010). Grounded congruency effects: Automaticity and "strategery" in cognition. Masters Thesis, Emory University. Available at http://web.library.emory.edu/. Accessed September 10, 2015.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*, 297.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*, 160–174.

Santiago, J., Roman, A., & Ouellet, M. (2011). Flexible foundations of abstract thought: A review and a theory. In T. Schubert & A. Mass (Eds.), *Spatial dimensions of social thought* (pp. 39–108). Berlin: Mouton de Gruyter.

Sarti, E. L. (2008). The effect of fillers in a conceptual height study. Honors Thesis, Emory University. Available at http://web.library.emory.edu/. Accessed September 10, 2015.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*, 90–100.

Tomasino, B., & Rumiati, R. I. (2013). At the mercy of strategies: The role of motor representations in language understanding. *Frontiers in Psychology*, *4*, 27, 1–13.

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia*, *49*, 1105–1127.

Yeh, W., & Barsalou, L. W. (2006). The situated nature of concepts. *American Journal of Psychology*, *119*, 349–384.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS ONE*, *7*, e51382.