Working-memory capacity and island effects: A reminder of the issues and the facts

JON SPROUSE

MATT WAGERS

COLIN PHILLIPS

University of California, Irvine

University of California, Santa Cruz

University of Maryland, College Park

1. UNDERSTANDING THE TERMS OF THE DEBATE. In their response to our article (Sprouse et al. 2012), Hofmeister, Staum Casasanto, and Sag (2012, henceforth HSS) level three primary criticisms at our studies: (i) that working-memory (WM) correlations are not predicted by reductionist theories, (ii) that the WM tasks that we chose are inappropriate for showing the predicted correlation, and (iii) that our results do, in fact, show the predicted correlation. We note that these three criticisms appear to be contradictory—if any one were true, the other two would very likely be false—but here we focus on an attempt to clarify why all of them are unjustified: (i) the WM correlation is exactly what a viable account of the island effects predicts; (ii) the diversity of noncorrelations that have now been documented forces the reductionist account to abandon independently motivated memory mechanisms in favor of currently undefined 'mystery' mechanisms; (iii) HSS's characterization of our results is selective and misleading, and it inaccurately describes the predictions of the reductionist account of island effects. But before addressing these specific arguments it is worthwhile to briefly remind the reader of the central question of this debate. This is already discussed at length in our original article, but a number of the issues raised by HSS appear to be orthogonal to the central question.

ISLAND EFFECTS are the sharp declines in acceptability that are found when WH-dependencies and similar linguistic dependencies span certain structural domains, such as relative clauses, interrogative clauses, complex subjects, adjoined phrases, and many others. These structural domains are commonly known as syntactic islands. The empirical fact to be explained is uncontroversial: extraction from island structures is much less acceptable than would be predicted by the simple summation of the acceptability cost of (i) WH-extraction and (ii) the presence of island structures. We refer to this as the SUPERADDITIVE property of island effects, and the focus of the debate is on the source of this property. A long-standing view is that the superadditivity reflects a linguistic constraint that specifically targets the combination of extraction and islands (i.e. island violations). A huge literature has arisen from this starting point. An equally long-standing competing view is that island effects reflect limits on language processing resources. According to this REDUCTIONIST view, island effects can be explained without appeal to domain-specific linguistic constraints. The superadditive property of island effects cannot, of course, be explained by simply summing the processing cost of extraction and island structures, so the simplest reductionist view is a nonstarter. But a variant of this reductionist hypothesis is more promising. Perhaps the superadditivity is a consequence of resource overload: to the extent that simultaneously processing extraction and an island structure exceeds the available memory resources, the extra acceptability cost (i.e. the superadditivity) might be explained as the cost for exceeding the available memory resources (Kluender & Kutas 1993). This account makes a straightforward prediction: the degree of superadditivity should vary as a function of the WM resources available to an individual. We tested this prediction in two experiments with over three hundred speakers, and the results were not encouraging for the reductionist approach: individual differences in WM capacity accounted for an average of only around 1% of the variance

in the size of superadditive effects, and in many cases the correlations were zero or the opposite of what the reductionist approach predicts. Similar conclusions have been reached in studies by Tokimoto (2009) and Michel (2012).

Crucially, the central question of this debate is whether linguistic constraints are needed in order to account for island effects, specifically the superadditive effect of combining extraction and island structures. The reductionist account of a given island constraint succeeds only if it eliminates the need for linguistic constraints entirely. What this means is that demonstrations that island-violating sentences are more or less difficult to understand under different circumstances, such as those in HSS's response and several of their other papers, are irrelevant to the central question. Showing that factor X modulates an island effect by some small amount clearly does not entail that it explains away that island effect completely. Showing individual examples that sound fairly acceptable does not explain why superadditivity is observed in tests using carefully controlled materials. We used lexically controlled materials, and observed variable superadditive effects across individuals. So HSS's concerns about the naturalness or discourse felicity of our materials have little to do with the debate, unless they wish to claim that discourse felicity will eliminate the superadditive effects (something explicitly tested in Sprouse 2007). Also, since the debate is about whether WM capacity limits are the cause of the superadditive effects that arise in a factorial definition of island effects (i.e. the relative acceptability of four sentence types), HSS's criticism regarding the complexities of the correlations between WM and the ratings of individual sentences falls wide of the mark. Finally, it should be noted that the question of whether the linguistic constraints on islands should be understood as syntactic, semantic, or pragmatic in nature is orthogonal to the current debate. In this debate, the notion 'grammatical constraint' should be understood in a broad sense, not as a narrow claim about syntax. With these preliminaries in mind, let us turn to HSS's primary objections.

2. The primary objections of HSS.

2.1. Do reductionist theories predict a correlation between WM resources and island effects? Absolutely. HSS argue that it is misguided to look for a correlation between WM and island effects because the link between WM and acceptability is currently unspecified. While it is true that there is no general theory of the link between WM and acceptability, this is not what is at stake here. We tested a specific explanation for the superadditive property of island effects. The reductionist theory that we tested draws a direct link between processes that conspire to overload the available WM resources, and the superadditive effect on acceptability (Kluender & Kutas 1993). We tested this reductionist theory because it is sufficiently explicit to suggest an account of the superadditive property of island effects.

HSS do not offer an alternative account of the superadditive property of islands. Instead they repeatedly point to the fact that island-violating sentences involve difficult-to-process properties. To repeat an important point: it is insufficient to simply observe that island-violating sentences often have multiple difficult-to-process properties, because that does not address the key fact about island acceptability. The fact to be explained is why combining those properties yields a result that is much less acceptable than the combined effect of each individual property would predict. The appeal to limited WM resources by Kluender and Kutas (1993) is the only concrete reductionist account of this effect that we are aware of.

Of course, it is possible that HSS aim merely to show that comprehension difficulty contributes to the acceptability of island-violating sentences, and that they are not really proposing a reductionist account, leaving linguistic constraints intact. In that case we

have no argument with them. But we doubt that this is their position, as it would make their broad attacks on linguistic island constraints somewhat puzzling. Instead, we interpret their position to be one of true reductionism for the island types under investigation here (*whether*, complex NP, subject, and adjunct islands), as they appear to be arguing that grammatical constraints are not necessary to account for the superadditivity of these particular island types. That being said, HSS also seem to be willing to assume a grammatical constraint for certain island types, such as the coordinate structure constraint. The adoption of a grammatical constraint for a single island type does not, however, negate the reductionist approach that they advocate for many other island types.

2.2. DO THE NONCORRELATIONS REFLECT AN INAPPROPRIATE CHOICE OF WM TASKS? HIGHLY UNLIKELY. HSS's second main objection assumes that their first objection is false, and argues that the WM tasks that we chose might fail to tap into the relevant WM constructs that are invoked by reductionist theories. The implication is that if we only had chosen the correct task, we would have found the predicted correlation. This concern is reasonable, in principle, but it is seriously undermined by what is known about the number of underlying WM constructs (very few) and the range of WM tasks that have been shown to fail to correlate with superadditivity in island effects (increasingly many).

It is important to distinguish between WM TASKS and the cognitive WM CONSTRUCTS that those tasks engage. Although there are a panoply of different tasks that go under the rubric of WM tasks, this does not mean there are just as many components to a WM architecture. In a hypothetical world in which each WM task taxes a distinct WM construct, we would expect different results based on the task chosen. However, in a world in which WM tasks tax the same set (or a substantially overlapping set) of WM constructs (Kane et al. 2004), perhaps in differing degrees, then it is less likely that the results should change based on the task chosen. The extent to which WM tasks overlap in the constructs that they tax is an empirical question. The serial-recall task and the *n*-back task used in our studies have been argued to tap into relatively different memory constructs (Kane et al. 2007). Moreover, the tasks we used are known to interact substantially with performance on explicit sentence memory tasks, giving rise to substantial individual difference correlations (word span, *n*-back; Roberts & Gibson 2002) or impairing performance under concurrent task conditions (*n*-back; Baddeley et al. 2009). Furthermore, in recent studies Dan Michel (2012) has also found noncorrelations between superadditivity and several additional tasks that we did not measure, including reading span, interference, and flanker attention tasks. These results have also received crosslinguistic support from studies in Japanese (Tokimoto 2009). The bottom line is that in order to find a correlation between island effects and a putatively more appropriate WM task, it would need to be the case that there are one or more underlying WM constructs that play a crucial role in the processing of extractions and island structures, and yet those constructs are entirely untapped by (i) *n*-back tasks, (ii) serial-recall tasks, (iii) reading span tasks, (iv) interference tasks, and (v) flanker tasks. This is a logical possibility, but we submit that it is highly unlikely.

2.3. DID OUR RESULTS DEMONSTRATE THE PREDICTED CORRELATION? ABSOLUTELY NOT, AND EFFECT SIZES DO MATTER. HSS's third main objection assumes that their first two objections are incorrect, and instead claims that our experiments did indeed detect the predicted negative correlation between WM and island effects, thus corroborating the reductionist theory. We regard this as a mischaracterization of our results, and it seems to be an optimistic revision of the predictions of the reductionist theory. Here, again, it is important to keep in mind that the debate is about whether WM resources can eliminate the need for linguistic island constraints, not whether island-violating sentences are

sometimes difficult to understand. Therefore, what matters is whether individual differences in WM capacity can account for a substantial proportion of the variance in the superadditive property of island effects. Our results, and the results of the replications that have been conducted since then (e.g. Michel 2012), suggest that this is not the case.

In describing our results, HSS mention only the three linear regressions for serial recall (experiment 1) that were significant in the direction predicted by the reductionist theory, failing to mention the two regressions that were significant in the opposite direction, and the seven others that trend in the opposite direction (sixteen regressions were run for serial recall alone; there are eighty total statistical analyses in our article if one includes the *n*-back regressions and resampling simulations). By setting aside 75% of the serial-recall regressions, HSS miss the fact that there was a clear contradiction in the direction of the slopes between experiment 1 and experiment 2, which primarily differed in ways that did not affect the basic superadditive property of the judgment data that we seek to explain. Furthermore, the 25% of serial-recall regressions that HSS choose to highlight are in some ways less relevant as tests of reductionist theories: the analyses they focus on are those that EXCLUDE PARTICIPANTS WHO DO NOT REPORT ISLAND EFFECTS, despite the fact that one of the most interesting predictions of the reductionist theory is that there should be individuals who fail to show island effects (Hofmeister & Sag 2010:403), that is, speakers who fail to show the typical superadditive effect in acceptability. It is not clear to us how such speakers can simultaneously be evidence for the reductionist theory and an impediment to observing its effects in a well-controlled experiment.

HSS also downplay the role of effect sizes (such as R^2) in the interpretation of statistical results. They make the claim that any significant correlation in the predicted direction, no matter how small, corroborates the reductionist theory. This is not true. Statistical significance is not equivalent to practical significance (see Cohen 1994 and Nickerson 2000 for accessible reviews, and see Gigerenzer 1994 for a discussion of the problem with ritualizing *p*-values alone). A classic example used to demonstrate this disassociation is a hypothetical drug that prolongs lifespan by a statistically significant margin, but that margin turns out to be a single day. Patients are unlikely to take this new drug (perhaps because it is expensive or dangerous) because one day is not a practically significant difference, despite being statistically significant. Effect sizes such as R^2 are one piece of information that researchers can use to evaluate the practical significance of a statistically significant result. In the case at hand, the question is whether the sometimes absent, sometimes extremely small effect sizes that we found (in both the predicted and opposite directions) are sufficient to justify the claim that island effects are explained away by effects of WM capacity. We do not believe that they are, given the current formulation of reductionist theories (i.e. a single explanatory predictor, WM, is responsible for the superadditive effect). For reductionist theories to account for the effect sizes that we observed, they would likely need to be reformulated to incorporate additional explanatory predictors of the superadditive effect, so that the role played by WM could be scaled down accordingly.

3. MOVING THE DEBATE FORWARD. The attraction of reductionist accounts of islands is that they offer the promise of removing complex constraints from grammatical theories. We appreciate this attraction, and in other domains we have argued that some linguistic phenomena are better understood as consequences of language processing mechanisms (e.g. Wagers et al. 2009, Phillips 2011, Sprouse et al. 2011). We are not 'grammar apologists'. Rather, we believe that each case must be evaluated on its merits. In the case of island effects we believe that the existing evidence is clear. There is no question that

language processing factors can modulate the acceptability of island-violating sentences, but that is not what is under debate. The debate is about the possibility of eliminating linguistic constraints on islands, using independently motivated constructs. The challenges to the reductionist approach are substantial.

(i) As discussed here, the WM capacity theory is the only reductionist account of the superadditive property of islands that we are aware of. But the noncorrelations found in our studies and others provide no support for this account.

(ii) As discussed in §2 of our article, the WM capacity theory is based on the assumption that island structures incur a substantial cost to acceptability, independent of extraction. Our judgment studies show that this assumption is not always true (see also §§2.3, 4.2, and 5.2).

(iii) Resource-capacity accounts link island phenomena to language processing difficulty, yet active gap creation is possible inside islands, when allowed by the grammar of parasitic gaps (Phillips 2006).

(iv) If island effects are due to independently motivated processing limitations, then similar effects should be found when other long-distance relations span island boundaries. Recent language processing evidence challenges this: island structures do not impede the active processing of cataphoric dependencies (Yoshida, Kazanina, Pablos, and Sturt 2012) or backward sluicing dependencies (Yoshida, Ackerman, Ward, and Purrier 2012). These phenomena—establishing coreference, resolving deletions—are formally distinct from the processes that create island-violating structures. They are functionally very similar, however, since they involve forward parsing processes across complex structural domains and subsequent retrieval from memory. The reductionist theory of islands seems to imply that dependencies that make similar functional demands should show similar processing profiles.

(v) Reductionist theories do not easily account for island effects that arise without displacement in WH-in-situ languages (Huang 1982, Malhotra 2009). It is plausible to expect WM demands to be reduced in displacement constructions in those languages, given that the WH-words in those languages appear in their thematic position. Additionally, even if covert WH-movement processes are invoked to account for difficulty effects in those languages, this would not entail simultaneous processing of island structures and long-distance dependencies, as needed by the reductionist account.

(vi) Reductionist theories have little to say about the constrained cross-language variation that is observed in island effects, nor about the various cases of constructions in certain languages that present the illusions of acceptable island violations.

Grammatical approaches to islands can account for all of these findings. They predict the noncorrelations with WM capacity observed in our experiments. They are consistent with the finding in our experiments that some island structures are not independently difficult. They also readily accommodate the findings about parasitic gaps in parsing, and about the difference between processing cataphora and backward sluicing. They are able to describe findings in WH-in-situ languages, and capturing constrained cross-language variation is one of their strengths. We take all of this, our findings and previous findings, to be evidence in favor of a grammatical approach to island effects. This, of course, does not preclude the logical possibility that a reductionist theory could be formulated that accounts for all of these facts equally as well as grammatical theo-

ries. But it is not accurate to claim that current reductionist theories are on equal empirical footing with grammatical theories. Instead, advocates of the reductionist approach should propose novel mechanistic theories that explicitly account for the same range of phenomena as grammatical theories. We believe this would clarify matters considerably and move the present debate forward.

REFERENCES

BADDELEY, ALAN D.; GRAHAM J. HITCH; and RICHARD J. ALLEN. 2009. Working memory and binding in sentence recall. *Journal of Memory and Language* 61.438–56.

COHEN, JACOB. 1994. The earth is round ($p < .05$). *American Psychologist* 49.997–1003.

GIGERENZER, GERD. 1994. Mindless statistics. *The Journal of Socio-Economics* 33.587–606.

HOFMEISTER, PHILIP, and IVAN A. SAG. 2010. Cognitive constraints and island effects. *Language* 86.366–415.

HOFMEISTER, PHILIP; LAURA STAUM CASASANTO; and IVAN A. SAG. 2012. How do individual cognitive differences relate to acceptability judgments? A reply to Sprouse, Wagers, and Phillips. *Language* 88.390–400.

HUANG, C.-T. JAMES. 1982. *Logical relations in Chinese and the theory of grammar*. Cambridge, MA: MIT dissertation.

KANE, MICHAEL J.; ANDREW R. A. CONWAY; TIMOTHY K. MIURA; and GREGORY J. H. COLFESH. 2007. Working memory, attention control, and the *n*-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33.615–22.

KANE, MICHAEL J.; D. ZACH HAMBRICK; STEPHEN W. TUHOLSKI; OLIVER WILHELM; TABITHA W. PAYNE; and RANDALL W. ENGLE. 2004. The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General* 133.189–217.

KLUENDER, ROBERT, and MARTA KUTAS. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8.573–633.

MALHOTRA, SHITI. 2009. On WH-quantifier interactions. College Park: University of Maryland, MS.

MICHEL, DAN. 2012. Content-addressable working memory and the grammar-processing island debate. Paper presented at the 86th annual meeting of the Linguistic Society of America, Portland, OR, January 5–8.

NICKERSON, RAYMOND. 2000. Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods* 5.241–301.

PHILLIPS, COLIN. 2006. The real-time status of island constraints. *Language* 82.795–823.

PHILLIPS, COLIN. 2011. Some arguments and nonarguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes*, published online July 2011; to appear in print.

ROBERTS, ROSE, and EDWARD GIBSON. 2002. Individual differences in working memory. *Journal of Psycholinguistic Research* 31.573–98.

SPROUSE, JON. 2007. *A program for experimental syntax*. College Park: University of Maryland dissertation.

SPROUSE, JON; SHIN FUKUDA; HAJIME ONO; and ROBERT KLUENDER. 2011. Grammatical operations, parsing processes, and the nature of WH-dependencies in English and Japanese. *Syntax* 14.179–203.

SPROUSE, JON; MATT WAGERS; and COLIN PHILLIPS. 2012. A test of the relation between working memory and syntactic island effects. *Language* 88.82–123.

TOKIMOTO, SHINGO. 2009. Island phenomenon in Japanese and working memory: Syntactic constraints independent from working memory constraints. Poster presented at the 22nd annual CUNY Sentence Processing Conference, University of California, Davis, March 26–28.

WAGERS, MATT; ELLEN FRANCES LAU; and COLIN PHILLIPS. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61.206–37.

YOSHIDA, MASAYA; LAUREN ACKERMAN; REBEKAH WARD; and MORGAN PURRIER. 2012. The processing of backward sluicing. Paper presented at the 25th annual CUNY Sentence Processing Conference, CUNY Graduate Center, March 14–16.

YOSHIDA, MASAYA; NINA KAZANINA; LETICIA PABLOS; and PATRICK STURT. 2012. On the origin of islands. *Language and Cognitive Processes*, to appear.

Sprouse
Department of Cognitive Sciences
University of California, Irvine
3151 Social Science Plaza A
Irvine, CA 92697-5100
[jsprouse@uci.edu]
[mwagers@ucsc.edu]
[colin@umd.edu]