Assessing the replication landscape in experimental linguistics

Kristina Kobrock*[,a], Timo B. Roettger[b]

[a] *University of Osnabrück, Institute of Cognitive Science, Wachsbleiche 27, 49090 Osnabrück, Germany*
[b] *Universitetet i Oslo, Department of Linguistics and Scandinavian Studies, Niels Henrik Abels vei 36, 0313 Oslo, Norway*

Abstract

Replications are an integral part of cumulative experimental science. Yet many scientific disciplines do not replicate enough and novel confirmatory findings are valued higher than direct replications and null results. A systematic assessment of the replication landscape in experimental linguistics is as of yet missing. The present study estimated replication rates for over 50.000 articles across 98 journals, using a combination of automatic string matching using the Web of Science search engine and in-depth manual inspections of over 200 papers. Across a representative set of experimental linguistic journals, we found a median rate of mentioning the search string "replicat*" of 1.6%. Manual analyses showed that only roughly half of those articles mentioning the term were actual replication studies. We found only eight direct replications, i.e. studies that arrive at the same scientific conclusions as an initial study by using exactly the same methodology. Moreover, our analyses suggest that only 1 out of 1600 experimental linguistic studies reports a direct replication performed by independent researchers. We conclude that, similar to the state-of-affairs in neighboring disciplines, experimental linguistics does not replicate enough and suggest possible ways forward. [185 words - goal: ~150]

keywords: replication, meta-research, journal impact factor, publishing guidelines

data availability statement: All data, analysis code and models used can be inspected at https://osf.io/9ceas/.

1. Introduction

Understanding the inner workings of human language and its cognitive underpinnings has been increasingly shaped by experimental data. With a field that builds its theories on a rapidly growing body of experimental evidence, it is of critical importance to evaluate and substantiate existing findings in the

*Corresponding Author
Email addresses: kkobrock@uni-osnabrueck.de (Kristina Kobrock), timo.b.roettger@gmail.com (Timo B. Roettger)

literature because evidence provided by a single study is limited (e.g., Amrhein et al., 2019). Scientists are trained to ensure the reliability and generalizability of scientific findings by conducting direct replication studies, i.e. studies that arrive at the same scientific conclusions as an initial study by collecting new data and completing new analyses but using the same methodology (see Barba, 2018 for a comprehensive overview of different terminological uses).

Replications are an integral part of cumulative experimental science (e.g., Campbell, 1969; Rosenthal, 1990; Zwaan et al., 2018). Yet many scientific disciplines do not replicate enough. Researchers from diverse fields such as psychology (Makel et al., 2012), educational science (Makel and Plucker, 2014), and economics (Mueller-Langer et al., 2019) report on very low numbers of published replications, ranging from 0.1% in economics to 1.6% in psychology.

One reason for the observed lack of replication studies is the asymmetric incentive system in academia that rewards novel confirmatory findings over direct replications and null results: Replication studies are not very popular because the necessary time and resource investments are not appropriately rewarded (e.g., Koole and Lakens, 2012; Nosek et al., 2012). Both successful replications (Madden et al., 1995) and repeated failures to replicate (e.g., Doyen et al., 2012) are rarely published. Even if they are, replications usually appear in less prestigious outlets than the original findings. These dynamics lead to an abundance of positive findings in the absence of possible conflicting negative evidence (see also Fanelli, 2010) and the widely held view that replications lack prestige, originality, or excitement (e.g., Lindsay and Ehrenberg, 1993).

The lack of replication studies threatens the very fabric of cumulative progress in experimental science because experimental results are often taken for granted without them ever being replicated. But this leads to a related problem: If we don't try, we won't fail to replicate our studies. However, if we try, we fail more often than we would like: Coordinated efforts to replicate published findings have uncovered alarmingly low rates of successful replications in fields such as psychology (Open Science Collaboration, 2015), economics (Camerer et al., 2016), and social sciences (Camerer et al., 2018), a state of affairs that has been referred to as the "replication crisis" (Fidler and Wilcox, 2018).

The replication crisis is not rooted in a singular cause, but pertains to a network of different practices and incentive structures, all of which conjointly lead to an increase in results that are not replicable. Researchers have identified practices that might have contributed to the wide-spread lack of replicability, including but not limited to too small sample sizes (e.g., Button et al., 2013), lack of data and materials sharing (e.g., Nosek et al., 2015), use of anti-conservative statistical methods (e.g., Yarkoni, 2019), large analytical flexibility (e.g., Simmons et al., 2011), and lack of generalizability across diverse contexts and populations (Henrich et al., 2010).

These limitations are present, and maybe even exacerbated in experimental linguistic research: Access to certain linguistic populations is often limited or too cost-intensive, making it difficult to collect sufficiently large samples. Experimental linguistic research is resource-intensive because of equipment cost and complexity, elaborateness of data collection procedures, and computational

requirements of data analysis and curation. This often results in studies with small sample sizes and, consequently, with low statistical power (e.g., Casillas, 2021; Kirby and Sonderegger, 2018). Statistical analyses in linguistics are often ignoring important assumptions (e.g., Winter and Grice, 2021) and are characterized by a large number of researcher degrees of freedom (Roettger, 2019). Moreover, claims about human language are often based on a small set of languages, limiting the generalizability of claims about human language (e.g., Levisen, 2019; Majid and Levinson, 2010).

In light of the large overlap in research practices between linguistics and neighboring disciplines for which low replication rates and failures of attempts to replicate have been attested, there are raising concerns about both replication rates and replicability in the field of experimental linguistics (e.g., Marsden et al., 2018; Roettger and Baer-Henney, 2019). A number of failed replication attempts reported in various subfields of linguistics indicate that these concerns have to be taken seriously (e.g., Chen, 2007; Morey et al., 2021; Nieuwland et al., 2018; Papesh, 2015; Stack et al., 2018; Westbury, 2018).

Moreover, there might be only very few published direct replications in linguistics. In their detailed assessment of replications in second language (L2) research, Marsden et al. (2018) explored 67 self-labeled L2 replication studies for a wide variety of characteristics. Their results indicate that for every 400 articles, only one replication study is published which translates into a replication rate of 0.25%. Moreover, their sample did not include a single direct replication study, i.e. a replication that strictly followed the design of the initial study, a state of affairs that is worrisome and warrants further investigation. To our knowledge, there is no systematic assessment of replication rates across experimental linguistics beyond Marsden et al. (2018). This paper aims at filling this gap. To gauge the past and current replication landscape in experimental linguistics, track progress over time, and calibrate future policy and training initiatives, it will be useful to assess the prevalence of replications across experimental linguistics and explore their contributing factors.

The present study assesses the frequency and typology of replication studies that have been published in a representative sample of experimental linguistic journals from 1988 to 2020. Our study aimed at answering two main questions: "How many direct replications are published in experimental linguistics?" and "Are there factors that affect the replication rates and are they either found at the journal level (e.g. journal policies, open access, journal impact factor, etc.) or at the study level (e.g. composition of authors, investigated language, etc.)?" The study consisted of two analyses: First, we assessed the frequency of articles mentioning the term replication (search string: replicat*) across 100 linguistic journals. Second, we categorized the type of replication studies (direct, partial, conceptual) in a subset of twenty journals. We then related their replication rates to factors like the years of publication, and the citations of both initial and replication study.

2. How often do journals mention the term replicat*?

The key dependent variable of the first part of this study was the rate of replication mention for journals relevant to the field of experimental linguistics.

2.0.1. Data availability

The article counts for all journals in the sample and calculated rates of replication mention can be inspected at https://osf.io/yefr8/. The journals coded for the factors journal policy, journal impact factors and open access publishing can be retrieved at https://osf.io/rukc7/.

2.0.2. Material and methods

The study design has been preregistered at 2021-03-08 and can be inspected at https://osf.io/9ceas/.

In order to determine the rates of replication mention for individual journals, we drew on a method introduced by Makel et al. (2012). First, a sample of 100 journals relevant to the field of experimental linguistics was identified by making use of the search engine "Web of Science" (https://webofknowledge.com) (access date: 2021-03-03). We restricted the search results to journals in the web of science category "Linguistics" which had at least 100 articles published and a high ratio of articles containing the term experiment* in title, abstract or keywords in order to ensure that the subset contained journals that are relevant for experimental linguistics research. Among those, all articles categorized as having been published in English and between 1945-2020 were taken into account.

The ratio between overall number of articles and those articles mentioning the term "experiment*" ranged between 6.1 and 60.3 (with a median of 11.5) across journals. The full sample of journals can be inspected at https://osf.io/q2e9k/ or in the appendix of this article.[1]

After journal selection, we obtained the total count of articles containing the search term replicat* in title, abstract or keywords for each journal via the Web of Science search. Following the method presented by Makel et al. (2012), the rates of replication mention are calculated by dividing the number of articles containing the term replicat* by the total number of eligible articles for each journal. As we were only interested in experimental linguistic studies, we only considered articles containing the search term experiment* as eligible.

Rates of replication mention were then related to three journal properties: journal policies with regards to replication studies, journal impact factor and whether the journal publishes open access or not. To gain an understanding of the journal policies with regards to replication studies, we examined the journals' submission guidelines adopting a method suggested by Martin and Clarke (2017). They grouped psychology journals into four categories dependent

[1]Two journals, namely "Language and Cognitive Processes" and "Language, Cognition and Neuroscience" had to be excluded because it turned out during analysis that both journals have been renamed in 2013 and that they have already been included in our sample under the new name. Our final sample thus included only 98 journals.

on whether they (explicitly or implicitly) encouraged replication studies or not in their "instructions to authors" and "aims and scope" sections on the journal websites. For our analysis, we only distinguished between those journals explicitly encouraging replication studies and those that do not. We extracted journal impact factors via Journal Citation Reports (https://jcr.clarivate.com).[2] Whether journals offered an open access publication or not was assessed via Web of Science. We distinguished between two access categories: those journals which are listed in the Directory of Open Access Journals (DOAJ) ("DOAJ gold"), and those journals with some articles being published as open access articles ("partial") and those journals with no option to publish open access ("no") grouped together.

2.1. Results and Discussion

Out of the 51272 articles in our sample, 8006 mentioned the term 'experiment*' in title, abstract, or keywords and were thus assumed to be articles presenting an experimental investigation. Out of these articles, 347 contained the term replicat* which amounts to a replication rate of 4.3% across experimental linguistic articles.

The distribution of the rate of replication mention substantially varies across journals ranging from 0 to 12.82%. The median rate of replication mention is as low as 1.6% (SD = 3.3), a rate that is comparable to that Makel et al. (2012) have reported in their assessment of replications in psychology. Almost half of all journals (n = 43) did not mention the term in any of their articles. Figure 1 illustrates the variation across those journals that exhibited at least one mention of the term.

Following preregistered protocol, we statistically estimated the rate of replication mention as predicted relative to the following factors: journal impact factors (continuous, henceforth jif), open access (binary: open access journal or not), and replication policies (binary: either explicitly encourage or not). We used Bayesian parameter estimation based on generalized linear regression models with a binomial link function. The model was fitted to the proportion of replication mentions per journal using the R package brms (Bürkner, 2016). We used weakly informative normal priors centered on 0 (sd = 2.5) for the intercept and Cauchy priors centered on zero (scale = 2.5) for all population-level regression coefficients. These priors are what is referred to as regularizing (Gelman et al., 2008), i.e. our prior assumption is agnostic as to whether the predictors affect the dependent variable, thus making our model conservative with regards to the predictors under investigation. Four sampling chains with 2000 iterations each have been run for each model, with a warm-up period of 1000 iterations. For relevant predictor levels and contrasts between predictor levels, we report the posterior probability for the rate of replication mention. We summarize these distributions by reporting the posterior mean and the 95% credible intervals (calculated as the highest posterior density interval).

[2]The 2019 journal impact factors are calculated by dividing the citations in 2019 to items published in 2017 and 2018 by the total number of citable items in 2017 and 2018.

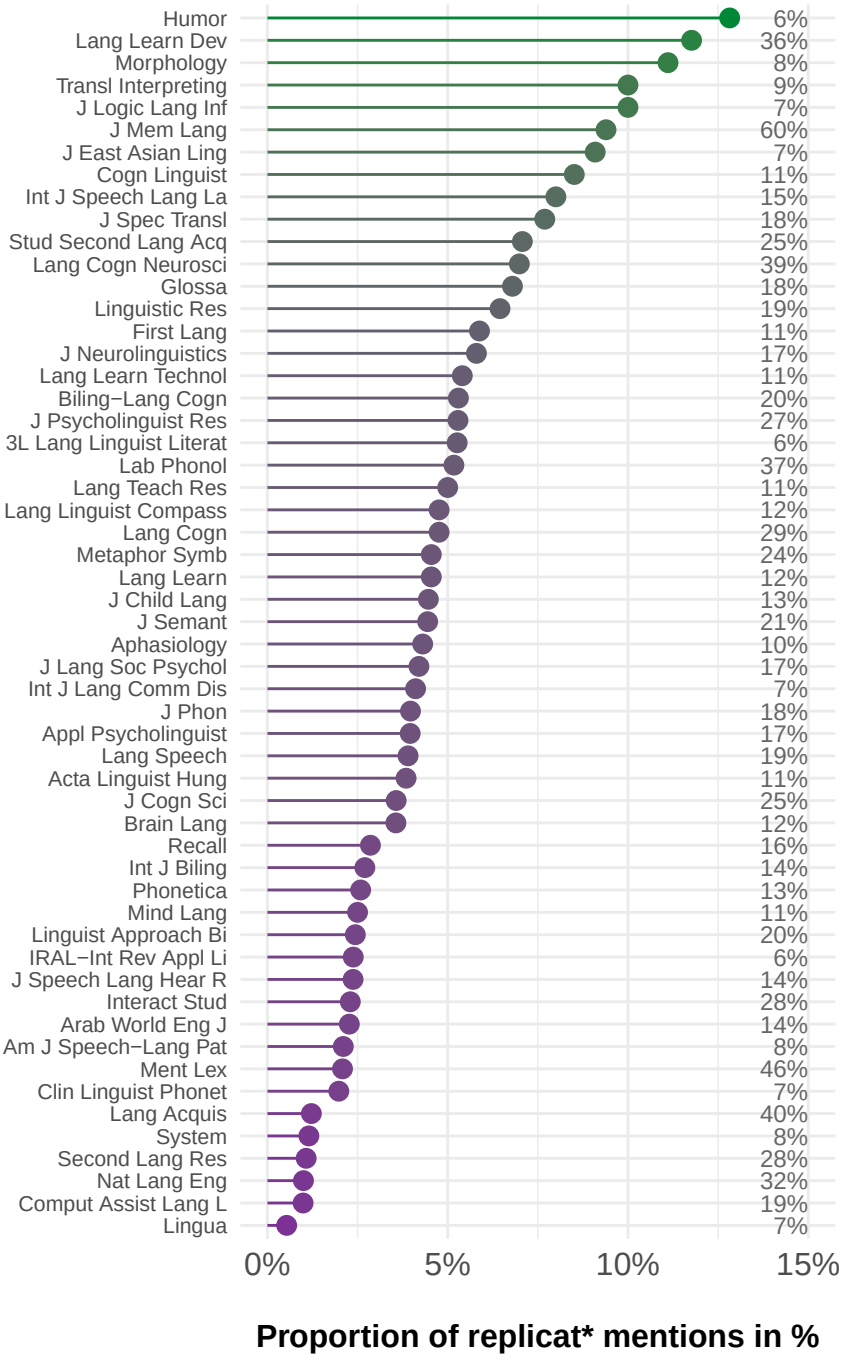

Figure 1: Variation in rate of replication mention across those journals that exhibited at least one mention of the term. Numeric values on the right indicate the observed proportion of articles containing the string experiment* in title, abstract or keywords.

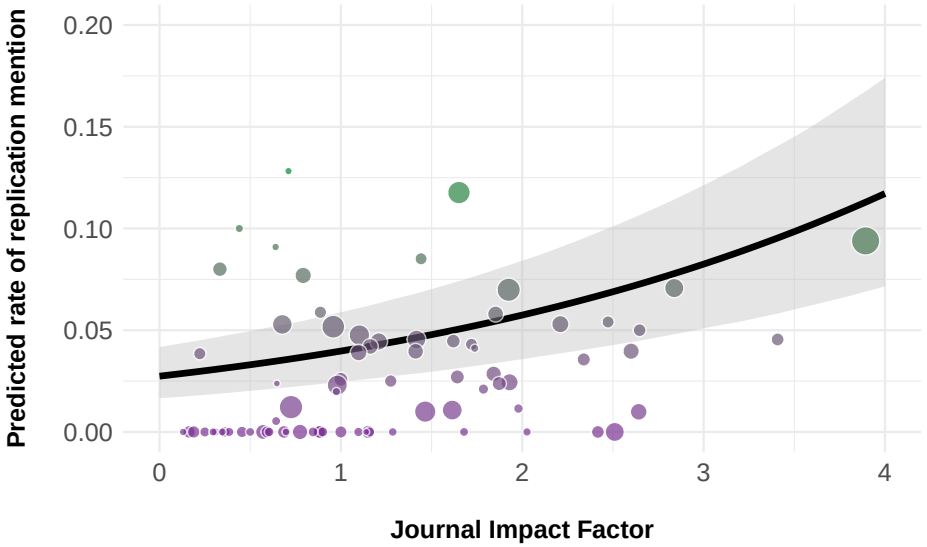

Figure 2: Rate of mentioning the term 'replicat*' across sampled journals plotted against their journal impact factor. Each point represents one journal. Point size indicates the proportion of papers categorized as experimental (i.e. larger points indicate journals with more experimental articles). Line and shading indicate model predictions and 95% credible intervals.

The model estimates the proportion of replication mentions as 2.7% [1.7, 4.2] at jif = 0 and estimates an increase of the proportion with each integer unit of jif (log odds = 0.39 [0.29, 0.49]). Figure 2 illustrates this relationship.

Further explorations, however, indicate that jif is correlated with the number of experimental studies reported in a journal (Spearman correlation = 0.43).[3] Given the observed correlation, it remains unclear if the term replicat* is really used more often in high impact journals or simply more common in journals that generally publish more experimental studies (which tend to have higher jifs).

The preregistered model estimates the impact of whether the journal allows for open access publishing or not and whether replications are explicitly encouraged or not both as positive, i.e. the term replication is mentioned more often in both open access journals and journals that explicitly encourage direct replications. However, the uncertainty around these estimates is substantial (open access: 0.41 [-0.41, 1.14]; policy: 0.24 [-0.27, 0.72]) due to the small number of journals that explicitly encourage direct replications (2 out of 98), and the relatively small number of open access journals (11 out of 98) and thus not informative.

[3]This exploratory analysis was not preregistered.

3. How many articles containing the term replicat* are actual replications?

The second part of the study aimed at investigating further what types of replication studies are published and whether replications are becoming more frequent over time. Because the string replicat* is commonly used in ambiguous ways, the articles that contained the search term required further analysis to determine whether the articles in question indeed reported a replication study or whether they used the term in a different way.

3.0.1. Data availability

The full list of coded articles can be obtained at https://osf.io/4rtvh/.

3.0.2. Material and methods

From the superset of 100 journals obtained above, the 20 journals with the highest proportion of experimental studies were selected for a more detailed analysis while excluding journals for which less than 2 hits (TS=replicat*) could be obtained (see at https://osf.io/f3yp8/ for a list of article counts per journal). Because of the skewed distribution of our sample (114 hits for Journal of Memory and Language, and less than 40 for all other journals, respectively), we randomly selected 50 out of the 114 articles for the Journal of Memory and Language (see at https://osf.io/6vfpe/ for details). The sampling procedure above resulted in 210 possible self-labeled replication studies.

We identified whether the article in question indeed presented a replication study or not. Parts of the papers that were examined were title and abstract of the paper, text before and after occurrences of the search term replicat*, the paragraph before the Methods section as well as the first paragraph of the Discussion section (following and adapting the procedure specified by Makel et al., 2016). If the authors explicitly claimed that (one of) their research aim(s) was to replicate the result or methods of an initial study, this article was treated as a replication and was submitted to further analysis according to the preregistered coding scheme which can be inspected at https://osf.io/ct2xj/.

When extracting number and types of changes made to the initial study, we assumed that the authors of a replication study did not make any drastic changes *without* reporting them. Following Marsden et al. (2018), replication studies were classified according to the number of changes made into three categories: direct replication (0 changes), partial replication (1 change) and conceptual replication (2 or more changes). We noted the nature of methodological changes as one of the following categories: experimental paradigm, sample, materials/experimental set-up, dependent variable, independent variable, and control. We also recorded the language under investigation. The information on whether the article was published open access as well as citation counts and years of publication for both studies were obtained from Web of Science. An author overlap was attested when at least one author was a (co-)author on both studies. During the coding procedure of the articles, we encountered edge cases that we did not anticipate in our preregistration: When several self-labeled replication studies were mentioned

in one article, we chose the first mentioned study for our analysis. If there were one independent, but also one or more inner-paper replications, i.e. experiments that replicated previously obtained results from the same article, we chose the independent study for analysis.

3.1. Results and Discussion

Out of the 210 articles in the subsample, 200 (95.2%) indeed presented experimental linguistics research. The remaining 10 (4.8%) were not experimental in nature, but comments, reviews or computational studies. Out of the 200 experimental studies, 116 were self-claimed replications according to our criteria. The remaining 84 mentions were articles that mentioned the term in other contexts or studies that did not specify the concrete aim of replicating an initial study's design or results. Moreover, many papers used the term "replicated" in a broad sense that roughly translates into "finding a similar result," thus not qualifying as a replication study as defined above. Out of the replication studies, we categorized 66 (56.9%) as conceptual, 42 (36.2%) as partial, and only 8 (6.9%) as direct replications.

About one third (31.6%) of the replications were published in the scope of the same paper as the initial study. Publishing multiple experiments within one article and replicating one's own previously obtained results thus seems to be common practice in the field of experimental linguistics.

Looking closer at direct replications, 3 studies were independent studies, i.e. there was no overlap between authors of the initial study and the replication study. Out of these independent direct replication studies, 2 were self-labeled as successful replications. In other words, our sample included only one failed independent and direct replication attempt. These low rates indicate that replication attempts, and especially direct replication attempts, are rather rare in the experimental linguistics literature - an observation that is in line with replication rates estimated for other research fields (Makel et al., 2012; Makel and Plucker, 2014; Mueller-Langer et al., 2019).

Figure 3 illustrates the development of replication studies throughout publication years. While the overall number of studies increased over the years, the proportion of direct replications remained stable at best. However, it seems as if there is an increasing number of partial and conceptual replications that was published within the last few years.[4]

One possible reason for the fact that (direct) replication rates are not increasing for the field according to our analysis could be that experimental linguistics predominantly replicates experimental findings across languages, making the studies by definition only partial replications. However, only one quarter of replications targeted a different language than the initial study (24.1%). The

[4]Given the small number of direct replications in our sample, both a descriptive assessment and an inferential assessment as preregistered are very uninformative. The reader is directed to the supplementary materials, if they are interested in the model outputs of the preregistered analysis.

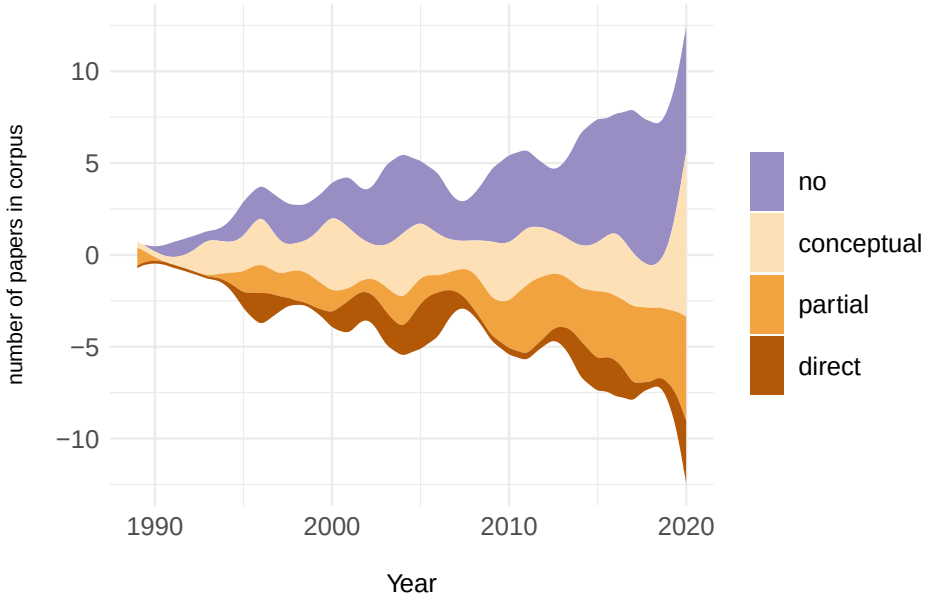

Figure 3: Development of amount of replication studies published over time.

majority of replication efforts were conducted within the same language as the initial study. In fact, 67.2% of all replication studies in our sample had one variety of English as the main language of investigation either in the replication or in the corresponding initial study.

The median number of years between an initial and a replication study is 7 years. Initial studies were on average 41.6 times cited before a replication was published which amounts to an average yearly citation rate of 5.9 citations. This average citation rate is well above the impact factor of core linguistic journals (median journal impact factor in superset: 1.1). Replication studies were on average only 17 times cited which amounts to an average yearly citation rate (calculated up to the time of analysis) of 0.6 citations. These results are in line with Marsden et al.'s (2018) assessment of second language research. They found that replication studies were on average conducted after more than six years and after over a hundred citations of the original study and concluded that replications are either only performed or only published after the original study had already substantially impacted the field. Our findings support this interpretation. The observed "drop" in the number of citations of replication studies compared to corresponding initial studies is in line with the lack of perceived value of replication studies reported in other fields (e.g., Koole and Lakens, 2012; Nosek et al., 2012).

3.1.1. Case study of Journal of Memory and Language

Due to the skewed sample, we conducted a subset analysis of articles published in Journal of Memory and Language (JML) which accounts for the largest part

of our sample compared to other journals and is the journal with the highest impact factor (3.9). We find that 34 (68%) of the 50 papers in our sample contain replication studies. Of these, 15 (44.1%) are conceptual, 16 (47.1%) are partial, and 3 (8.8%) are direct replication studies which is in line with the results for the whole sample. Only 2 of the studies published in JML were independent direct replication studies (one of which was successful). We conclude that we have little reason to believe that the large proportion of JML articles substantially skewed our results (for the better or worse).

4. General discussion

The current study aimed at providing a comprehensive survey of published replications in experimental linguistic research. By analyzing the publication history of over 50000 articles across 100 journals that publish experimental linguistic research, our study found that 4.3% of experimental linguistic publications used the term replicat* in title, abstract or key words. A more thorough analysis of 210 sampled experimental articles containing the term replicat* revealed that only around half of the hits represented actual replication studies, reducing the effective replication rate to 2.4%. This rate is slightly higher than reports of comparable investigations in psychology (1.6%, Makel et al., 2012), educational science (0.1%, Makel and Plucker, 2014), and economics (0.1%, Mueller-Langer et al., 2019). The higher rate might be due to a methodological choice, however. Due to large plurality of methods in linguistics, we calculated the replication rate based on only those articles that contained the term experiment* (as opposed to all articles in the sample), reducing the denominator substantially.

A closer look at the nature of replication studies revealed that the majority of replication studies were studies that diverted from the initial study by at least one design choice. Only 6.9% were direct replications, i.e. studies that directly repeated an initial study without self-reported changes to the design and only three of these were replications conducted by an independent team of researchers. Taken together, 0.06% of experimental studies are independent direct replications in the field of linguistics. Or in other words, only 1 in over 1600 experimental linguistic articles is an independent direct replication. This clearly indicates that replication attempts, and especially independent direct replication attempts, are very rare in the experimental linguistics literature.

Before making our recommendations, there are important caveats to our results. If research articles were not framed as experimental, then they were not included in the analysis. If experimental articles were not framed as replications, then they were not categorized as such. While a clear limitation of our method, it also applies more generally: If studies are not framed as replications by using the term replication, readers' ability to connect research to its intellectual precedents is severely limited. To circumvent this methodological problem, the entire sample would have to undergo manual coding which is not feasible for a large-scale assessment. Future research using alternative assessment methods or more in-depth investigation of subfields or specific journals might result in different replication rates. Moreover, our assessment of replication types relied

on two assumptions. First, we assume that the authors disclosed changes to the initial study in a transparent way. Second, we assume that if changes are disclosed, we were able to extract and interpret these changes accurately. Neither of these assumptions must hold entirely, thus any rates that are generated here, are necessarily only a rough proxy of the true replication rate. Nevertheless, given that our findings seem to align well with evidence from other fields as well as an in-depth analysis of a subfield of linguistics (Marsden et al., 2018), we are confident that our conclusion holds.

Although the present study is the first systematic assessment of replication rates in linguistics, our conclusions are hardly surprising. Academic incentive systems do not reward replication studies. Neither journals nor funders encourage them. For example, Martin and Clarke's (2017) survey results suggest that in 2015 only 3% of psychology journals explicitly state that they will consider publishing replications. Similarly, out of the 98 journals in our sample, only 2 encouraged direct replications. And even if one manages to publish a replication, replication studies are characterized by much lower yearly citation counts compared to corresponding initial studies, leading to a lack of perceived prestige (e.g., Koole and Lakens, 2012; Marsden et al., 2018; Nosek et al., 2012). Direct replications simply do not seem worth their costs.

In order to overcome the asymmetry between the cost of direct replication studies and the presently low academic payoff for it, we must re-evaluate the value of direct replications. Funding agencies, journals, but also editors and reviewers, need to start valuing direct replication attempts as much as they value novel findings. For example, we could either dedicate existing journal space to direct replications (e.g. as its own article type) or create new journals that are specifically dedicated to replication studies. Journals could help normalizing replication studies by calls for special issues dedicated to replications of influential findings like the Journal of Memory and Language.[5] Another alternative is the Pottery Barn rule, implemented by for example Royal Society Open Science: Once a journal has published a study, it commits to publish all direct replications of this study.[6]

At the same time, we should attempt to find more resource-efficient ways to both identify replication targets and conduct replication studies. We believe, most people would agree that not every study is worth replicating. Take for example the McGurk effect, i.e. perceiving a sound that lies in-between an auditory presented component of one sound and a visually presented component of another one (McGurk and MacDonald, 1976). This phenomenon is probably replicated in dozens of linguistic classrooms every semester across the globe. Finding convenient yet effective tools to identify worthwhile replication targets is an active meta-scientific field at the moment (e.g., Coles et al., 2018; Hardwicke et al., 2018; Isager et al., 2021a) and feasible algorithms are currently actively

[5] https://www.journals.elsevier.com/journal-of-memory-and-language/call-for-papers/replicating-influential-findings

[6] https://royalsociety.org/blog/2018/10/reproducibility-meets-accountability/

developed and tested (Isager et al., 2021b). When it comes to more accessible ways to conduct replication studies, several authors have suggested involving our students more rigorously (Frank and Saxe, 2012; Grahe et al., 2012; e.g., Leeuw et al., 2019; Roettger and Baer-Henney, 2019), possibly creating a rich learning experience for our students while at the same time reducing the resource costs of replication studies. Alternatively, resources can be pooled across multi-lab replication efforts, effectively reducing the costs for individual researchers and labs (e.g., Frank et al., 2017; Nieuwland et al., 2018; Open Science Collaboration, 2015).

We are confident that the field of linguistics can function as a role model for neighboring fields. Although major meta-scientific discourses are held in other fields, linguistics has demonstrated quick uptake of methodological reforms time and time again. A point in case is the swift uptake of Registered Reports[7], a new article form in which a study proposal is reviewed before the research is undertaken. While the uptake across disciplines is slow, linguistics has at least 12 high-impact journal outlets that offer Registered Reports. Moreover, an increasing number of reproducibility initiatives founded in the field during the last few years give hope that the field is continuing to evaluate their past, current, and future practices and succesfully face the challenges ahead. This paper was an attempt to contribute to this development. We hope our assessment allows future efforts to track progress over time and calibrate policies across experimental linguistics.

5. Appendix

Table 1: The full sample of journals sorted by their ratio of experimental linguistics articles.

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
JOURNAL OF MEMORY AND LANGUAGE	2,012	1,214	60.34

[7] http://cos.io/rr

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
LANGUAGE AND COGNITIVE PROCESSES	783	399	50.96
MENTAL LEXICON	105	48	45.71
LANGUAGE ACQUISITION	207	82	39.61
LANGUAGE COGNITION AND NEUROSCIENCE	590	229	38.81
LABORATORY PHONOLOGY	155	58	37.42
LANGUAGE LEARNING AND DEVELOPMENT	141	51	36.17
NATURAL LANGUAGE ENGINEERING	312	100	32.05
LECTURE NOTES IN COMPUTER SCIENCE	150	46	30.67
LANGUAGE AND COGNITION	144	42	29.17

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
INTERACTION STUDIES	312	87	27.88
SECOND LANGUAGE RESEARCH	338	93	27.51
JOURNAL OF PSYCHOLINGUISTIC RESEARCH	1,691	454	26.85
STUDIES IN SECOND LANGUAGE ACQUISITION	389	99	25.45
COMPUTATIONAL LINGUISTICS	521	130	24.95
JOURNAL OF COGNITIVE SCIENCE	114	28	24.56
METAPHOR AND SYMBOL	278	66	23.74
LECTURE NOTES IN ARTIFICIAL INTELLIGENCE	113	26	23.01

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
JOURNAL OF SEMANTICS	218	45	20.64
LINGUISTIC APPROACHES TO BILINGUALISM	204	41	20.10
BILINGUALISM LANGUAGE AND COGNITION	753	151	20.05
COMPUTER ASSISTED LANGUAGE LEARNING	531	101	19.02
LINGUISTIC RESEARCH	166	31	18.67
LANGUAGE AND SPEECH	1,521	282	18.54
JOURNAL OF SPECIALISED TRANSLATION	141	26	18.44
GLOSSA A JOURNAL OF GENERAL LINGUISTICS	561	103	18.36

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
JOURNAL OF PHONETICS	1,389	252	18.14
JOURNAL OF NEUROLINGUISTICS	806	138	17.12
APPLIED PSYCHOLINGUISTICS	1,202	202	16.81
JOURNAL OF LANGUAGE AND SOCIAL PSYCHOLOGY	711	119	16.74
RECALL	214	35	16.36
PHONOLOGY	190	31	16.32
INTERPRETING	131	20	15.27
EURASIAN JOURNAL OF APPLIED LINGUISTICS	115	17	14.78
INTERNATIONAL JOURNAL OF SPEECH LANGUAGE AND THE LAW	171	25	14.62

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
JOURNAL OF LANGUAGE AND EDUCATION	145	21	14.48
LINGUISTICS VANGUARD	146	21	14.38
ARAB WORLD ENGLISH JOURNAL	952	132	13.87
JOURNAL OF SPEECH LANGUAGE AND HEARING RESEARCH	3,389	463	13.66
INTERNATIONAL JOURNAL OF BILINGUALISM	542	74	13.65
PHONETICA	862	116	13.46
JOURNAL OF CHILD LANGUAGE	1,711	224	13.09
PROCESAMIENTO DEL LENGUAJE NATURAL	107	14	13.08

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
APPLIED LINGUISTICS RESEARCH JOURNAL	177	23	12.99
LITERARY AND LINGUISTIC COMPUTING	247	32	12.96
NATURAL LANGUAGE SEMANTICS	145	18	12.41
JOURNAL OF QUANTITATIVE LINGUISTICS	258	32	12.40
BRAIN AND LANGUAGE	3,680	449	12.20
LANGUAGE AND LINGUISTICS COMPASS	178	21	11.80
LANGUAGE LEARNING	1,314	154	11.72
CORPUS LINGUISTICS AND LINGUISTIC THEORY	156	18	11.54

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
REVIEW OF COGNITIVE LINGUISTICS	182	21	11.54
LANGUAGE TEACHING RESEARCH	524	60	11.45
INTERPRETER AND TRANSLATOR TRAINER	231	26	11.26
POZNAN STUDIES IN CONTEMPORARY LINGUISTICS	322	36	11.18
MIND LANGUAGE	728	80	10.99
FIRST LANGUAGE	312	34	10.90
PRAGMATICS COGNITION	193	21	10.88
ACTA LINGUISTICA HUNGARICA	243	26	10.70

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
SYNTAX A JOURNAL OF THEORETICAL EXPERIMENTAL AND INTERDISCIPLINARY RESEARCH	150	16	10.67
COGNITIVE LINGUISTICS	443	47	10.61
JOURNAL OF RESEARCH IN APPLIED LINGUISTICS	283	30	10.60
LANGUAGE LEARNING TECHNOLOGY	352	37	10.51
APHASIOLOGY	1,999	209	10.46
DIGITAL SCHOLARSHIP IN THE HUMANITIES	389	38	9.77
PROBUS	157	15	9.55

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
INNOVATION IN LANGUAGE LEARNING AND TEACHING	168	16	9.52
INTERNATIONAL JOURNAL OF ENGLISH LINGUISTICS	786	71	9.03
TRANSLATION INTERPRETING THE INTERNATIONAL JOURNAL OF TRANSLATION AND INTERPRETING	114	10	8.77
ACROSS LANGUAGES AND CULTURES	164	14	8.54
MORPHOLOGY	106	9	8.49

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
AMERICAN JOURNAL OF SPEECH LANGUAGE PATHOLOGY	1,132	95	8.39
REVUE ROUMAINE DE LINGUISTIQUE ROMANIAN REVIEW OF LINGUISTICS	205	17	8.29
INTERCULTURAL PRAGMATICS	245	20	8.16
CHILD LANGUAGE TEACHING THERAPY	249	20	8.03
LANGUAGE AWARENESS	262	21	8.02
GESTURE	143	11	7.69
JOURNAL OF THE INTERNATIONAL PHONETIC ASSOCIATION	221	17	7.69

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
SYSTEM	1,131	87	7.69
METAPHOR AND SYMBOLIC ACTIVITY	134	10	7.46
IBERICA	203	15	7.39
LINGUA	2,551	187	7.33
ANNUAL REVIEW OF APPLIED LINGUISTICS	151	11	7.28
LINGUISTICA ANTVERPIENSIA NEW SERIES THEMES IN TRANSLATION STUDIES	138	10	7.25
TERMINOLOGY	127	9	7.09
ANNUAL REVIEW OF LINGUISTICS	101	7	6.93
JOURNAL OF LOGIC LANGUAGE AND INFORMATION	146	10	6.85

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
JOURNAL OF FRENCH LANGUAGE STUDIES	117	8	6.84
CLINICAL LINGUISTICS PHONETICS	1,480	101	6.82
LANGUAGE AND LINGUISTICS	281	19	6.76
INTERNATIONAL JOURNAL OF LANGUAGE COMMUNICATION DISORDERS	1,080	73	6.76
NORDIC JOURNAL OF LINGUISTICS	150	10	6.67
JOURNAL OF EAST ASIAN LINGUISTICS	338	22	6.51
LANGUAGE AND LITERATURE	246	16	6.50

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
3L LANGUAGE LINGUISTICS LITERATURE THE SOUTHEAST ASIAN JOURNAL OF ENGLISH LANGUAGE STUDIES	293	19	6.48
BABEL REVUE INTERNATIONALE DE LA TRADUCTION INTERNATIONAL JOURNAL OF TRANSLATION	264	17	6.44
HUMOR INTERNATIONAL JOURNAL OF HUMOR RESEARCH	607	39	6.43

Journal	Total Number of Articles	Experimental Linguistic Articles	Ratio of Experimental Linguistic Articles in %
INTERNATIONAL JOURNAL OF CORPUS LINGUISTICS	239	15	6.28
IRAL INTERNATIONAL REVIEW OF APPLIED LINGUISTICS IN LANGUAGE TEACHING	671	42	6.26
INTERNATIONAL JOURNAL OF APPLIED LINGUISTICS	163	10	6.13

References

Amrhein, V., Trafimow, D., Greenland, S., 2019. Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. The American Statistician 73, 262–270.

Barba, L.A., 2018. Terminologies for reproducible research. CoRR abs/1802.03311.

Bürkner, P.-C., 2016. Brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software 80, 1–28.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafo, M.R., 2013. Power failure: Why small sample size undermines the reliability of neuroscience. Nature reviews neuroscience 14, 365–376.

Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., Wu,

H., 2016. Evaluating replicability of laboratory experiments in economics. Science 351, 1433–1436. doi:10.1126/science.aaf0918

Camerer, C.F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johanesson, M., Kirchler, M., Nave, G., Nosek, B.A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E.-J., Wu, H., 2018. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nature 2, 637–644. doi:10.1038/s41562-018-0399-z

Campbell, D.T., 1969. Reforms as experiments. American Psychologist 24, 409.

Casillas, J.V., 2021. Interlingual interactions elicit performance mismatches not "compromise" categories in early bilinguals: Evidence from meta-analysis and coronal stops. Languages 6, 9.

Chen, J.-Y., 2007. Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). Cognition 104, 427–436.

Coles, N.A., Tiokhin, L., Scheel, A.M., Isager, P.M., Lakens, D., 2018. The costs and benefits of replication studies. Behavioral and Brain Sciences 41.

Doyen, S., Klein, O., Pichon, C.-L., Cleeremans, A., 2012. Behavioral priming: It's all in the mind, but whose mind? PloS one 7, e29081.

Fanelli, D., 2010. Do pressures to publish increase scientists' bias? An empirical support from US states data. PLoS ONE 5, e10271. doi:10.1371/journal.pone.0010271

Fidler, F., Wilcox, J., 2018. Reproducibility of scientific results.

Frank, M.C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J.K., Hannon, E.E., Kline, M., Levelt, C., others, 2017. A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. Infancy 22, 421–435.

Frank, M.C., Saxe, R., 2012. Teaching replication. Perspectives on Psychological Science 7, 600–604.

Gelman, A., Jakulin, A., Pittau, M.G., Su, Y.-S., others, 2008. A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics 2, 1360–1383.

Grahe, J.E., Reifman, A., Hermann, A.D., Walker, M., Oleson, K.C., Nario-Redmond, M., Wiebe, R.P., 2012. Harnessing the undiscovered resource of student research projects. Perspectives on Psychological Science 7, 605–607.

Hardwicke, T.E., Tessler, M.H., Peloquin, B.N., Frank, M.C., 2018. A bayesian decision-making framework for replication. Behavioral and Brain Sciences 41.

Henrich, J., Heine, S.J., Norenzayan, A., 2010. The weirdest people in the world? Behavioral and brain sciences 33, 61–83.

Isager, P.M., Aert, R. van, Bahník, Š., Brandt, M.J., DeSoto, K.A., Giner-Sorolla, R., Krueger, J.I., Perugini, M., Ropovik, I., Veer, A.E. van't, Vranka, M., Lakens, D., 2021a. Deciding what to replicate: A decision model for replication study selection under resource and knowledge constraints. Psychological Methods.

Isager, P.M., Veer, A. van't, Lakens, D., 2021b. Replication value as a function of citation impact and sample size. doi:10.31222/osf.io/knjea

Kirby, J., Sonderegger, M., 2018. Mixed-effects design analysis for experimental phonetics. Journal of Phonetics 70, 70–85.

Koole, S.L., Lakens, D., 2012. Rewarding replications: A sure and simple way to improve psychological science. Perspectives on Psychological Science 7, 608–614.

Leeuw, J.R. de, Andrews, J., Livingston, K., Franke, M., Hartshorne, J., Hawkins, R., Wagge, J., 2019. Using replication studies to teach research methods in cognitive science. Perspectives on Psychological Science 7, 600–604.

Levisen, C., 2019. Biases we live by: Anglocentrism in linguistics and cognitive sciences. Language Sciences 76, 101173.

Lindsay, R.M., Ehrenberg, A.S., 1993. The design of replicated studies. The American Statistician 47, 217–228.

Madden, C.S., Easley, R.W., Dunn, M.G., 1995. How journal editors view replication research. Journal of Advertising 24, 77–87.

Majid, A., Levinson, S.C., 2010. The language of perception across cultures, in: Talk Presented at the XXth Congress of European Chemoreception Research Organization, Symposium on "Senses in Language and Culture". Avignon, France.

Makel, M.C., Plucker, J.A., 2014. Facts are more important than novelty: Replication in the education sciences. Educational Researcher 43, 304–316.

Makel, M.C., Plucker, J.A., Freeman, J., Lombardi, A., Simonsen, B., Coyne, M., 2016. Replication of Special Education Research: Necessary but Far Too Rare. Remedial and Special Education 37, 205–212.

Makel, M.C., Plucker, J.A., Hegarty, B., 2012. Replications in psychology research: How often do they really occur? Perspectives on Psychological Science 7, 537–542.

Marsden, E., Morgan-Short, K., Thompson, S., Abugaber, D., 2018. Replication in Second Language Research: Narrative and Systematic Reviews and Recommendations for the Field. Language Learning 68, 321–391. doi:gc3h3b

Martin, G.N., Clarke, R.M., 2017. Are Psychology Journals Anti-replication? A Snapshot of Editorial Practices. Frontiers in Psychology 8. doi:10.3389/fpsyg.2017.00523

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.

Morey, R.D., Kaschak, M.P., Díez-Álamo, A.M., Glenberg, A.M., Zwaan, R.A., Lakens, D., Ibáñez, A., García, A., Gianelli, C., Jones, J.L., others, 2021. A pre-registered, multi-lab non-replication of the action-sentence compatibility effect (ACE). Psychonomic Bulletin & Review.

Mueller-Langer, F., Fecher, B., Harhoff, D., Wagner, G.G., 2019. Replication studies in economics—how many and which papers are chosen for replication, and why? Research Policy 48, 62–83.

Nieuwland, M.S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Zu Wolfsthurn, S.V.G., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D.J., Rousselet, G.A., Ferguson, H.J., Bush-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, M.E., Donaldson, D.I., Kohút, Z., Rueschemeyer, S.-A., Huettig, F., 2018. Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. eLife

7, e33468. doi:10.7554/eLife.33468.001

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., others, 2015. Promoting an open research culture. Science 348, 1422–1425.

Nosek, B.A., Spies, J.R., Motyl, M., 2012. Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. Perspectives on Psychological Science 7, 615–631.

Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science 349. doi:10.1126/science.aac4716

Papesh, M.H., 2015. Just out of reach: On the reliability of the action-sentence compatibility effect. Journal of Experimental Psychology: General 144, e116–e141. doi:10.1037/xge0000125

Roettger, T.B., 2019. Researcher degrees of freedom in phonetic research. Laboratory Phonology: Journal of the Association for Laboratory Phonology 10.

Roettger, T.B., Baer-Henney, D., 2019. Toward a replication culture: Speech production research in the classroom. Phonological Data and Analysis 1, 1–23.

Rosenthal, R., 1990. Replication in behavioral research. Journal of Social Behavior and Personality 5, 1.

Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological science 22, 1359–1366.

Stack, C.M.H., James, A.N., Watson, D.G., 2018. A failure to replicate rapid syntactic adaptation in comprehension. Memory & Cognition 46, 864–877. doi:10.3758/s13421-018-0808-6

Westbury, C., 2018. Implicit sound symbolism effect in lexical access, revisited: A requiem for the interference task paradigm. Journal of Articles in Support of the Null Hypothesis 15, 1–12.

Winter, B., Grice, M., 2021. Independence and generalizability in linguistics. Linguistics 59, 1251–1277.

Yarkoni, T., 2019. The generalizability crisis. Behavioral and Brain Sciences 1–37.

Zwaan, R.A., Etz, A., Lucas, R.E., Donnellan, M.B., 2018. Making replication mainstream. Behavioral and Brain Sciences 41, E120. doi:10.1017/S0140525X17001972