Remembering Generalizations: Memory Mechanisms Underlying the Generic Recall Bias

Griffin Pion[1], Sophie H. Arnold[2], Elliot Schwartz[1], Julia Johnson[2], Eric Mandelbaum[1,3] & Marjorie Rhodes[2]

[1]Cognitive Science Program and Department of Philosophy, CUNY Graduate Center

[2]Department of Psychology, New York University

[3]Department of Psychology, CUNY Graduate Center; Departments of Psychology and Philosophy, Baruch College

Abstract

Humans express generalizations using both generics (e.g., "Dogs bark") and explicit quantifiers (e.g., "All dogs bark"). Quantified statements are systematically misremembered as generics more than vice versa, a phenomenon known as the Generic Recall Bias. Yet, the memory processes underlying this bias remain unclear. We investigate whether this phenomenon arises during memory encoding, retention, or retrieval. Here, we taught adults (N = 1,189) generic or quantified statements about a novel social category, and assessed recall immediately and one week later. Replicating and extending prior findings, we found that "all", "most", and "some" statements were each significantly misremembered as generics. Critically, longitudinal analyses revealed that misrecall patterns were stable across time and consistent within items, supporting an encoding-based account: quantified information is often encoded without its specific proportional information from the outset. Overall, these results clarify the memory mechanisms underlying human generalization.

Keywords: generics; quantifiers; memory; generalization

Introduction

Humans have the capacity to form concepts representing an endless range of categories, including inanimate objects, biological kinds, functional kinds, and abstract ideas. Furthermore, we are able to use these representations to form *generalizations* about the categories as a whole, in a way that often goes far beyond our experience. For example, upon encountering a small number of cuddly puppies, a child may infer that puppies are *generally* cuddly, extending this trait to never before seen puppies.

Adults possess multiple linguistic means of expressing generalizations. For instance, *generics* express generalizations about kinds as a whole (e.g., "Puppies are cuddly") without specifying exactly what proportion of the kind's members possess the predicated property. In this respect, generics differ from another way of expressing generalizations, *quantified statements* (e.g., "All puppies are cuddly," "Most puppies are cuddly"); unlike generics, quantified statements tend to explicitly encode how many members of a kind (e.g., puppies) satisfy the predicate in question (e.g., the majority).

In some respects, generics appear to be a privileged way of expressing generalizations, and indeed may reflect a cognitively default way of forming generalizations (Leslie, 2007, 2008). First, developmental studies have shown that across multiple, unrelated languages, young children understand generics at a younger age than quantified expressions (Gelman & Tardif, 1998; Gelman et al., 2016; Hollander et al., 2002; Mannheim et al., 2010), and parents commonly use generics but not quantifiers in child-directed speech (Gelman et al., 2000, 2014). Beyond child behavior, Leslie et al. (2011) demonstrate that adults often treat quantified expressions as generics in reasoning tasks. Additionally, as we discuss in detail below, adults and children asymmetrically misremember quantified statements as generics, an effect known as the *Generic Recall Bias* ("GRB"). Taken together, the existing literature supports the view that generic statements play a privileged role in expressing generalizations about categories.

The Generic Recall Bias

Both children and adults misrecall explicitly quantified statements as generics more than the other way around. For example, Leslie and Gelman (2012) asked adults and children to memorize novel facts about animals, manipulating whether those facts were presented generically ("Bees have five eyes") or quantificationally ("[All/most/some] bees have five eyes"). Adults misrecalled "all" and "most" quantificational facts as generic more than vice versa, and children misrecalled every type of quantificational fact as generic more than vice versa. Subsequent work has replicated this finding and extended it to novel social categories and additional quantifiers (Gelman et al., 2019; Sutherland et al., 2015). Furthermore, additional studies have ruled out competing explanations, such as the fact that generics involve fewer words than the quantificational statements (Gelman et al., 2016).

While the GRB is well documented, its underlying mechanisms are not yet well understood. We propose and test concrete mechanisms that could underlie the GRB. We use a novel experimental design with a two-session recall paradigm (Fig. 1). We taught participants a series of statements about a novel social category ("Zarpies") using either generic statements or statements with one of three quantifiers (ALL, MOST, SOME). We then tested participants on their recall memory. Half of our participants were tested immediately and again one week later. The other half was tested exclusively after one week. This left us with three memory test conditions: immediate, delay, delay-only. (Details are included below.)

This design allows us to pinpoint where the GRB arises. To accurately remember any given statement, you must (i) encode a representation from working memory into long-term memory, (ii) retain that representation over time, (iii) suc-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

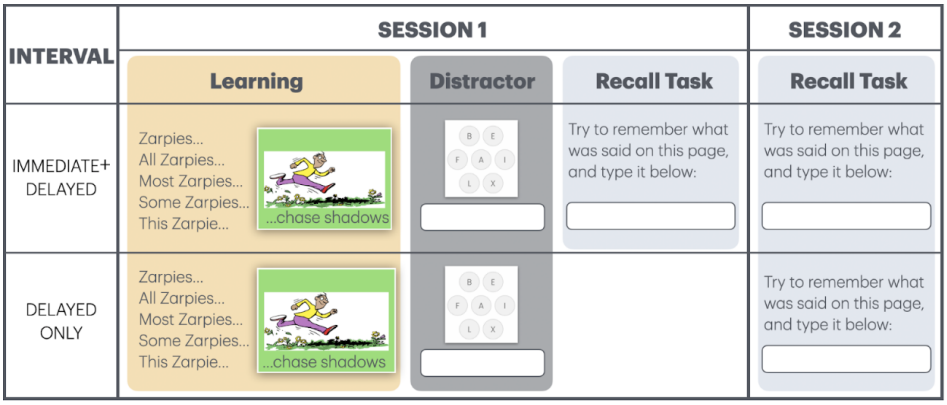

Figure 1: *Schematic of our two-session recall paradigm.* During a learning phase, participants hear 16 sentences about a novel social category, Zarpies, in one of five language conditions (GENERIC, ALL, MOST, SOME, SPECIFIC); they also hear 16 filler sentences. There were no words on the screen during the learning phase. Then, after a brief distractor, participants in the IMMEDIATE+DELAYED interval are given a recall task for what they remember about Zarpies. One week later, participants in both the IMMEDIATE+DELAYED and DELAYED ONLY interval return to take the memory test. The recall task included the image seen during the learning phase with the prompt in the figure above.

cessfully retrieve the representation when prompted (Tulving & Thomson, 1973). The GRB could arise from lapses at any (or potentially multiple) of these stages, leading to three hypothetical mechanisms:

Encoding: When people hear a quantified statement ("All Zarpies chase shadows"), they tend to encode generic information ("Zarpies chase shadows") into long-term memory.

Retention: While people successfully encode a quantified statement ("All Zarpies chase shadows") into long-term memory, they forget the specific quantificational information over time and, as a result, generic information ("Zarpies chase shadows") remains.

Retrieval: When prompted to recall a quantified statement successfully preserved in long-term memory ("All Zarpies chase shadows"), people sometimes fail to retrieve the specific quantificational information and, as a result, report generic information ("Zarpies chase shadows").

If the GRB is due to lapses during **encoding**, then participants who misrecall a quantifier as a generic during an immediate memory test should also misrecall it one week later. In other words, **encoding** predicts a trial-level correlation between items misrecalled as generic. Conversely, if the GRB is due to lapses in **retrieval**, then misrecalling a generic immediately should not predict whether it is misrecalled after a delay. Finally, if the GRB is due to **retention**, we should expect participants to misremember quantifiers as generics more often when tested one week later than when tested immediately. Thus, each memory hypothesis makes a distinct prediction within our longitudinal design.[1]

Experiment

Participants

The final sample consisted of 1,189 U.S. adults recruited through Prolific (M_{age} = 40.2 years, SD_{age} = 13.3 years; 648 women, 534 men, 5 prefer not to say, 2 missing). An additional 511 participants were excluded for the following preregistered reasons: completing fewer than half of the dependent measures in either session 1 (participants in the IMMEDIATE+DELAY interval only; see Procedure below; n= 5) or session 2 (participants in the DELAY ONLY interval; n = 19); having no useable memory data from session 1 (for participants in the IMMEDIATE+DELAY interval; n = 157; an additional 137 participants' data were excluded from second session for this same reason, these participants are still included in the first session sample; see Procedure below for additional details); or having no useable memory data from session 2 (for participants in the DELAY ONLY condition; n = 330). While this exclusion rate is high, this is due to our use of an open-ended free recall task and a longitudinal design, both of which make data quality and retention harder than a typical Prolific study. The research questions, hypotheses, and methods were preregistered on OSF (https://osf.io/wk8su/overview?view_only=6557ceec0b1f4f16ba93c8f8e6dc2793).

The present data are part of a larger project where participants were assigned to one of five between-subjects language conditions (GENERIC, ALL, MOST, SOME, SPECIFIC) and one of two between-subjects intervals (IMMEDIATE+DELAY, DELAY ONLY).[2] These intervals represent three memory conditions: participants in the IMMEDIATE+DELAY responded in both ses-

[1] In principle, these hypotheses are compatible: all three could contribute to the GRB. However, to generate unique predictions, our design pits **encoding** against **retrieval**. Thus, we are able to detect **retention** and either **encoding** or **retrieval**. We provide some supplemental analyses below suggesting that even between **encoding** and **retrieval**, only the former contributes to the GRB.

[2] We use small caps for conditions, and italics for response types. This allows us to easily distinguish between those in, e.g., the GENERIC language condition from those who respond with a *generic*.

sion one (i.e., IMMEDIATE) and session two (i.e., DELAY) while those in the DELAY ONLY condition only responded to questions in session two. Given that the focus of the present paper is on the GRB, which concerns a memory asymmetry between generics and quantified statements, we focus on participants in the GENERIC, ALL, MOST, and SOME conditions, not the SPECIFIC condition. We include mention of the SPECIFIC condition to clarify the study procedure and data processing, but do not use any data from that condition in the present study. Therefore, all analyses reported below were conducted on the 977 participants in the GENERIC and QUANTIFIERS conditions (M_{age} = 42.7 years, SD_{age} = 13.0 years; 542 women, 429 men, 4 prefer not to say, 2 missing).

Procedure

The present experiment consisted of 2 sessions. In the first session, all participants completed a learning phase where they heard 16 statements about a novel group ("Zarpies") consistent with the language condition they were assigned to (e.g., GENERIC: "Zarpies chase shadows"). There were 16 filler statements about animals and artifacts using the four other language forms so all participants heard all language forms. Then, they completed a short distractor task (three NYTimes Spelling Bee puzzles) to reduce the possibility for rehearsal. Participants in the IMMEDIATE+DELAY interval then completed the recall task (see below). Participants in the DELAY ONLY condition stopped the session after the distractor task. All participants were invited back one week later to complete session two, where they completed the recall task.

Recall Task Participants were prompted to freely recall what they heard for each of the 16 statements, presented in a random order. These freely recalled responses were then coded into one of six categories based on the linguistic form of the recalled statement: *generic, all, most, some, specific,* or *other*. For *all, most,* and *some* statements, the statement was coded as the corresponding quantifier if the statement started as the quantifier (e.g., *all*) and ended with a statement about the novel category, regardless of the accuracy for the statement content (e.g., "all Zarpies run fast"). Statements were coded as *specific* if they started with the same construction as presented to them in the learning phase (e.g., "this Zarpie"), regardless of the statement content. Statements were coded as *generic* if they started with a bare plural. While the name of the novel category was Zarpies, we allowed any group bare plural to count (e.g., zuzus, garbos, scwigglies) as we were not interested in memory for the novel category name. All other recalls were coded as *other* (e.g., "I don't know", "oh no, where did I leave my car"). Other responses were removed prior to analyses.

We created two scores for our questions of interest. First, as an initial look at our data, we created a single proportion accuracy score for each participant. This score represents the proportion of accurately recalled statements (e.g., proportion of statements recalled as *generic* in the GENERIC condition) such that 1 = *only recalled statements accurately* and 0 = *never recalled statements accurately*. Second, to address questions about the GRB, we created proportion memory mismatch scores. For the quantifier conditions (ALL, MOST, SOME), this score represents the proportion of *generics* recalled out of total codeable memory errors. For example, for the ALL condition, this is the proportion of *generics* recalled out of that condition's errors (i.e., responses other than *all*; *generic, most, some*, or *specific* recalls). For the GENERIC condition, we created three proportions, one for each quantifier response. We then looked at the proportion of recalling that quantifier out of total codeable errors. For example, we calculated the proportion of *all* responses from individuals in the GENERIC condition, out of that condition's errors (i.e., responses other than *generic*; *all, most, some*, or *specific* recalls).

Results

H0: Recall Accuracy by Language and Memory Conditions We examined how language condition and memory condition jointly shape participants' ability to accurately recall the statements they heard. Given the nature of the dependent variable (proportion), we constructed a mixed beta regression with proportion accuracy regressed on the language condition (4 levels: GENERIC, ALL, MOST, SOME), memory condition (3 levels: IMMEDIATE, DELAYED, DELAY-ONLY), and the interaction between the two condition variables with a random intercept for participant.

There was a significant main effect of memory condition such that participants were most accurate when tested immediately ($M = 0.66$), followed by delayed ($M = 0.55$), and least accurate when tested only at a delay ($M = 0.48$) (main effect of memory condition: $\chi^2(6)= 24.73$, $p < .001$; all contrast ps $< .003$). There was also a significant effect of language condition such that participants in the GENERIC condition were most accurate ($M = 0.71$), followed by those in the SOME condition ($M = 0.61$), then the ALL condition ($M = 0.52$), and lastly the MOST condition ($M = 0.40$) (main effect of language condition: $\chi^2(3)= 145.94, p < .001$; all contrast ps $< .002$).

These main effects were further qualified by a two-way interaction between language and memory conditions ($\chi^2(6)= 24.73, p < .001$; see Fig. 2). This interaction was driven largely by the GENERIC condition. For the GENERIC condition, participants' accuracy remained stable across memory conditions (contrast ps $< .71$). All other language conditions showed some drop in accuracy between either the IMMEDIATE and DELAY ONLY (SOME condition, $p < .001$) or between both the IMMEDIATE and DELAY ONLY as well as the IMMEDIATE and DELAYED conditions (MOST and ALL conditions, contrast ps $< .001$).

H1: Generic Recall Bias (GRB) We examined whether participants in a quantifier condition were more likely to recall a *generic* than participants were to recall the matched quantifier. Given the nature of the dependent variables (proportions), we constructed mixed beta regressions with proportion memory mismatch regressed on the language condition (1 = GENERIC, 0 = QUANTIFIER), memory condition (IMMEDIATE,

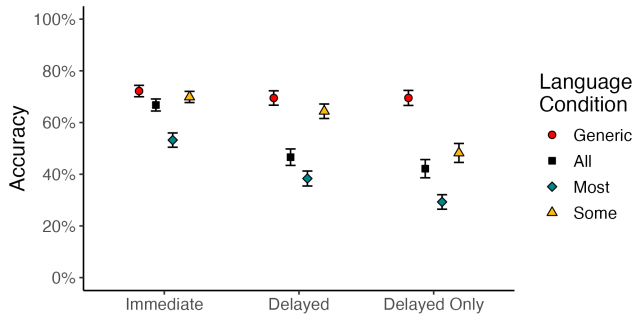

Figure 2: *Recall Accuracy by Language and Memory Conditions.* Note: Shapes are predicted means; error bars are standard errors.

DELAYED, DELAY-ONLY), and a significant two-way interaction between the two condition variables with a random intercept for participant. There were separate analyses for each quantifier by generic comparison: GENERIC VS. ALL, GENERIC VS. MOST, and GENERIC VS. SOME. The GRB would be supported by a significant main effect of language condition (reported below).

Replicating and extending prior work, we found evidence for Generic Recall Bias for all quantifier comparisons: 28% of memory errors in the GENERIC condition were *all* whereas 69% of the memory errors in the ALL condition were generic ($\chi^2(1)= 32.59, p < .001$), 5% of the memory errors in the GENERIC condition were *most* whereas 47% of the memory errors in the MOST condition were *generic* ($\chi^2(1)= 56.02, p < .001$), and 42% of the memory errors in the GENERIC condition were *some* whereas 59% of the memory errors in the SOME condition were *generic* ($\chi^2(1)= 5.01, p = .025$).

H2: Memory Mechanisms for the GRB If the GRB is due in part to **retention** (i.e., an increase of misrecalling quantified statements as generic over time), we should see a significant interaction between language condition and memory condition in the model outlined above. We found no such evidence of moderation over time for any of the three comparisons (GENERIC VS. ALL language × memory: $\chi^2(2)= 3.13, p = .21$; GENERIC VS. MOST language × memory: $\chi^2(2)= 0.14, p = .93$; GENERIC VS. SOME language × memory: $\chi^2(2)= 1.48, p = .47$, suggesting that the GRB is stable over a one-week delay (see Fig. 3).

To adjudicate between whether the GRB is due to **encoding** or **retrieval**, we examined trial-level correlations in recall responses. The results supported **encoding** most clearly: there was high trial-level correlation between recall at the first and second sessions (92% agreement; 4,470 matched trials out of 4,860), and across the four language conditions, participants who correctly recalled the language form in the first session always recalled correctly in the second session (3,579/3,579 trials). As additional evidence against **retrieval**, those participants who misrecalled a quantified statement as *generic* in the first session never recalled the correct quantifier in the second session (0/615 trials).

Discussion

These results refine our understanding of the Generic Recall Bias (GRB), the finding that quantified statements are systematically misremembered as generics more than vice versa. First, we found evidence for the GRB in a wider range of quantifiers than previously documented, including "some", which had not been previously found in adults.[3] Furthermore, our experiment clarifies the stage in the memory process at which the GRB arises. Our data are consistent with the effect being driven by **encoding**, the hypothesis that the GRB arises because people initially encode quantified statements as generics in long-term memory (more than vice versa). Conversely, the data are inconsistent with both **retrieval** (the hypothesis that the GRB arises because people fail to retrieve stored proportion information) or **retention** (the hypothesis that the GRB arises because people fail to retain stored proportion information over time).

First, we distinguished **encoding** and **retrieval** by looking at trial-level correlations. According to a **retrieval**-based hypothesis, the GRB arises because people sometimes fail to retrieve the proportional information successfully encoded by a quantified statement, and so report a generic which does not encode this information. Accordingly, a **retrieval**-based hypothesis does not predict a trial-level correlation between responses at IMMEDIATE and DELAYED tests: if someone is failing to retrieve stored proportional information from long-term memory immediately, they may still successfully retrieve the stored proportional information one week later. In contrast, an **encoding**-based hypothesis predicts a high trial-level correlation between items recalled at IMMEDIATE and DELAYED tests: if someone immediately stores a quantified statement as a generic, then they should tend to report the same misremembered item at both tests. Our data clearly supported the first hypothesis: on 92% of trials where participants were tested twice, they gave the same response. In further contradiction of a **retrieval** hypothesis, we found that participants in a quantifier condition who misrecalled a generic at **immediate** test never recalled the correct quantifier one week later (0/615 trials). Lending additional support to an **encoding** hypothesis, we found that participants who recalled correctly in the first session always recalled correctly in the second session (3,579/3,579 trials).

Second, a **retention**-based hypothesis holds that the GRB arises because proportional information is forgotten over time; once forgotten, individuals report generics, which lack such information. This hypothesis uniquely predicts that the magnitude of the GRB should increase over time. However, we found no interaction between language condition and memory condition to support this hypothesis. Thus, we conclude that

[3] Leslie and Gelman (2012) found a GRB for "all" and "most" for adults, Sutherland et al. (2015) showed the GRB in "all" and "many", and Gelman et al. (2016) showed it in the Spanish translations of "all" (="todos") and "many" (="muchos").

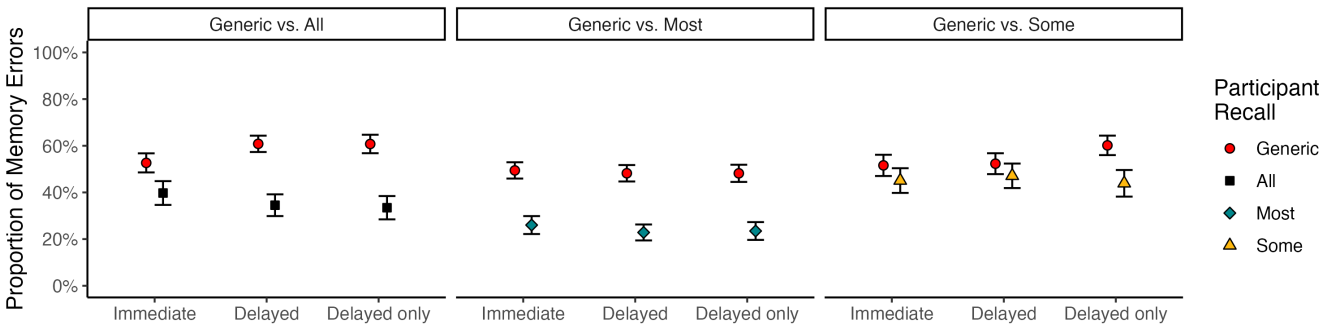

Figure 3: *Adult recall mismatch by generic-quantifier comparison*. Note: Shapes represent predicted proportion mismatch (e.g., for GENERIC vs. ALL, recall *generic* in ALL condition and recall *all* in GENERIC condition); error bars are standard errors. "Immediate" (N = 617), "Delayed" (N = 456) and "Delayed only" (N = 360) correspond to responses in the memory test sessions for the two memory intervals, IMMEDIATE+DELAYED (in which participants were tested twice, one week apart) and DELAY ONLY (in which participants were tested once, a week after learning). All 456 participants in the "Delayed" condition (session 2) were the same participants from the "Immediate" condition (session 1).

the GRB arises solely due to **encoding**: individuals misrecall quantifiers as *generics* more than vice versa because they initially encode quantifiers into memory as generic information. While it was antecedently possible that misrecalling quantifiers as *generics* was driven by multiple underlying mechanisms on different occasions, the results surprisingly only support one across the board. Thus, on our view, quantifiers are not forgotten from long-term memory; rather, they are never represented at all.

Although we have explained the GRB as arising from a memory encoding process, an alternative explanation is that the GRB arises through a truth-conditional hedging strategy: when individuals forget proportional information but are required to recall a statement, they may respond with a generic rather than a quantifier because generics are less proportionally committal, and so may be more likely to communicate what is true (Goldsmith et al., 2002). While intuitive, this strategy makes predictions that are inconsistent with existing data: participants in each language condition should report sentence types with truth-conditionally weaker proportional information. In particular, participants in the ALL condition should report generalizations involving *generics, most, some,* and *specifics*; participants in the MOST condition should report generalizations involving *generics, some,* and *specifics*; and so on. However, these predictions are not generally borne out. For example, participants in the ALL condition rarely report generalizations with *most* (6% of misrecalled statements), *some* (18%), or *specifics* (6%), especially compared to generics (69%). Beyond making predictions that are inconsistent with the data, this alternative explanation cannot explain the full suite of findings. In particular, it does not predict the GRB for "some". Given that existential statements do not entail generics (e.g., *some Canadians are left-handed* does not entail *Canadians are left-handed*), one would not be ensuring truthfulness by recalling a generic having heard an existential. Yet, participants in the SOME condition show the GRB, often misrecalling generics. Thus, given that this truth-conditional hedging strategy makes inaccurate predictions and does not explain the full range of effects, we do not find it persuasive.

In one notable respect, our findings differ from previous examinations of the GRB. Leslie and Gelman (2012) found that while children show a GRB for "all", "most", and "some", adults only show a GRB for the first two quantifiers. Here, however, we found that even adults can show the GRB for "some". These authors argued that this contrast was explained by the fact that adults, unlike children, possess the *but not all* scalar implicature for "some", which blocks a kind-level generalization. Our results, however, suggest that the GRB can occur for "some", even in adults.

The divergence between our results and Leslie & Gelman's may be explained by two differences in experimental design. First, Leslie & Gelman varied language-type as a within-subject factor (e.g., giving a single participant "some" and "generic" statements), while our study varied language-type between-subjects.[4] It is possible that alternating language-types for a single individual makes the *but not all* scalar implicature triggered by "some" more salient. If so, participants in our study would be less likely to draw that implicature compared to those in Leslie & Gelman's study. Second, we used statements about a novel social category ("Zarpies"), whereas Leslie & Gelman used statements about familiar biological kinds (e.g., "bears"). Novel kinds may be less likely to generate implicatures as speakers are assumed to be less knowledgeable about the kind's properties (e.g., a speaker may be assumed to know that some Zarpies chase shadows but be ignorant as to whether they all do). If these two design factors do mitigate the implicature, we would predict exactly the pattern seen in our data: we should see a GRB for SOME

[4]To be clear, our experiment included fillers of different language-types. However, non-condition language-types were not about novel kinds, and so there would have been no salient quantificational contrast for critical, Zarpie trials.

(reflecting the trials on which the implicature is not drawn), but it should be smaller than that of the other quantifiers (reflecting the trials on which the implicature is drawn).

Open Questions

While our results suggest that the GRB arises from errors in memory encoding, they do not directly speak to the directionality of these encoding errors; that is, they do not explain why memory of generalizations is biased towards generics rather than quantifiers. We see three compatible possibilities. First, it may be that generic statements express simpler mental representations than quantified statements do. For example, perhaps understanding a quantifier requires computing relations between sets (e.g., the Zarpies are a subset of the shadow-chasers), whereas understanding a generic only involves predicating a property to a kind (e.g., predicating shadow-chasing of Zarpies). Participants may tend to encode information in representationally simpler formats and therefore erroneously encode generics as quantifiers (Amalric et al., 2017; Quilty-Dunn et al., 2023). Second, participants could pragmatically infer that—although the statements employ quantifiers—they are being used to express properties of kinds (Alba & Hasher, 1983, p.209). While the statement "All Zarpies chase shadows" literally means every single Zarpie does so, it is unlikely that a speaker would have actual knowledge about every instance. Instead, a speaker may be interpreted as attempting to communicate information about a kind, which is more naturally expressed as a generic. Finally, a third possibility situates these effects within schema-based accounts of memory, according to which encoding is guided by expectations about what parts of an input is relevant (see Alba and Hasher (1983) for a review; see also, Koriat et al. (2000)). For instance, participants may be guided by a schema of what types of information are relevant when learning about social categories, which may lead them to omit the proportional information conveyed by quantifiers during schema theory's "selection" process. Further research is required to address these possibilities.

Future Directions

One important direction for future research is to examine the mechanisms underlying the GRB as it emerges across development. Prior work shows that 3-5-year-old children exhibit an even more robust GRB than adults (Gelman et al., 2016; Leslie & Gelman, 2012). This raises the question of whether children's GRB is driven by the same **encoding** asymmetry observed here, by asymmetries at a different stage, or by a combination of mechanisms. Addressing this question bears on a broader issue in cognitive development: whether memory mechanisms are stable over ontogeny, with changes in performance reflecting fine-tuning of existing mechanisms, or whether the operative mechanisms themselves change. We are currently collecting child data ($N \approx 750$) using an adapted version of our paradigm. In this ongoing work, we are extending prior developmental investigations by examining a wider age range (4-9-year olds), allowing us to track developmental changes in both the magnitude of the GRB and the mechanisms that give rise to it.

Another important direction for future research is to probe the GRB using measures of recognition memory (Tulving & Thomson, 1973; Tulving & Watkins, 1973; Yonelinas et al., 2022). While recall, which we tested here, requires participants to explicitly report previously encountered information, recognition merely involves identifying information as familiar. Recognition memory is typically more robust and enduring than explicit recall (Larzabal et al., 2018; Mitchell, 2006). In future work, we plan to implement a recognition-based variant of the present paradigm. Combined with our two-session design, this approach will allow us to determine whether there is a recognition GRB, and if so, whether it arises at **encoding**, **retention**, or **retrieval**. Answering these questions bears directly on a broader theoretical motivation to understand how generalizations are stored in long-term memory. Many memories—perhaps most—cannot be articulated explicitly but can nevertheless be recognized. As a result, recognition is perhaps a more comprehensive measure of memory for generalizations, and at the very least an important source of complementary evidence. Thus, extending our research into recognition memory will give a fuller picture of how generalizations are remembered.

Broader implications

Our results have implications for the nature of human generalization. Over the past few decades, an influential research program has advanced the position that generics express a cognitively privileged way of generalizing about categories, a view known as the Generics-as-Default hypothesis (Lerner & Leslie, 2016). In showing that the GRB holds for even more quantifiers than previously known — indeed, every quantifier we tested — the present study provides further evidence that generics express default generalizations.

Beyond generalization, the structure of the present study could also be used to discover the underlying mechanisms driving other memory asymmetries. Memory asymmetries similar to the GRB have been documented in other domains. A well-known example comes from psycholinguistics: people are more likely to misremember a negated statement (e.g., "The dog is *not* barking") as its affirmative counterpart ("The dog is barking") than to mistakenly insert a negation where none was present (Cornish & Wason, 1970; Maciuszek & Polczyk, 2017; Meyer, 1975). These asymmetries have been observed for decades, yet the mechanisms that underpin them remain poorly understood. Our longitudinal design offers a template to investigate these phenomena and to determine their underlying causes. One interesting possibility is that many memory asymmetries will pattern the same. If this is true, then there may be a unified cognitive principle governing how certain types of information are preferentially preserved or lost. Such a finding would illuminate a basic structural feature of human memory.

Acknowledgments

We would like to thank Spencer Caplan, Susan Carey, Sandeep Prasada, Jared Vasil, Marianna Zhang, and the members of the CDSC Lab for their helpful comments and feedback.

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203.

Amalric, M., Wang, L., Pica, P., Figueira, S., Sigman, M., & Dehaene, S. (2017). The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLoS computational biology*, *13*(1), e1005273.

Cornish, E. R., & Wason, P. C. (1970). The recall of affirmative and negative sentences in an incidental learning task. *Quarterly Journal of Experimental Psychology*, *22*(2), 109–114.

Gelman, S. A., Hollander, M., Star, J., & Heyman, G. D. (2000). The role of language in the construction of kinds. In *Psychology of learning and motivation* (pp. 201–263, Vol. 39). Elsevier.

Gelman, S. A., Leslie, S.-J., Gelman, R., & Leslie, A. (2019). Do children recall numbers as generic? a strong test of the generics-as-default hypothesis. *Language Learning and Development*, *15*(3), 217–231.

Gelman, S. A., Tapia, I. S., & Leslie, S.-J. (2016). Memory for generic and quantified sentences in spanish-speaking children and adults. *Journal of Child Language*, *43*(6), 1231–1244.

Gelman, S. A., & Tardif, T. (1998). A cross-linguistic comparison of generic noun phrases in english and mandarin. *Cognition*, *66*(3), 215–248.

Gelman, S. A., Ware, E. A., Kleinberg, F., Manczak, E. M., & Stilwell, S. M. (2014). Individual differences in children's and parents' generic language. *Child development*, *85*(3), 924–940.

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General*, *131*(1), 73.

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental psychology*, *38*(6), 883.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual review of psychology*, *51*(1), 481–537.

Larzabal, C., Tramoni, E., Muratot, S., Thorpe, S. J., & Barbeau, E. J. (2018). Extremely long-term memory and familiarity after 12 years. *Cognition*, *170*, 254–262.

Lerner, A., & Leslie, S.-J. (2016). Generics and experimental philosophy. *A companion to experimental philosophy*, 404–416.

Leslie, S.-J. (2007). Generics and the structure of the mind. *Philosophical perspectives*, *21*, 375–403.

Leslie, S.-J. (2008). Generics: Cognition and acquisition. *Philosophical review*, *117*(1), 1–47.

Leslie, S.-J., & Gelman, S. A. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive psychology*, *64*(3), 186–214.

Leslie, S.-J., Khemlani, S., & Glucksberg, S. (2011). Do all ducks lay eggs? the generic overgeneralization effect. *Journal of Memory and Language*, *65*(1), 15–31.

Maciuszek, J., & Polczyk, R. (2017). There was not, they did not: May negation cause the negated ideas to be remembered as existing? *PLoS One*, *12*(4), e0176452.

Mannheim, B., Gelman, S. A., Escalante, C., Huayhua, M., & Puma, R. (2010). A developmental analysis of generic nouns in southern peruvian quechua. *Language Learning and Development*, *7*(1), 1–23.

Meyer, D. E. (1975). Long-term memory retrieval during the comprehension of affirmative and negative sentences. *Studies in long-term memory. London: John Wiley & Sons*, 289–312.

Mitchell, D. B. (2006). Nonconscious priming after 17 years: Invulnerable implicit memory? *Psychological Science*, *17*(11), 925–929.

Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2023). The best game in town: The reemergence of the language-of-thought hypothesis across the cognitive sciences. *Behavioral and Brain Sciences*, *46*, e261.

Sutherland, S. L., Cimpian, A., Leslie, S.-J., & Gelman, S. A. (2015). Memory errors reveal a bias to spontaneously generalize to categories. *Cognitive science*, *39*(5), 1021–1046.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*(5), 352.

Tulving, E., & Watkins, M. J. (1973). Continuity between recall and recognition. *The American Journal of Psychology*, 739–748.

Yonelinas, A. P., Ramey, M. M., Riddell, C., Kahana, M., & Wagner, A. (2022). Recognition memory: The role of recollection and familiarity. *AD Kahana, MJ, & Wagner (Ed.), The Oxford handbook of human memory*.