A Weak Bias Against Strong Synonymy?

Aja Marie Altenhof (s2009154@ed.ac.uk)
University of Edinburgh

Abstract

Is there a cognitive bias against absolute synonymy? The current work explored this question via a miniature mini-artificial language experiment featuring iterated learning, which can amplify weak cognitive biases, as languages are sifted through multiple adult learners. Participants were taught two novel synonymous verbs in positive or negative sentences. Afterwards, they had to generalize the words to new positive or negative sentences. These sentences then served as the input sentences for the next participant in the diffusion chain and so on. Despite inconsistent input with regard to positive or negative meaning, participants differentiated the verbs. More than half of participants strongly differentiated them, specializing at least one term. However, transmission did not increase differentiation overall, suggesting that a bias against synonymy may encourage a minimally distinctive difference, not necessarily a systematic one, between synonyms.

Keywords: synonymy; semantic differentiation; iterated learning; language evolution

Introduction

While synonymy is extremely common, absolute synonymy — where two words could be substituted for one another in any context, without a change to truth value, communicative impact, or connotational meaning — seems rare (Cruse, 1986). The synonymous pair *amazing* : *spectacular* superficially appears to meet this requirement. Swapping them for one another in sentences like "I had an ______ night!" doesn't change the intended meaning of either. Nevertheless, both spectacular and amazing carry meanings in addition to "really good" or "impressive". Amazing things are astonishing and shocking, whereas something spectacular is large or obvious. A football team could experience a "spectacular loss", but given a poor record, it might not be an amazing one. Conversely, learning that ants can't be microwaved is pretty amazing, but not certainly not spectacular. Most synonymous words pattern in this manner and are deemed near-synonyms, denoting largely the same meaning but differing in their stylistic and semantic effect, often across multiple dimensions simultaneously (DiMarco, Hirst, & Stede, 1993).

One hypothesis suggests that there could be a general cognitive bias against synonymy (Altenhof & Roberts, under revision). When presented with one unfamiliar and one familiar object, children will apply a new label to the unfamiliar object; learners generally assume a single object has a single label (Markman & Wachtel, 1988; E. Clark, 1992; Lewis, Cristiano, Lake, Kwan, & Frank, 2020). Adults also abide by mutual exclusivity (Savage & Au, 1996) despite prior knowledge of synonymy — that objects can (and often do) have more than one label.

Alternatively, near-synonyms could arise through distributional sampling effects as differences in the perceived contextual distribution of candidate synonyms become amplified and lexicalized during learning. This is best understood through the historical example of English words for animals and their respective meat (e.g., *sheep* : *mutton*, *cow* : *beef*, *pig* : *pork*, etc.) Following the Norman conquest of England, French became the language of power and the rich spoke about the food they ate using the appropriate French forms. The peasants, who labored in the fields and raised these animals, continued to use their native Old English names. Eventually, this distributional bias — mostly French in food contexts and mostly English in farming ones — became a semantic one as the words were pushed farther apart over time (Clark, Parikh, & Ryant, 2007).

Adults are highly sensitive to co-occurrence statistics, even for relatively infrequent mappings, maintaining multiple hypotheses about the meaning of a word and assigning likelihoods to possible candidates (Vouloumanos, 2008). In a now seminal study, Hudson Kam and Newport (2005) exposed both adults and young children to an artificial language where particles probabilistically followed nouns. During test, adult learners recreated the unpredictable variation found in their input, matching the particle usage frequency in their output. Children, in contrast, regularized — imposing patterns that reduced the variation in their input.

Yet under some circumstances, adults will regularize. Samara, Smith, Brown, and Wonnacott (2017) taught both children and adults on one of three artificial languages where nouns were obligatorily followed by meaningless particles that were conditioned on speaker identity, probabilistically, or not at all. Though probability matching was still the dominant behavior for adults, some successfully regularized on the basis of speaker identity, and in the absence of this, lexically conditioned on the noun. Brown, Smith, Samara, and Wonnacott (2021) used a similar experimental paradigm, featuring two different languages with either fully consistent or partially consistent semantic cues. In both cases, adult learners regularized, generalizing the words to novel contexts though this behavior was much weaker for the partially predictive cues.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

In fact, even when adults display probability matching behavior, there is a small bias towards regularization (Smith & Wonnacott, 2010). Smith and Wonnacott (2010) exposed learners to a semi-artificial language with unpredictable plural marking. Nearly all participants' probability matched, reproducing this variability. However, in a second experiment, when the output languages created by this first group of learners was passed on to a second group, and so on, a small bias in favor of regularization was revealed. As transmission filtered the language through multiple adult learners, unpredictable variation was removed, resulting in more systematic languages. These languages did not exhibit zero variability keeping all but a single plural marker (cf. Reali & Griffiths, 2009). Instead, they displayed lexical conditioning with each noun associated with a particular plural.

Mirroring existing work on morphological variation (Brown et al., 2021; Samara et al., 2017), Altenhof and Roberts (under revision) examined the origin of near-synonymy by conditioning lexical items on positive or negative meaning. Probing both the cognitive bias theory and distributional effects, they conducted two novel word learning experiments where participants were taught two potentially synonymous unfamiliar verbs and asked to generalize them to new positive or negative sentences. Intended to manipulate the amount and consistency of information about possible positive or negative meaning and designed to mimic the variability common to natural languages (Mirkovic, MacDonald, & Seidenberg, 2005), participants were exposed to these words in a number of conditions.

In the Random condition, participants were given conflicting data about meaning with each verb appearing equally in both positive and negative contexts. In the Neutral condition, participants were given no data about meaning — the verbs only appeared in neutral sentences. Lastly, in the Consistent condition participants were given clear data about meaning — one word appeared solely in negative contexts, while the other appeared solely in positive ones.

Across conditions, most participants differentiated the verbs, and a large portion did so despite receiving random or neutral information during learning, suggesting they were quite resistant to treating the words as exactly synonymous. While many of these learners matched the probabilities present in their input, a good proportion did not, adding systematicity to their input and conditioning the words on context. As a result, participants' differentiation was not linearly tied to their input. The authors considered this outcome to be compatible with mutual exclusivity, where learners are biased against absolute synonymy.

However, they did not rule out a role for distributional effects. Though differentiation still occurred in the under-informative conditions, learners only fully differentiated the verbs, restricting each verb to a single context, positive or negative, in the Consistent condition. Here, participants' input was clear — there was no contextual overlap for the verbs. As a result, input quality may matter more than quantity for the strength of differentiation.

An iterated, semi-artificial word learning experiment paired with accessible contexts for lexical conditioning, like Smith and Wonnacott (2010), could elucidate the bias against synonymy suggested by Altenhof and Roberts (under revision). Weak differentiation at the participant-level could lead to stronger differentiation at the population-level over time, as the proposed bias is magnified during cultural transmission. Couching this in terms of Altenhof and Roberts's (under revision) findings, we might see greater rates of full differentiation, regardless of input quality, after a few generations of learners.

The present experiment aims to assess this proposal by applying an iterated learning paradigm to Altenhof and Roberts (under revision) language learning task. As in Altenhof and Roberts (under revision), participants were exposed to two new, potentially synonymous verbs, *snater* and *fincur*, embedded in English sentences connoting positive and negative meaning. Afterwards, learners were tasked with extending the verbs to new positive and negative sentences by filling in the blank. Participants' output during this task was then passed on to the next learner, and so on. Critically, for the first participant in each chain, the verbs appeared equally in both sentence types, providing mixed information about positive or negative meaning.

Method

75 participants were recruited from Prolific. All participants were native English speakers and were compensated £2.50 for their time.

Stimuli

Stimuli consisted of 100 sentences from Altenhof and Roberts (under revision). Of these sentences, 36 featured a novel noun, *murp*. The remaining 64 were written for one of two verbs, *snater* or *fincur*. Only sentences taking a verb were further divided into Positive (e.g., "I'm proud to be the best at ______!") and Negative (e.g., "I hope he breaks his habit of ______ all the time") contexts.

Positive and negative valence were established in an independent norming study in Altenhof and Roberts (under revision). Participants were shown each potential stimuli sentence (e.g., "I would love to see you neert sometime!") with a random novel verb (other than *snater* or *fincur*). After the presentation of each sentence, they were asked to rate the meaning of the emphasized verb on a scale from 0, the most negative interpretation, to 50, neutral, and finally 100, the most positive. Re-analyzing the sentences from Altenhof and Roberts (under revision), the mean positivity score for each context type was as follows: Positive, 68.514; Negative, 29.188. Likewise, a one-way ANOVA revealed that the difference in mean positivity score for each context was indeed significant, $F(1,58) = 1593, p < .001$.

Procedure

Language exposure and generalization Participants were taught three "new English slang words" — a noun, *murp*, and two target verbs, *snater* and *fincur* — in a mini-artificial language experiment conducted via JsPsych (Leeuw, 2015). Before beginning the experiment, they were told that "two of these words, **snater** and **fincur**, have the same definition. That is, *snater* and *fincur* are both verbs and they both refer to the same action." Participants were further instructed that they would see all three novel words across a variety of sentences (*Exposure phase*) before using the words to fill in the blank for new sentences (*Generalization phase*). The definition of *snater* and *fincur* was purposefully left undefined. This decision served two purposes. First, guessing their shared meaning served as a cover task for our participants — participants were informed in the initial instructions that they would have to infer the verbs' definition at the end of the experiment. Second, specifying the meaning of the verbs would have introduced a potential for participants' opinions about the activity denoted by them (or the English verb presented as equivalent to them) to influence their responses.

After reading these instructions, but before moving on to the Exposure phase, participants completed an understanding check. This check included a series of questions intended to ensure participants recognized the words they would be learning, understood their respective parts of speech and acknowledged that *snater* and *fincur* were synonymous. Participants received corrective feedback or reinforcement on each question before proceeding to the next.

During the Exposure phase, participants saw each word (*murp*, *snater* and *fincur*) 12 times, for a total of 36 sentences. Sentences were presented individually, remaining on screen for ten seconds before a button appeared that allowed participants to move ahead.

Next, during the Generalization phase, participants were shown 36 unseen sentences with a blank. Participants were tasked to fill the blank for each sentence via a forced response, multiple choice question. Of these 36 sentences, one-third featured positive frames (e.g., "I'm so excited to ______ later."), one-third featured negative frames (e.g., "Don't ______ near me.") and one-third took a noun (e.g., "I lent her my ______.") The sentences used for Generalization were selected randomly from a larger pool of 64. Exposure sentences for the prior participant in a chain were also recycled back into the sentence pool to be used for Generalization. This combination of randomization and recycling ensured that different participants viewed unique sets of sentences during both Exposure and Generalization. The inclusion of sentences requiring a noun and the addition of murp as a choice for sentences requiring a verb doubly functioned as an attention check and helped to mask the true focus of the experiment. Consistently placing verbs in contexts written for nouns (or vice versa) could suggest inattentiveness or failure to fully comprehend the instructions. Moreover, if participants thought the task was actually about parts of speech, it may reduce demand characteristics.

Initial input language All participants in generation one, the first of their respective chains, received exactly the same input language — 36 sentences, 12 for the noun and 24 for the verbs. The 24 sentences for verbs were divided equally between *snater* and *fincur*. Verb sentences were further divided into 12 positive and 12 negative frames. Critically, *snater* and *fincur* occurred equally in both contexts following the distribution used by Altenhof and Roberts (under revision) Random Condition. This particular distribution was specifically chosen to ensure there was no biasing information in the initial distribution to suggest a particular meaning for each verb, positive or negative.

Diffusion Chain Design All 75 participants were organized into 15 diffusion chains of five people. The initial participant in each chain was trained on the input as specified above. Each subsequent learner directly received the prior participant's output during the Generalization phase (the sentence plus the filled-in blank) as their input during the Exposure phase.

To prevent confusion about part of speech and center variability related to positive or negative meaning, part of speech errors were not passed on to the next participant in the chain. When sentences intended for verbs were instead filled with a noun, they were omitted for the next generation, reducing the number of sentences in the Exposure phase for the following learner. However, participants who made more than three of these errors during the Generalization phase were excluded and subsequently replaced. Likewise, regardless of a participants' choice during Generalization, sentences written for nouns were transmitted with the appropriate noun for the next participant's Exposure phase.

Scoring

For each participant, both target verbs were given a context score. The context score for each word (*fincur context score* and *snater context score*, respectively) was calculated as the proportional frequency of that word in its majority context divided by the sum of its frequencies in both contexts. Whatever context, positive or negative, the word was used in a simple majority of times was selected as the majority context. If a word occurred equally in both contexts, it received a context score of 0.5

A *Differentiation score*, the product of the two context scores, was then calculated for each participant to measure the extent to which they differentiated the verbs based on context. Greater differentiation is indicated by higher Differentiation scores. A score of 0 reveals that a single verb was used across all contexts (while the other was not used at all). A score of 0.25 serves as the chance level — both verbs occurred equally as often in both positive and negative contexts, undifferentiated. In contrast, a score near 0.67 suggests *partial differentiation* — one of the words was used exclusively in a single context (positive or negative) while the other appeared in both . Finally, a score of 1.0 represents *full differentiation*, where each verb is used only in one context, with no overlap. Scoring examples for all the above ideal cases can be seen in Figure 1.

A - No differentiation (Chance level)

Fincur is used 50% of the time in positive contexts and 50% of the time in negative contexts.

$$Fincur\ context\ score = \frac{Frequency\ of\ fincur\ in\ majority\ context}{Frequency\ of\ fincur\ in\ all\ contexts}$$

$$Fincur\ context\ score = \frac{0.5}{0.5+0.5} = \frac{0.5}{1} = 0.5$$

Snater is used 50% of the time in positive contexts and 50% of the time in negative contexts.

$$Snater\ context\ score = \frac{Frequency\ of\ snater\ in\ majority\ context}{Frequency\ of\ snater\ in\ all\ contexts}$$

$$Snater\ context\ score = \frac{0.5}{0.5+0.5} = \frac{0.5}{1} = 0.5$$

$$Differentiation\ score = Fincur\ context\ score * Snater\ context\ score = 0.25$$

B - Partial differentiation

Fincur is used 50% of the time in positive contexts and 0% of the time in negative contexts.

$$Fincur\ context\ score = \frac{Frequency\ of\ fincur\ in\ majority\ context}{Frequency\ of\ fincur\ in\ all\ contexts}$$

$$Fincur\ context\ score = \frac{0.5}{0.5+0} = \frac{0.5}{0.5} = 1$$

Snater is used 50% of the time in positive contexts and 100% of the time in negative contexts.

$$Snater\ context\ score = \frac{Frequency\ of\ snater\ in\ majority\ context}{Frequency\ of\ snater\ in\ all\ contexts}$$

$$Snater\ context\ score = \frac{1}{1.0+0.5} = \frac{1}{1.5} = 0.667$$

$$Differentiation\ score = Fincur\ context\ score * Snater\ context\ score = 0.667$$

C - Full differentiation

Fincur is used 100% of the time in positive contexts and 0% of the time in negative contexts.

$$Fincur\ context\ score = \frac{Frequency\ of\ fincur\ in\ majority\ context}{Frequency\ of\ fincur\ in\ all\ contexts}$$

$$Fincur\ context\ score = \frac{1}{1.0+0} = \frac{1}{1} = 1.0$$

Snater is used 0% of the time in positive contexts and 100% of the time in negative contexts.

$$Snater\ context\ score = \frac{Frequency\ of\ snater\ in\ majority\ context}{Frequency\ of\ snater\ in\ all\ contexts}$$

$$Snater\ context\ score = \frac{1}{1.0+0} = \frac{1}{1} = 1$$

$$Differentiation\ score = Fincur\ context\ score * Snater\ context\ score = 1$$

Figure 1: Ideal cases for differentiation

Results

Data are available here. Analyses were performed using R (R Core Team, 2013) and graphs were created using ggplot2 (Wickham, 2011). Figure 2 displays the Differentiation scores for each participant in all fifteen diffusion chains. Three participants were removed for applying the noun to more than three verb sentence frames and were subsequently replaced. Eleven participants who made one error and two participants who made three errors were not removed. A single participant was removed for the complete exclusion of one of the verbs. A one-way ANOVA revealed there was no significant effect of generation on differentiation score, $F(4,70) = 0.977, p = 0.426$. That is, Differentiation scores did not change significantly over the course of each diffusion chain. Likewise, a repeated measures ANOVA did not reveal a significant effect for chain position on participant Differentiation scores, $F(2.89,40.48) = 1.51, p = 0.228$ (Greenhouse-Geisser correction). This suggests that most participants were matching their input by copying the distribution of *snater* and *fincur* from the prior learner. Participants were also not applying *fincur* and *snater* randomly. The mean Differentiation score across participants was 0.56 ($SD = 0.27$). One-sample Wilcoxon Signed-Rank tests for each generation revealed the mean for each generation was significantly higher than chance, represented by a Differentiation score of 0.25, (all $p < 0.01$).

We completed further analysis of participants' behavior with respect to their input. Following Altenhof and Roberts (under revision), which aimed to mirror prior work studying the regularization of unpredictable variation by Hudson Kam and Newport (2009), participants were grouped into the following three categories: matchers, under-matchers and over-matchers. Matchers replicated their Exposure phase distribution in their Generalization phase output. Over-matchers extended their input frequencies introducing new systematicity to their output. Lastly, under-matchers fell short of their input choosing words less systematically than their predecessors. Participants' categories were decided using a randomization test with 10,000 replications which compared a participant's actual Differentiation score to an expected one based on probability matching given the distribution found in a participant's input. 52 participants (69%) were classified as matchers, $14(19\%)$ were classified as over-matchers and the final nine participants (12%) were classified as undermatchers. That is, while the majority of participants replicated their input probabilities, more than a quarter of learners changed their input in some way either adding or reducing regularity during Generalization.

To get a better understanding of the degree to which participants differentiated the verbs — fully, partially, or not at all — we performed a k-means clustering analysis on all participants. Given a set of observations and a predetermined parameter k, the algorithm attempts to find the best way to break up the observations into k clusters (Steinley, 2006). For this analysis, the gap statistic identified three as our ideal value for k. One cluster (non-differentiators) had a center at a Differentiation score of 0.33, a second was centered around 0.59 (partial differentiators) and the final cluster was centered at a Differentiation score of 0.95 (full differentiators). There were 20 full differentiators, 18 partial differentiators and 37 non-differentiators. Figure 3 shows each participant's Differentiation score with respect to their input Differentiation score and their appropriate cluster. Participants closest to the center line, x = y, are largely matching their Exposure distribution, while those far above or below it are adding or removing systematicity, respectively.

Discussion

The present experiment explored the differentiation of expressive near-synonymous pairs where at least one member carries a positive or negative connotation (e.g., *steadfast* : *stubborn*). To do so, participants organized into simple diffusion chains were taught three novel words—*murp*, a distractor noun and *fincur* and *snater*, two potentially synonymous verbs. They were then tasked to extend the verbs to new positive and negative sentences. Despite varying levels

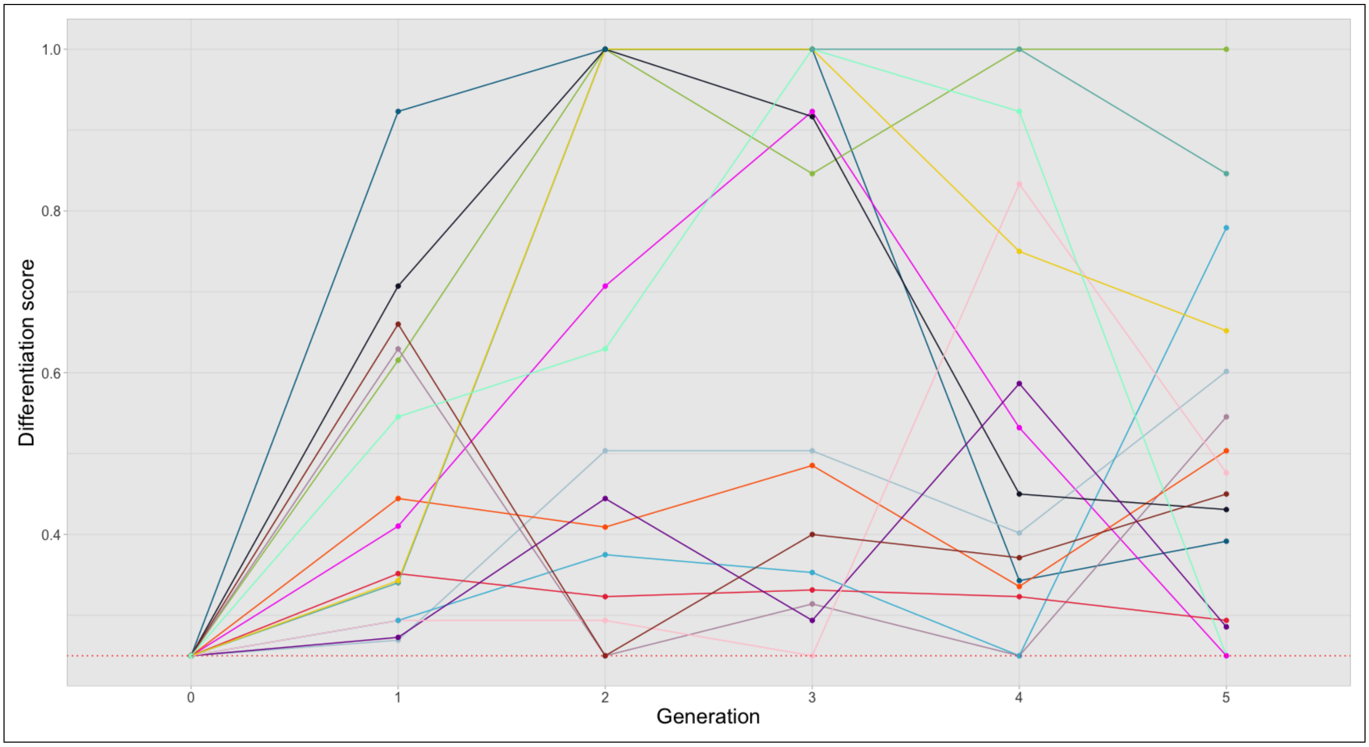

Figure 2: Participant Differentiation scores. Each line represents an individual diffusion chain. The dotted line at 0.25 indicates a differentiation score at chance.

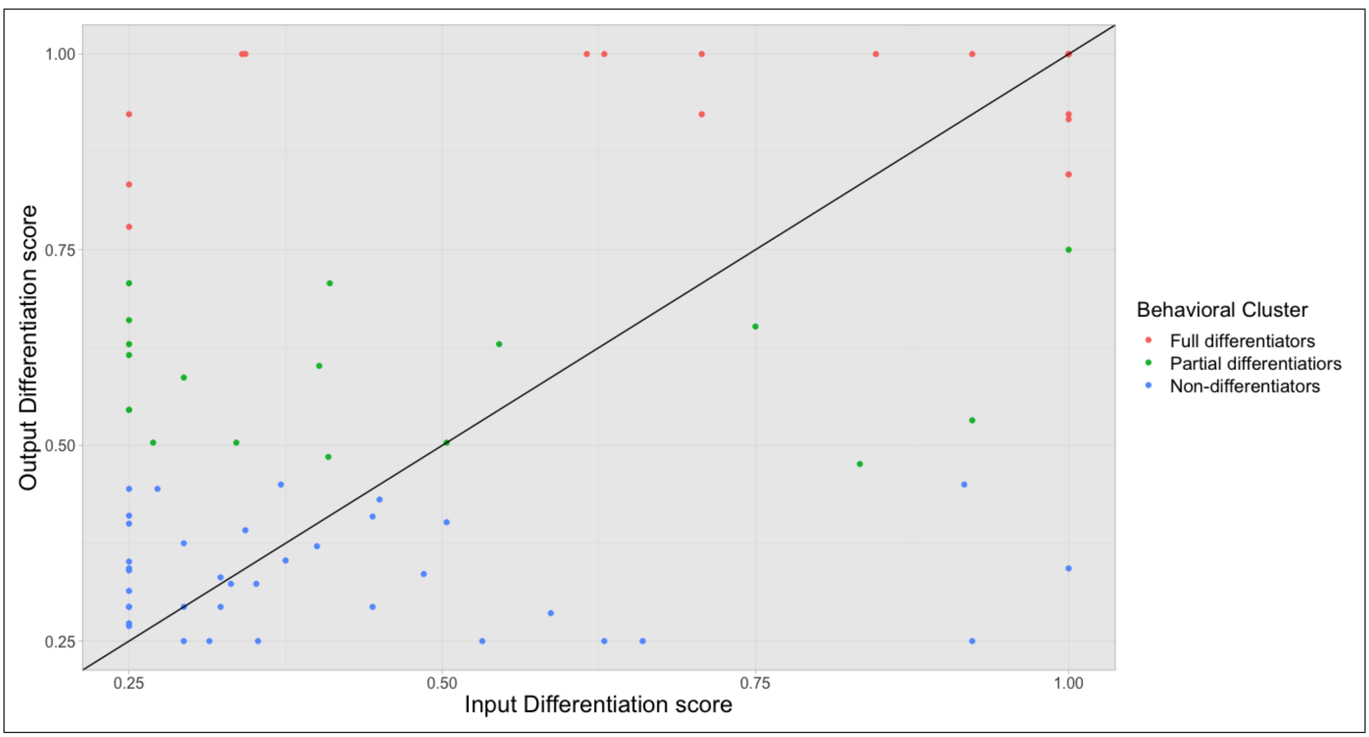

Figure 3: Participant Differentiation scores and behavior (full, partial, non- differentiators) with respect to their input. Participants closer to the line are matching their input, while those above and below it are introducing and reducing systematicity, respectively

of consistency in their exposure input, slightly more than half of participants strongly differentiated the verbs. These participants were either full differentiators — applying one word to one context, positive or negative, with no overlap (e.g., *thrifty* : *stingy*) — or partial differentiators, where one word functioned as a default in both contexts and the other, specialized word dominated remaining context (e.g., *vintage* : *old*).

Interestingly, cultural transmission did not lead to a general increase in regularity. Only a handful of chains converged on something like full differentiation. Participants in later generations did not display stronger differentiation and instead exhibited the heterogeneity characteristic of single learners. In fact, Altenhof and Roberts's (under revision) non-iterated paradigm, featuring similarly varying levels of input, reported a nearly identical pattern of differentiation; nearly half of participants displayed full or partial differentiation. At first, this finding appears potentially at odds with prior literature (Smith & Wonnacott, 2010; Smith et al., 2017) where cultural transmission reduces unpredictable variation. In these experiments, though individual participants closely matched their input frequencies when choosing potential variants, iterated learning served to amplify small biases toward regularization, culminating in more systematic languages.

However, simplifications at the individual level, as found here, do not always result in language simplification at the population level. Atkinson, Smith, and Kirby (2018) taught a second generation of learners a morphologically complex language based on the output of the first, manipulating both the complexity of the input and the number of participants in the previous generation (from which the input was drawn). Regardless of where input complexity came from (consistently reproduced versions of the complex target language from prior learners or mixing multiple simplified systems from prior learners), second generation learners produced complex languages in their output. The authors concluded that simplification (regularization) from initial adult learners is often idiosyncratic in its complexity and cannot solely provide an account for systematicity at the population level (Atkinson et al., 2018). Consequentially, it is helpful to draw a distinction between differentiation as a behavior shown by single learners and the collection of participants as a whole. While a majority of participants did in fact probability match, a third significantly altered their input. Over-matchers increased the contextual consistency of the verbs, conditioning on positive or negative meaning. This behavior is somewhat surprising given research reaffirming adults' tendency to probability match (Hudson Kam & Newport, 2005; Smith & Wonnacott, 2010). Of course, that is not to say adults never over-match — when presented with sufficiently complex input, adults display regularization. Adults will regularize when unpredictable alternative forms are numerous and infrequent or novel (Hudson Kam & Newport, 2009; Wonnacott & Newport, 2005). The participants who regularized in this study were likely not responding to distributional complexity or increasing demands on memory. Learners may have been responding to a complexity of a different kind — semantic complexity. Previous work on lexical conditioning suggests that all types of conditioning are not created equal (Samara et al., 2017). Although our stimuli sentences were written and normed to introduce a relatively straightforward distinction for conditioning by linguistic context — positive or negative meaning — participants may have identified or introduced subtle connotative meanings. This semantic ambiguity is both a bug and a feature — more faithful to natural language but a challenge to operationalize.

There were also a surprising number of under-matchers, who reduced the systematicity present in their Exposure phase input. Their behavior may be partially understood as resulting from differential attention (or lack thereof). Participants varied in their completion times which ranged from five minutes to greater than twenty. Perhaps as participants pay less attention to the task, their responses become more noisy, which can then be capitalized on by the next (possibly) inattentive participant in the chain, rapidly and dramatically impacting Differentiation scores, as seen in some of the most unstable chains. This claim seems supported by Brown et al. (2021) where participants were only able to regularize from unpredictable semantic cues if they were also able to verbalize the relationship between the cue and their choices.

Though some degree of instability due to participant heterogeneity characterizes iterated learning (Navarro, Perfors, Kary, Brown, & Donkin, 2018), future work could aim to decrease distortions caused by extreme learners. Participants were forced to wait before clicking on to the next sentence during the Exposure phase but not during Generalization. A delay before answering and simple attention checks could help slow down, reorient and refocus distracted participants. Additionally, to account for individual variability, more participants could be included in each generation and the mean Differentiation score for that group could be transmitted to the next (cf. Atkinson et al., 2018).

Manipulating the communicative demands of the experiment could also further increase systematicity. In an artificial language experiment, Fehér, Wonnacott, and Smith (2016) found that interaction reduced unpredictable variation in word order; participants understood the "counter-functional" nature of this variation when trying to communicate successfully. Fehér, Ritt, and Smith (2019) report similar results. When "variable" singular marking partners interacted with their "categorical" marking counterparts, they accommodated by increasing the frequency of their marker usage. But "Categorical" markers didn't do the same by becoming more variable. Fehér et al. (2019) concluded that this asymmetry could push a population toward categorical use of a variable — relative to this study, the complete positive or negative application of the verbs (i.e., full differentiation).

Alternatively, the most stable outcome for synonymy, that we might expect to find at the end of a diffusion chain, may not be fully systematic, or in this context, fully differentiated. Though the bar for full and partial differentiation is high, most participants displayed some kind of differentiation, resulting in a mean differentiation score well above chance. Learners were hesitant to treat *fincur* and *snater* as exact synonyms, aligned with a potential bias against absolute synonymy, but were simultaneously reluctant to completely condition on context. It could be that this bias encourages words to be distinctive, following learners' expectations of contrast (E. Clark, 1992) but minimally so, explaining the "fuzzy", connotative, and frequently individual, differences that characterize many synonymous terms. This semantic uncertainty has arisen in other artificial language experiments. For example, in Fedzechkina, Hall Hartley, and Roberts (2022), participants were exposed to a miniature language with uninformative word order and two of its dialects — one with case and one without. Participants socially biased toward the no case dialect ultimately dropped case marking, even though doing so increased ambiguity and resulted in a less communicatively efficient language.

Ultimately, the paradigm used in the current work represents a vast simplification of natural language use. While the choice to use three words — only two of which were synonymous — was simplistic enough to avoid overwhelming participants, real synonyms only sometimes occur in pairs. Often they exist in far larger groups, spanning multiple intertwined semantic dimensions (e.g., *lie* : *fib* : *falsehood* : *untruth* : *misrepresentation* : *alternative fact*). Future studies could explore more complex manipulations to semantic space (such as force, formality, or intensity, etc.) or increase the number of synonymous competitors. Our participants additionally had to juggle multiple hypotheses about meaning — the positive/negative distinction as well as a working definition for *snater* and *fincur*. Though outside the scope of the current study, choosing a set meaning could change the nature of differentiation.

Similarly, both *snater* and *fincur* were verbs, though synonyms that convey different expressed attitudes can transcend lexical categories (e.g., "she's the boss", "she's bossy", "girlboss"). Moreover, syntactic framing can influence the interpretation of, and serve as the basis for, synonymous differentiation, as is the case for collocational near-synonyms (Edmonds, 1999). These near-synonyms differ in how they interact with their surrounding context, due to selectional restrictions based on denotation (e.g., both planes and birds can land, but only the latter can perch), expectations about recurrent word combinations (e.g., revisions are a daunting task, not a daunting job) and the way grammatical roles are assigned (e.g., Sebastian teaches linguistics to students, *Sebastian instructs linguistics to students). Given that participants viewed a variety of sentences with different lengths, frames and elements (propositions, exclamations, etc.), our learners may have conditioned on syntactic context. For example, "I hope he breaks his habit of ______ all the time" contains a gerund, with the verb operating as a noun, whereas "Gross! She's ______!" takes a progressive verb. Casenhiser (2005) found that syntactic category impacts children's understanding of homonyms; children are better able to accept a new meaning for a homonym when the syntactic context indicates a new meaning is required. Nonetheless, the present study was intended to investigate expressive near-synonyms and any effects arising from syntactic context are unintentional (see Hudson Kam, 2015, where children conditioned variation on a variable unexpected by the researchers). Further edits and norming to the stimuli sentences with these considerations in mind could prove a fruitful avenue for further research.

Finally, our model of cultural transmission was also limited. To begin, it is important to note that the connection between iterated learning and learner biases is not always straightforward. Weak biases can play a large role in shaping language structures but the converse is also true: strong biases can have a weak or minimal impact (Smith et al., 2017). For our purposes, within each diffusion chain, each generation contained a single learner transmitting information in a single direction. In natural language environments, learners receive input from and interact with multiple diverse speakers. Simulations from Smith (2009) showed that when learners receive information from all speakers in the previous generation, iterated learning with Bayesian agents does not necessarily converge on the learners' prior. Rather, the population came to use a single, dominant language based on the original distribution of languages, potentially in conflict with the prior. However, when learners exposed to multiple teachers account for the possibility of divergent hypotheses from each one, iterated learning reflects learners' initial biases, as it would with a single teacher (Burkett & Griffiths, 2010). Population size could be particularly interesting to manipulate in the present study; synonymy requires that participants explicitly deal with multiple possible candidate meanings. Transmission also only proceeded in one direction. Naturalistic interaction is far more nuanced and reciprocal, and this kind of interaction may be required for the emergence of some forms of communication, such as graphic symbol systems (Garrod, Fay, Rogers, Walker, & Swoboda, 2010).

Despite these limitations, the present study is an informative step in the direction of understanding the dynamics of synonymy. In spite of conflicting information about positive or negative meaning, participants differentiated two potentially synonymous verbs in support of a possible cognitive bias against synonymy. With many chains converging on middling degrees of differentiation, this bias may encourage minimally distinctive differentiation between competing synonyms suggesting that partial differentiation might be the stablest outcome for synonymy.

References

Altenhof, A., & Roberts, G. (under revision). Quality, not quantity, impacts the differentiation of near-synonyms.

Atkinson, M., Smith, K., & Kirby, S. (2018). Adult learning and language simplification. *Cognitive Science*, *42*(8), 2818–2854. doi: 10.1111/cogs.12686

Brown, H., Smith, K., Samara, A., & Wonnacott, E. (2021). Semantic cues in language learning: An artificial language study with adult and child learners. *Language, Cognition and Neuroscience*, *37*(4), 509–531. doi: 10.1080/23273798.2021.1995612

Burkett, D., & Griffiths, T. (2010). Iterated learning of multiple languages from multiple teachers. In A. Smith, M. Schouwstra, B. Boer, & K. Smith (Eds.), *The evolution of language: Proceedings of the 8th international conference* (pp. 58–65). Singapore: World Scientific.

Casenhiser, D. M. (2005). Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language*, *32*(2), 319–343. doi: 10.1017/s0305000904006749

Clark, Parikh, P., & Ryant, N. (2007). *Evolving linguistic defaults.* University of Pennsylvania.

Clark, E. (1992). Conventionality and contrast: Pragmatic principles with lexical consequences. In A. Lehrer & E. Kittay (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 171–188). Lawrence Erlbaum Associates, Inc.

Cruse, D. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.

DiMarco, C., Hirst, G., & Stede, M. (1993). The semantic and stylistic differentiation of synonyms and near-synonyms. In *AAAI Spring Symposium on Building Lexicons for Machine Translation* (pp. 114–121).

Edmonds, P. (1999). *Semantic Representations of Near-Synonyms for Automatic Lexical Choice* (Unpublished doctoral dissertation). University of Toronto.

Fedzechkina, M., Hall Hartley, L., & Roberts, G. (2022). Social biases can lead to less communicatively efficient languages. *Language Acquisition*, 1–26. doi: 10.1080/10489223.2022.2057229

Fehér, O., Ritt, N., & Smith, K. (2019). Asymmetric accommodation during interaction leads to the regularisation of linguistic variants. *Journal of Memory and Language*, *109*, 104036. doi: 10.1016/j.jml.2019.104036

Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*, *91*, 158–180. doi: 10.1016/j.jml.2016.06.002

Garrod, S., Fay, N., Rogers, S., Walker, B., & Swoboda, N. (2010). Can iterated learning explain the emergence of graphical symbols? *Experimental Semiotics*, *11*(1), 33–50. doi: 10.1075/is.11.1.04gar

Hudson Kam, C. (2015). The impact of conditioning variables on the acquisition of variation in adult and child learners. *Language*, *91*(4), 906–937. doi: 10.1353/lan.2015.0051

Hudson Kam, C., & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*(2), 151–195. doi: 10.1207/s15473341lld0102_3

Hudson Kam, C., & Newport, E. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66. doi: 10.1016/j.cogpsych.2009.01.001

Leeuw, J. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. doi: 10.3758/s13428-014-0458-y.

Lewis, M., Cristiano, V., Lake, B., Kwan, T., & Frank, M. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, *198*, 104191. doi: 10.1016/j.cognition.2020.104191

Markman, E., & Wachtel, G. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.

Mirkovic, J., MacDonald, M., & Seidenberg, M. (2005). Where does gender come from? Evidence from a complex inflectional system. *Language and Cognitive Processes*, *20*(1-2), 139–167. doi: 10.1080/01690960444000205

Navarro, D., Perfors, A., Kary, A., Brown, S., & Donkin, C. (2018). When extremists win: Cultural Transmission via iterated learning when populations are heterogeneous. *Cognitive Science*, *42*(7), 2108–2149. doi: 10.31234/osf.io/7npq6

Reali, F., & Griffiths, T. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*(3), 317–328. doi: 10.1016/j.cognition.2009.02.012

Samara, A., Smith, K., Brown, H., & Wonnacott, E. (2017). Acquiring variation in an artificial language: Children and adults are sensitive to socially-conditioned linguistic variation. *Cognitive Psychology*, *94*, 85–114. doi: 10.1016/j.cogpsych.2017.02.004

Savage, S., & Au, T.-F. (1996). What Word Learners Do When Input Contradicts the Mutual Exclusivity Assumption. *Child Development*, *67*(6), 3120. doi: 10.2307/1131770

Smith, K. (2009). Iterated learning in populations of Bayesian agents. In N. Taatgen & H. Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 697–702). Austin, TX: Cognitive Science Society.

Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711). doi: 10.1098/rstb.2016.0051

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449. doi: 10.1016/j.cognition.2010.06.004

Steinley, D. (2006). K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 1–34.

Team, R. C. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for

Statistical Computing.
Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729–742. doi: 10.1016/j.cognition.2007.08.00
Wickham, H. (2011). ggplot2. *Wiley interdisciplinary reviews: computational statistics*, *3*(2), 180–185.
Wonnacott, E., & Newport, E. (2005). Novelty and Regularization: The Effect of Novel Instances on Rule Formation. In A. Brugos, M. Clark-Cotton, & S. Ha (Eds.), *Proceedings of the 29th annual Boston University Conference on Language Development.* Somerville, Massachusetts: Cascadilla Press.