Title: “Concrete similarity among exemplars facilitates the initial implicit learning of a constrained category”

Matt A. Johnson
Department of Psychology, Green Hall
Princeton University
Princeton, NJ 08540
Office: (609) 258-0870
Cell: (707) 477-5185
Email: majthree@princeton.edu

Adele E. Goldberg
Department of Psychology, Green Hall
Princeton University
Princeton, NJ 08540
Office: (609) 258-8772
Email: adele@princeton.edu

Header: “Input advantages in category learning”

Corresponding author: Matt A Johnson
Department of Psychology, Green Hall
Princeton University
Princeton, NJ 08540
Abstract

Within both linguistic and non-linguistic category learning, there have been mixed findings as to what type of input best facilitates the generalization process. While certain work emphasizes the importance of similarity among exemplars during learning, other work demonstrates that more variability is best. We attempt to reconcile these apparent conflicts by recognizing two distinct components of the categorization process: the recognition of an abstract, constrained category and the broadening of a category. We find that while concrete similarity is helpful for the former, variability in the input facilitates the latter. We also provide experimental results demonstrating that category learning is sensitive to whether the learning is explicit or implicit, which reconciles additional conflicts arising research on second language learners.
Introduction

For the latter half of the 20th Century, the biggest debate in language acquisition was whether essential aspects of language were learned in the way that other sorts of implicit knowledge is learned, or whether instead, language is “acquired” by the triggering of preexisting, i.e., innately specified, knowledge. Although the latter view still has its defenders (Fodor 1998; Yang 2002; Snyder 2001; Culbertson, Smolensky, Legendre to appear), a growing number of researchers now adopt the position that most if not all linguistic knowledge is in fact learned via domain-general learning abilities (e.g., Saffran, Aslin and Newport 1996; Langacker 1987; Tomasello 2003; Hauser, Chomsky and Fitch 2002; Kam and Newport 2005; Kuhl 2000; Goldberg 2006).

The learning perspective immediately raises the issue of exactly how language is learned and whether there are ways to structure the input that can facilitate language learning. If learning crucially involves, at least in part, generalizing or categorizing over witnessed exemplars, we can turn to the field of non-linguistic categorization for insights into the process. In fact, we know a great deal about categorization (see e.g., Murphy 2002 for a review). In the present work, we focus on the following two well-established principles:

1) Concrete similarity among exemplars facilitates generalization or the recognition of an abstract category (Markman and Gentner 1993; Elio and Anderson 1984; Sloutsky 2003)
2) Increased variation among types facilitates generalization (Bybee 1995; Marchman and Bates 1994; Abbot-Smith and Tomasello 2006)

While both principles have empirical support, they may appear to be at odds with one another insofar as the first would seem to imply that shared concrete similarity (less variation) encourages generalization, while the second implies that higher type frequency (more variation) encourages generalization. In fact, recent language acquisition studies have found relevant results that appear to contradict one another. In particular, certain previous findings have found that, when type and token frequency are held constant, exposing learners to skewed input, such that tokens share a good deal of concrete similarity, provides an advantage in category formation over exposing learners to input in which tokens are more varied (Casenhiser and Goldberg 2005; Goldberg et al. 2004; Goldberg et al. 2007). On the other hand, other work has found an advantage of exposing learners to more balanced input over skewed (Siebenborn, Krajewski and Lieven forthcoming), or no advantage for either (Year and Gordon 2010).

The present paper aims to reconcile the two principles and the existing research findings by recognizing two components of the generalization process. In addition, two means of forming generalizations can also be contrasted: implicit generalization over witnessed exemplars and explicit hypothesis testing. We provide experimental evidence that the two means of forming generalizations can lead to different effects, depending on the type of input that is witnessed. First, we review evidence for the categorization principles in (1) and (2).
Concrete similarity among exemplars facilitates category learning

Several researchers have noted that exposure to low-variance input or to prototypical instances before more varied input facilitates the learning of non-linguistic categories (Elio & Anderson, 1984; Homa et al., 1991; Nosofsky, 1988; Mervis and Pani, 1980; Hupp and Mervis, 1981; Mirman, 1978). Since prototypical exemplars are defined to be those exemplars that share more attributes with other category members and fewer attributes with non-category members, exposure to them tends to increase concrete similarity.

Elio and Anderson (1984) have provided relevant support. In a non-linguistic category learning experiment, they exposed participants to one of two conditions: in the skewed (“centered”) condition, subjects were initially exposed to more frequently represented, more prototypical instances, with the study sample growing gradually to include the full range of members in the category; in the balanced (or “representative”) condition, subjects were trained on a fully representative sampling from the start. Without explicit instruction, learners were more accurate in the skewed condition, yielding better typicality ratings and accuracy during the test phase on new instances.

Markman & Gentner (1993) note that processes of analogy required for generalization (their “structural alignment”) is facilitated when instances being compared are similar to one another. Gentner, Loewenstein, & Hung (2007) illustrated this idea by showing children a particular picture of a Martian to be used as a standard for comparison as well as two alternative Martian creatures. The standard Martian and one of the alternatives shared one body part, while the distinct Martian didn't. Children were asked, “This one has a blick; which one of these has a blick?” The results demonstrated that if the two alternatives were highly similar to the standard, children were better able to pick out the relevant shared body part; when they were only weakly similar, finding the body part was more difficult. In addition, Gentner et al. demonstrated that children who were tested in the high-similarity condition first, were subsequently more successful on the low-similarity items than children who had the same amount of experience with only low-similarity items. Moreover, categories that are identifiable with a salient type of stable feature are easier to learn than categories in which the feature is instantiated in different ways, even when the variability is relevant to the feature dimension (Markman & Maddox, 2003).

Observations in language learning also support a facilitating effect for concrete similarity. It is often noted that learners first recognize generalizations related to particular verbs before they come to generalize across verbs to learn completely abstract constructions such as the transitive. This suggests that learners recognize generalizations across instances that share concrete similarity more easily than they recognize more purely abstract generalizations (see Tomasello 2000 for review; also Brandt et al. 2011). The concrete similarity may not necessarily require a shared verb, since there is also some evidence that the transitive construction involving fixed pronouns, “He X’ed it” is well-learned before the more abstract construction (Childers & Tomasello, 2001). Rost and McGregor (2011) have recently shown that exposure to exemplars that emphasize a prototypical exemplar leads to better word learning than exposure to more varied exemplars. Sloutsky and Fisher (2004) offer a general model of category learning that is based on shared similarity among exemplars; they suggest that the well-known advantage that linguistic labels provide in category learning (e.g.,
Gelman and Markman 1986) exists because the label itself provides a feature that is shared among all category members (cf. also Lupyan 2008).

In these ways, it is clear that shared similarity among exemplars facilitates generalization in that it encourages the initial recognition of a category. At the same time, increased variability among exemplars facilitates generalization in that it encourages the creation of a broader category.

**Increased variation among types facilitates generalization**

The importance of highly variable types in categorization has been much discussed within the field of language acquisition. Bybee (1995) emphasized that productive morphological categories depend on the morpheme being witnessed with a great variety of types. Connectionist models of morphological productivity rely on this principle to account for regular “rules” (e.g., Plunkett and Marchman 1993). There have also been demonstrations within phonology for the advantage of witnessing multiple types (Richardsmeier, Gerken, and Ohala 2010).

A key concept that is often cited in this context is entropy - the degree of variation or uncertainty in a given distribution (Shannon and Weaver 1949). Greater variation—higher entropy—leads to greater uncertainty in what to expect. In artificial pattern learning tasks, high entropy can lead to the creation of an open slot that can be filled by a variety of items. Gomez (2002), for example, found that 18 month olds could learn non-adjacent dependencies across an open slot when the elements that occurred in the slotted position were sufficient varied. Matthews and Bannard (2010) found similar effects in 2 and 3 year olds that were tested on how likely they were to successfully reproduce novel instances of otherwise familiar four word strings (*A bowl of gooks*). They found that children were more likely to be successful when the string’s final position was difficult to predict from the preceding 3 words (high entropy; e.g., *out of the water/box/air*), than when the context was more semantically constraining (low entropy; e.g. *lets have a look*). This result makes sense if the higher entropy positions were more likely to lead to the creation of an open slot that was more readily filled with novel items.

**Distinguishing two components of generalization: vertical and horizontal**

We have reviewed evidence that shared concrete similarity (less variation) encourages generalization, and that higher type frequency (more variation) encourages generalization. The apparent paradox can be resolved by recognizing two different aspects of generalization: whether an abstract and appropriately constrained category is formed at all, and the extent to which a category is extended broadly. The learning of any category requires generalization, since categories are by their nature abstract, allowing new instances to be recognized as exemplars of the category. Moreover, part of learning a category is learning what the category is not: i.e., learning what distinguishes the category from other possible categories (Rosch and Mervis 1975). We can refer to this initial, constrained abstraction as *vertical* generalization or *generalization V*. All categories require *generalization V*, e.g., learning the categories of “hamster” and “animal” each require *generalization V*, in that they allow for new instances to be treated as members of the respective categories.
On the other hand, *horizontal* generalization (generalization_H) extends an abstraction to include a larger range of instances. For example, “animal” describes a broader generalization_H than “hamster.” The degree of generalization_H therefore captures how variable a category is.

With this distinction in mind, it becomes clear that concrete similarity among items encourages generalization_V in that it allows for an abstract category to be recognized. Variation in types of items encourages generalization_H in that learners come to know an abstraction applies broadly if they see it being applied broadly. In this way, categories that are fairly broad rely more on generalization_H than narrowly defined categories. Clearly there are degrees of generalization_H so it is a relative notion: a North American hamster is less general than “hamster” which is less general than “mammal,” which is less general than “animal,” which is in turn, less general than “thing.”

As noted at the outset, different uses of the word “generalization” have led to some apparently conflicting results in the language acquisition literature. The relevant studies have employed a version of the Elio and Anderson (1984) paradigm described above in contrasting skewed input, in which one type of exemplar accounts for the preponderance of tokens, with representative or balanced input in which exemplars reflect the broader variation of the category. In this work, overall type and token frequencies are controlled for, but exemplars in the skewed conditions are less variable since one prototypical token is repeated more often than the rest. The first principle predicts that less variable (here, skewed) input should facilitate generalization: an abstract but constrained generalization. The second principle predicts that more variable (here, balanced) input should facilitate a broader generalization.

Casenhiser and Goldberg (2005) tested six year old children’s ability to learn to pair a novel constructional meaning with a novel form. The novel form involved familiar nouns and a nonsense verb, arranged in a non-English word order (NP_1 NP_2 V). The meaning of the phrasal pattern was that of APPEARANCE, (a meaning novel for English phrasal patterns): the entity named by the first noun phrase comes to exist in the place named by the second noun phrase. For example, the intended meaning for the sentence *the sailor the pond neebod* was ‘the sailor sailed onto the pond from out of sight.’

In the balanced frequency condition, subjects heard 5 novel verbs, each with a relatively low token frequency: 3 novel verbs occurring four times and two novel verbs occurring twice (4-4-4-2-2). In the skewed condition, subjects again heard the same 5 novel verbs, but this time one had especially high token frequency of 8, while the other four verbs appeared twice each (8-2-2-2-2). The high token frequency exemplar was designed to represent the meaning of the construction in that it was used with scenes of appearance regardless of manner. This was inspired by the fact that argument structure constructions are often most frequently instantiated by a prototypical verb (e.g., give in the ditransitive; put in the caused-motion construction; go in the intransitive motion construction). Note that five different novel verbs and 16 total examples were presented in both experimental conditions. The control condition watched the same film but with the sound turned off.

Results demonstrated that children in the skewed group were most accurate on a forced choice task that required them to identify new exemplars of the novel construction and distinguish them from novel instances of a familiar construction that also involved novel verbs. Participants in the control (no-sound) condition did no better than chance at
choosing the correct scene. The balanced group showed a statistically significant improvement over the control condition, but did not perform as well as the skewed group.

Using this same general paradigm, support for the idea that skewed input facilitates category formation has been found with adults (Goldberg et al. 2004), as well as six year olds, and in the learning of random dot patterns (Casenhiser and Goldberg 2006). Moreover, it was found that witnessing the skewed exemplars before other exemplars also led to greater accuracy (Goldberg, Casenhiser and White 2007).

Note that the construction used in the Casenhiser and Goldberg studies that showed an advantage for skewed input was associated with a fairly specific meaning: an entity appeared on or in a location. There are limited ways in which entities can appear on a scene: they can roll in, slide in, rise up from below, fall down from above, or simply magically appear. Exemplars involved all of these possibilities, but the meaning was always one of something appearing in a location. As noted above, in the skewed condition, the most frequent nonsense verb was used without regard to manner of appearance and thus could be interpreted as denoting the prototypical meaning of the construction. In this way, the generalization that was required was primarily generalizationV. As expected, then, the lower variance skewed input showed an advantage because it helped learners to form a constrained abstract category.

Siebenborn, Krajewski and Lieven (forthcoming) exposed 5 and 7 year olds to a novel word order (SOV) that was assigned the meaning “pretend.” That is, the construction was assigned the meaning that the agent only pretended to perform the action. A BALANCED group witnessed 6 familiar verbs presented either 8 or 11 times each (11-11-8-8-8-8), a SKEWED group witnessed the same verbs with the distribution 23-15-4-4-4-4, and a SUPER-SKEWED group saw the verbs with the distribution 31-7-4-4-4-4. There are several differences between the input used in this study and that used in the Casenhiser and Goldberg studies. Siebenborn et al. used real verbs, and witnessed exemplars included six quite varied pretended actions, e.g., drinking, cutting, throwing, knitting, washing, and wiping. Exemplars that were used as test items were also quite distinct from the witnessed exemplars, e.g., building, folding, catching, drilling, mixing and tying. Thus the slot was quite general, in that the construction allowed any action that could be pretended. In addition, the most frequent exemplar was not in any sense a prototypical exemplar, so that the skewed input was not centered the way the Elio and Anderson (1984) and Casenhiser and Goldberg categories were. In fact, if skewed input serves to anchor the category such that the high frequency exemplars serve as a prototype, learners in the Siebenborn et al. study were arguably misled to expect that the construction was centered around one or two particular actions such as pretending to DRINK, although the particular verbs were counterbalanced across groups.

And yet, as was the case in the Casenhiser and Goldberg work, the skewed group out-performed the balanced group when the form of the construction involved the novel SOV word order. Interestingly, the super-skewed group performed the worst, most likely because it did not provide sufficient variation, since there were 31 tokens of the high frequency exemplar, compared with only 4 (or in the case of one verb, 7) tokens of the other five exemplars. Here, despite the fact that the slot was fairly general, Siebenborn et al. found that some degree of skew in the input facilitated the recognition of an abstract category (generalizationV).
At the same time, Siebenborn, Krajewski and Lieven (forthcoming) ran another study that only differed in terms of the form of the novel construction. Instead of using a non-English word order, the novel construction was indicated with typical English SVO order, while each familiar verb was prefixed with a morpheme, va-. In this case, the results were markedly different. In particular, they found that balanced distribution was more helpful than skewed input; (once again super-skewed input was the least helpful). This finding appears to contradict their first experiment as well as the Casenhiser and Goldberg studies in which learning was facilitated by skewed (although not super-skewed) input. What exactly is the difference between the two forms of the novel construction? Why should learners do better with balanced input when a prefix was used to indicate the novel construction, and worse when word order was used?

Note that attaching the prefix, va- to familiar verbs provided a dimension of concrete similarity among all exemplars which was a not present when word order alone was used. In this way, the prefix can be expected to facilitate generalization\textsubscript{V}. Without such concrete similarity, some degree of skewedness in the input was needed to identify a pattern in the noisy input.

At the same time, since the “pretended action” category is quite broad, requiring generalization\textsubscript{H}, learners ultimately benefit from varied input. This predicts that children who heard prefixed forms presented with a balanced distribution of verbs should have outperformed children who heard skewed input with the novel word order. Although this comparison did not reach statistical significance, it did trend numerically in the predicted direction (p=0.096; Krajewski, personal communication 8/4/11).

Thus, the difference between narrowly or broadly defined categories allows us to reconcile principles 1 and 2 and certain apparently contradictory findings in the language acquisition literature. Concrete similarity among exemplars (lower, but non-zero entropy) is advantageous for initial recognition of an abstract, constrained category: generalization\textsubscript{V}. High variability (high entropy) leads to the broadening of categories: generalization\textsubscript{H}.

There are other results in second language learning that lend themselves to a different explanation that depends on the means of forming generalization. In particular, Year and Gordon (2009) performed a study of second language learners, specifically Korean middle school students who were exposed to two English dative constructions: ditransitives and prepositional datives. As in the other studies, overall type and token frequency were held constant, but two groups differed in whether they received skewed or balanced input. In the skewed group, give occurred much more frequently than other verbs, while the balanced group witnessed input in which all verbs occurred the same number of times. Year and Gordon discuss production and acceptability judgment data for sentences like (3), and find that no pairwise comparisons between the skewed and balanced groups show statistically significant differences.

(3) Peter sent his boss a gift.

That is, while it is clear that both groups have learned something about the ditransitive construction -- they both perform significantly better than controls on the two tasks-- neither group had an advantage over the other (cf. also McDonough and Nekrasova-
In this case, a prototypical instance, *give*, was used as the most frequent exemplar in the skewed condition and the construction’s semantics is abstract but constrained.

Why did this study fail to find an advantage for skewed input? The explanation may lie in the fact that these studies taught older children or adults a construction in a classroom as a second language learning task. There are several lines of evidence suggesting that while first language learning is largely implicit, second language learning relies on a much more explicit mode of learning (Krashen, 1981; Bley-Vroman, 1991; DeKeyser, 2003; R. Ellis, 1994, 2004; Reber & Allen, 2000). This distinction is noteworthy, as learning mode has been shown to have important consequences for language acquisition (Hulstijn, 2005), as well as for category formation more generally.

Explicit hypothesis testing can lead to a focus on a single dimension as the learner tries to anticipate which dimensions are relevant, because correlations among dimensions are challenging to reason about explicitly. For example, in explicit sorting paradigms that ask participants to sort exemplars into categories, there exists a strong, domain-independent bias towards sorting on the basis of a single dimension, even with categories that are designed to resist such one-dimensional sorts in favor of a sort based on a family resemblance (Medin et al. 1987). One-dimensional sorting has been found even with large numbers of relevant dimensions (Smith, 1981), holistic stimuli, and stimuli for which a straightforward multidimensional descriptor was available (Regehr and Brooks, 1995). The emphasis on a single dimension can lead to quick and efficient learning, but it can also lead to errors when an irrelevant dimension is judged important or when relevant dimensions are overlooked. In particular, while shared similarity among exemplars can encourage generalization, it can also lead to an explicit focus on the shared similarity itself; if the shared similarity does not determine the relevant dimension for generalization, explicit instructions to form a category may paradoxically lead to less accurate generalization.

Elio and Anderson (1984) in fact found there to be an interesting interaction between the way generalizations are formed—whether implicitly or explicitly—and whether skewed or balanced input was helpful. The advantage of skewed input held when learning was implicit, where implicit learning involves knowledge that is not accessible to consciousness, fairly complex and abstract, and is an incidental consequence of some task demand (Seger, 1994). When participants were explicitly asked to form hypotheses about what criteria governed category membership, the advantage of learning the centered instances before other exemplars disappeared. Elio and Anderson reason that the interaction is based on the fact that when one is in a more explicit-hypothesis testing mode, participants’ memory is less strong for the instances that lead to the hypothesis (Trabasso and Bower 1966). As just noted, when participants are explicitly instructed to learn a category, they tend to strategically attend to one or another single dimension of the input so that the exemplars are retained in a more schematic fashion.

The learning involved in the original Goldberg et al. (2004) and Casenhiser & Goldberg (2005) was implicit since no direct instructions were given to the subjects regarding what they were expected to learn during the exposure and participants were generally unable to articulate what the category meant after exposure.
In a study detailed below, we set out to determine whether we could replicate the advantage for skewed input first reported for adults in Goldberg, Casenhiser and White (2007), while investigating whether the advantage of skewed input over balanced would reverse if we explicitly asked participants to learn the novel construction.

Methods

Participants
52 participants, randomly and equally assigned to one of the four groups, took part in the study, in exchange for credit towards fulfilling a course requirement. All participants were Princeton University undergraduate students (aged 18-22), and were native speakers of English.

Procedure
The experimental procedure consisted of an exposure phase and a testing phase. During exposure, participants viewed short video clips (6-10 seconds) of various scenes, overlaid with a linguistic narration. In each of the scenes, an object would come to appear on or in a certain location. For example, a king would fall out of the sky into a chair, or a ball would roll onto a stage. The linguistic narration described this scene with the following form:

\[
\text{noun phrase}_{\text{theme}} \quad \text{noun phrase}_{\text{location}} \quad \text{nonsense verb}
\]

The audio description was given in the simple present tense at the start of the scene and was repeated in the past tense at the end of the scene. For example, given a scene where a spot appears on the king’s nose, subjects heard, The spot the king moopo... The spot the king moopoed. The entity named by the first noun phrase appears in the place named by the second noun phrase.

The experimental design was 2 (explicit vs. implicit instructions) X 2 (balanced vs. skewed exposure) in a between-subjects design. In order to manipulate whether learners were likely to learn the category explicitly or implicitly, different instructions were given before the exposure was provided. In particular, in the IMPLICIT condition participants were simply told, “Please watch the videos, and pay attention to the language that describes them.” In the EXPLICIT condition, they heard: “In these videos, there is a pattern between the actions and the language that describes them. As you watch, try to figure out what the pattern is.”

There were eight separate scenes of appearance, with each scene being repeated twice for a total of 16 trials. In the BALANCED exposure conditions, participants heard five novel verbs, each of which had a relatively low token frequency (4-4-4-2-2). In the skewed frequency condition, participants heard the same five novel verbs, but a single verb accounted for half of the sixteen tokens (8-2-2-2-2). A highly frequent verb was paired with actions depicting various manners of appearance, so that it could be construed to have the prototypical meaning of the construction “to appear.”

After this exposure, participants were given a test to gauge how well they were able to learn the new construction. This took the form of 12 forced-choice trials, in which two videos played side by side, with one overlapping narration. Participants were instructed
to select which video the narration referred to. The videos consisted of either a novel act of appearance (a balloon falling onto a monster) or an intransitive action (a monster biting a balloon). The first 6 trials (henceforth ‘appearance’ trials) involved new instances of the novel construction in which selecting the scene of appearance was the correct choice. The final six trials were narrated by a transitive utterance, also involving a novel verb (e.g. the monster *eats the balloon*), thus making the transitive scene the right choice. The procedure was intended to measure how well participants were able to distinguish the novel construction from a construction that they are already familiar with (the transitive).

Results

Novel construction test trials

A 2X2 ANOVA (Explicit v Implicit Instruction X Skewed v Balanced), reveals a significant main effect for instruction (F(1,77)=9.25, p<.003), with an advantage for implicit (µ = 75.6, σ = 13) over explicit (µ = 63.4, σ = 22).

We do not find a significant main effect of skewing (F(1,77)=.231, p=.632); i.e., the difference between balanced (µ = 68, σ = 13), and skewed (µ = 70, σ = 23) presentations is not significant overall. But importantly, there is a significant interaction between the two variables (F(1,77)=7.4, p=.008). Within the implicit instruction group, participants who received a skewed presentation were more accurate than those in the balanced condition (F(1,25)=7.3, p=.012). This was not the case in the explicit condition, where performances did not differ significantly depending on skew (F(1,25)=1.07 p=.311). Here in fact, the means if anything trend in the opposite direction, with a numerical advantage for a balanced input (see Figure 1). Balanced presentation types do not differ significantly across the different instruction conditions (F(1,25) = .056, p=.816).

As determined by one sample t-tests against chance (.5), correcting for multiple comparisons using Bonferroni(α/4), all groups except Explicit/Skew were above chance: Explicit/Balanced: µ = 67.9, σ = 24; t(12) = 2.69, p = .01; Explicit/Skew: µ = 73, σ = 21, t(12) = 3.9, p = .002; Implicit/Balanced: µ = 70, σ = 21, t(12) = 3.4, p = .005; Implicit/Skew: µ = 75, σ = 14, t(12) = 6.325, p<.001, a 2X2 ANOVA reveals that none of these differences approached significance (F(1,47) = .297, p = .878).

Transitive test trials

The effects on trials of the novel construction contrast with results on the transitive trials, where performance is consistent across the different manipulations. While all sub-groups performed significantly above chance level, considering Bonferroni’s(α/4) correction (Explicit/Balanced: µ = 67.9, σ = 24; t(12) = 2.69, p = .01; Explicit/Skew: µ = 73, σ = 21, t(12) = 3.9, p = .002; Implicit/Balanced: µ = 70, σ = 21, t(12) = 3.4, p = .005; Implicit/Skew: µ = 75, σ = 14, t(12) = 6.325, p<.001), a 2X2 ANOVA reveals that none of these differences approached significance (F(1,47) = .297, p = .878).

Discussion
Within the implicit group, we find an advantage for the skewed input, replicating the findings of Casenhiser and Goldberg (2005) and Goldberg et al. (2004). This effect disappears when explicit task instructions are given. This interaction mirrors the results of Elia and Anderson (1984), who also found an advantage for skewed input but only when implicit instructions were given, in a non-linguistic task.

No differences among conditions are evident when participants were tested on the familiar transitive construction. This suggests that the effect of skewed input in the implicit condition is not due to potentially broad effects such as level of motivation, but to a specific effect of the learning process.

As noted above, in the skewed/explicit condition, participants are more likely to attend to an irrelevant feature dimension—in this case, quite possibly the repeated nonsense verb. When they hear more representative portions of the sample and discover that the high frequency nonsense verb is not relevant to the generalization, it may be difficult to reformulate their hypothesis due to relatively poor memory for the witnessed exemplars.

Relevantly, first language learning is an excellent example of implicit learning, since it is largely learned below the level of consciousness, is very complex and ultimately quite abstract, is a consequence of trying to communicate, and is preserved in cases of amnesia (Goldberg, Casenhiser, and Sethuraman, 2004: note 4).

At the same time, second language learning by adults in a classroom is typically explicit learning insofar as the learners are actively attempting to form hypotheses to account for the exemplars they witness. Thus the lack of an advantage on acceptable instances in the Year and Gordon (2010) study may have been due to learners adopting explicit learning strategies.

There is another interesting aspect of the second learner results that raises the possibility that learners in the balanced condition were more likely to generalize broadly as we might expect from the more varied (high entropy) input. Put differently, there is some evidence that a more appropriately constrained category was in fact learned in the skewed condition, at least initially. Year and Gordon (2009) only reported acceptability judgments from acceptable sentences like (3), but Year (2009) reports a wider set of results from the same experiment that are relevant here. In addition to collecting judgments for sentences for acceptable sentences, she also collected judgments that violate the semantic constraints on the construction including examples like (4):

(4) ?? Peter moved the library the books.

While after a full 8 weeks of exposure, there was no difference between groups, early on there was. In particular, in early testing—after 4 days of exposure—-the skewed group performed significantly better on examples like that in (4), assigning significantly lower ratings to ungrammatical sentences than the balanced group. That is, the skewed input group formed a more appropriately constrained category. The balanced group formed a broader category than was warranted by the data; they were at least initially more prone to overgeneralization. This makes sense insofar as higher variability, as in the balanced group, encourages generalization.

Another recent study involving Thai second language learners of English shows a similar effect. McDonough and Nekrasova-Becker (to appear) exposed young adults to
the English ditransitive construction in three groups: skewed first, skewed, and balanced exposure conditions. In the skewed conditions one verb, *send*, was heard 8 times. The other verbs-- *pass, owe, teach, fix*--were witnessed three times each. In the balanced condition, the same five verbs were witnessed four times each. Thus overall type and token frequency were controlled for: all learners were exposed to 20 double object sentences with five different verbs.

At test, learners were presented with 10 double object expressions that involved two inanimate NPs, e.g., *Tom showed the newspaper the photographs*, or *Steve brought the car some oil*. Participants were asked “What did Tom show?” or “What did Steve bring?” This was done for the purpose of determining whether participants realized the first NP was intended to be the recipient (unlike in the Thai construction, see note 2). This makes the results difficult to interpret, because examples with inanimate recipients are not only non-prototypical, they are not acceptable, unless the inanimate N is understood metaphorically or metonymically to stand for some person or group of people. We do not know whether participants assigned such interpretations to the sentences or not.

Importantly, the exposure items did not involve inanimate recipients in this way, but included much more natural double object expressions such as *Tom fixed them an apple pie* (pg 18), that involved animate and pronominal recipients. If participants learned (correctly) that the double object construction requires the interpretation of an animate recipient—that is overwhelmingly expressed as a pronoun—then they should not consider the examples with inanimate lexical recipients as instances of the construction. That is, the finding can be viewed as being in line with what Year (2009) found: that learners in the skewed condition were more aware of the construction’s constraints than learners in the balanced condition. Whether the test examples are considered grammatical or not is therefore debatable, since it determines whether we conclude that participants in the balanced or in the skewed condition performed better. But in either case, we can conclude that the participants in the balanced condition were more likely to extend their category broadly, as is expected if variability encourages to generalization.

Thus a closer look at the language acquisition studies indicates that two distinct components of the generalization process and two distinct manners of learning may lead to different outcomes. Concrete similarity, skewed input, or lower entropy encourages the recognition of an abstract but constrained category (vertical generalization). High variability, balanced input, or higher entropy encourages broadening of a category or even an open slot (horizontal generalization). Thus whether skewed input is better or not depends in part on what type of category needs to be learned.

Experimental results reported here demonstrate that the manner of learning matters as well. If learning is implicit without a conscious attempt to form a generalization, skewed input can provide an advantage. But if learners are explicitly trying to form an hypothesis, skewed input may result in learners forming inaccurate hypotheses that are harder to correct. Thus it is not at all clear that skewed input can provide an advantage in a classroom setting because the setting itself encourages learners to reason more explicitly.

**Conclusion**
Two types of findings call into question the general facilitation of skewed input. In work on artificial grammar learning, it is often pointed out that the greater variability or greater entropy leads to generalization. This appears to be at odds with the idea that skewed input helps learning insofar as skewed input provides a frequently repeating exemplar, which involves less variability and lower entropy. Extant findings can be reconciled, however, by recognizing two components of the generalization process and two means of generalizing. The two components of the generalization process include a) recognizing that an abstract but constrained category exists, and b) recognizing a more fully open slot. The two ways to generalize from instances are to implicitly generalize from accruing exemplars and to explicitly test hypotheses.

A greater recognition of these distinctions may have important practical implications as well. As discussed earlier, second language learning tends to rely on more explicit learning strategies, and as such, responds differently to different types of input. Efforts to teach broad grammatical categories in a second language would therefore be best suited by employing a more variable distribution of input.

Together with the findings in in Elio and Anderson (1984), the present work supports the link between linguistic and non-linguistic categorization. These strong parallels suggest that both processes are accomplished by, or rely on, at least in part, a common mechanism.

Acknowledgements

We wish to thank Patricia Brooks and Grzegorz Krajewski for helpful discussion. We are grateful to Berlin’s Einstein Foundation and the Freie University of Berlin for financial support.
References


Figure 1: Performance across groups on ‘appearance’ trials (* = p<.05).
Footnotes:

1 If items were identical, no generalization would need to be formed, so both some similarity and some variety is required for generalization.

2 The degree of skew was also arguably more extreme, with the most frequent exemplar occurring almost 8 times as often as the lower frequency items (31 vs 4); in the Goldberg and Casenhiser studies, the ratio was at most 4 to 1.

3 This fact is also discussed as a commentary in Boyd and Goldberg (2009)

4 Although participants had studied English for an average of 11 years, they still underperformed on a pretest to determine whether they interpreted double object expressions correctly. The authors explain that in Thai, a lexical verb “give” (hai) is used with the order V Theme (kae) Rec. Kae is an optional preposition “to” that can appear between the theme and recipient arguments. Other verbs obligatorily use hai between the Theme and Rec, perhaps as a serial verb construction. The consistent ordering in Thai of Theme Recipient may be what makes learning the English construction difficult. Give was not used in the experiment because of the parallel in Thai.

5 McDonough and Kim frame the results as evidence that participants learned better in the balanced condition because they assume that their test sentences should be treated as acceptable by the learners. There found no difference between skewed and skewed first conditions.