Adjusting the Use of Generalizations Based on Audience Expertise

Jeff Coon (jcoon1@uci.edu)
Department of Cognitive Sciences
2201 Social & Behavioral Sciences Gateway
Irvine, CA 53706 USA

Alexander Etz (alexetz@utexas.edu)
Department of Psychology
108 East Dean Keaton St
Austin, TX 78712 USA

Gregory Scontras (g.scontras@uci.edu)
Department of Language Science
Social Science Plaza B
Irvine, CA 53706 USA

Barbara W Sarnecka (sarnecka@uci.edu)
Department of Cognitive Sciences
2201 Social & Behavioral Sciences Gateway
Irvine, CA 53706 USA

Abstract

Generalizations are a fundamental linguistic tool for efficiently passing along information. To interpret the intended strength of a generalization, listeners rely on prior knowledge. Experienced and inexperienced listeners may interpret the same generalization differently, potentially leading to miscommunication. Speakers could mitigate such miscommunication by avoiding generalizations that inexperienced listeners are likely to misinterpret. However, experienced speakers may struggle to understand the perspective of an inexperienced listener. The present study examined whether experienced speakers adjust their use of generalizations based on the expertise of their intended audience. Results showed that any such adjustments are minimal and insufficient to avoid miscommunication as operationally defined. Future research may clarify the practical impact of such miscommunication by examining how generalizations are used in relation to speakers' and listeners' goals.

Keywords: Generalizations; Expertise; Curse of Knowledge; Pragmatics; Prior Knowledge; Esports; Bayesian Modeling

Introduction

One nice part of being human is that we do not have to figure everything out for ourselves. We have access to a trove of knowledge, assembled as experts pass on their knowledge to less experienced people. A particularly efficient way to pass on knowledge is via generalizations. Rather than referring to specific instances, generalizations refer to a whole category, licensing learners to make inferences about novel category members (Gelman, Star, & Flukes, 2002). For example, to explain ducks to someone who had never encountered them, we might use simple generalizations like "ducks are birds," "ducks lay eggs," or "ducks carry avian flu." These statements reference ducks in general and thus encourage inferences about previously unencountered ducks.

These examples also demonstrate how flexible generalizations are. "Ducks are birds" applies to every duck, but "ducks lay eggs" only applies to mature female ducks, and "ducks carry avian flu" only applies to a tiny minority of ducks. Upon hearing such statements, the challenge for a listener is in figuring out how many members of the category are described by each statement.

Past research indicates that listeners who are familiar with the subject under discussion can interpret generalizations flexibly, adjusting their application as appropriate (Tessler & Goodman, 2019b; Coon, Etz, Scontras, & Sarnecka, 2021). Someone who already knows about avian flu and how it relates to ducks is unlikely to think "ducks carry avian flu" applies to every duck. Yet listeners who lack directly relevant experience consistently interpret generalizations as applying broadly throughout the category (Cimpian, Brandone, & Gelman, 2010; Coon et al., 2021). Thus, when experienced speakers use generalizations to teach inexperienced listeners, there can be miscommunication. Experienced speakers sometimes use generalizations that they interpret narrowly but which their inexperienced audience interprets broadly.

If experienced speakers know they are speaking to an inexperienced audience, they might only use broadly applicable generalizations, dismissing more narrowly applicable generalizations as too nuanced for inexperienced listeners to worry about. And if experienced speakers do make such an adjustment, inexperienced listeners would be fully justified in interpreting generalizations as broadly applicable. However, it can be challenging for people to put themselves in the shoes of someone with less experience, a difficulty referred to as the "curse of knowledge" (Camerer, Loewenstein, & Weber, 1989). As a result of this phenomenon, experienced speakers might neglect a listener's lack of experience when evaluating whether a generalization is worth making.

In the present study, we examined two questions: (1) do experienced speakers adjust their use of generalizations in response to the expertise of their audience and (2) if so, is it sufficient to preclude miscommunication?

We operationally defined miscommunication as a speaker and listener failing to align their understanding of how broadly a generalization applies, and we discuss the implications of alternative definitions. The applicability of a generalization is often an important part of its message. For example, underestimating the applicability of "ducks carry avian flu" could lead people to ignore a relevant danger. On the other hand, overestimating its applicability could cause a needlessly exaggerated fear of ducks. Speakers might prevent such miscommunication by avoiding generalizations that their audience would interpret differently. However, speakers might not realize that their audience will interpret some generalizations differently than they would, as predicted by the curse of knowledge.

Experimental Setting

We used the esport *League of Legends* as a naturalistic setting in which to observe how people use and interpret generalizations. *League of Legends* players are often highly motivated

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

Figure 1: Matchup example. Note that in this condition, the participant's intended audience is a new player. "Laning phase" refers to the early part of a game.

to learn about this environment and develop varying degrees of authentic expertise.

The structure of the game is also well suited to this type of study. Each game is a competition between two teams composed of five players each. Prior to a game, there is a drafting phase during which each player selects the character they will use. The primary objective for each team is to destroy the other team's base. Since *League of Legends* is a team game, teammates need to align their expectations about how a game will progress. Discussions of strategy thus often involve generalizations that describe how a game will likely play out while crucially allowing for exceptions. We can measure how broadly a generalization applies in terms of how frequent the exceptions are across a set of games; *League of Legends* is iterative in that the environment resets to the same initial state before each game.

While the generalizations examined in many of the past studies discussed above are generic statements (Cimpian et al., 2010; Tessler & Goodman, 2019b), the generalizations examined in this study are more specifically habituals. The distinction is in the entity about which the statement is generalizing. Generic statements apply a trait to a category, generalizing the trait over instances of that category (e.g., "ducks carry avian flu"; Carlson, 1977; Carlson & Pelletier, 1995). Habituals apply a trait to an individual, generalizing that trait over time. For example, "Toby climbs mountains" generalizes about the behavior of an individual. In terms of the semantic challenge they present, habituals can be thought of as a subset of generic statements (Carlson, 2006; Tessler & Goodman, 2019a). In both cases, it is unclear how broadly the statement should be applied. How many ducks need to carry avian flu for the category to be described as carrying avian flu? How often does Toby need to climb mountains to be described as someone who climbs mountains?

In the present study, we asked participants to reason about how various character matchups would play out in the first part of a game. For approximately the first 15 minutes, commonly known as the "laning phase," players generally stay in their designated positions, competing against the player from the opposing team who is directly opposite them. Players can choose characters who tend to excel in this early part of the game or characters who tend to excel later on. Importantly for the present study, there are also exceptions to these general trends.

Methods

Participants We recruited experienced participants (n=75) from online *League of Legends* forums. To ensure that experienced participants were well acquainted with the domain, we only included those ranked in the 27th percentile (Silver tier) of players or higher. This sample was mostly men (84.1%).

We also recruited inexperienced participants (n=25) from the undergraduate pool of a research university. We excluded 10 additional participants because they indicated some experience with *League of Legends*. It is worth noting that this sample was mostly women (80%), but we have no theoretical reason to expect gender to impact results.

Materials Participants were first given a basic overview of *League of Legends*, including definitions of key terms related to the game (e.g., phases of the game, resources such as gold). Experienced participants then completed an online survey in which they were shown a series of matchups between two opposing characters (Figure 1). They were asked about which of the two gets more gold in the laning (i.e., early) phase of the game. Because collecting more resources confers an advantage, and gold is a fundamental resource, collecting more gold was our operational definition of gaining an early advantage.

First, experienced participants were asked whether they would endorse a generalization saying that a given character "gets more gold" in the early game. Participants could answer "yes," "no," or "I don't know." We coded "I don't know" responses as participants not making the generalization. Between subjects, we varied the expertise of the audience to which the generalizations would be directed. One group of participants was asked whether they would make the generalizations in explaining the matchup to a "new player." The other group was asked about generalizations made to an "experienced player." To emphasize this manipulation, the audience's experience was highlighted and italicized.

Next, experienced participants were asked to estimate how often they would expect the referenced character to get more gold in the early game if the matchup was played 100 times. Finally, experienced participants were told that an expert speaker had made the generalization based on having seen the matchup play out 100 times. Experienced participants were asked to interpret the generalization in terms of how often they thought the speaker had seen the referenced character get more gold in the early game.

Inexperienced participants were only asked to interpret generalizations. All participants were asked to describe their playing experience. Participants who indicated that they had prior experience were asked to provide more detail in terms of their ELO ranking (Elo, 2008), time played, knowledge of the characters included in the experiment, and how frequently they play in the types of matchups under discussion.

We removed experienced participants who said they would make a generalization that they think applies to fewer than 10% of examples, and those who said they would not make a generalization that they think applies to more than 90% of examples. Our logic in doing so was that such outliers would have an outsized impact on our model. These criteria excluded 7 additional experienced participants.

Design Participants viewed 12 matchups in total. In some of the matchups, one character had a distinct advantage, while others were more evenly matched. In even matchups, it made no difference which character was referenced in the generalizations, so participants only saw 4 such matchups. For the imbalanced matchups, the generalizations referenced the character with a clear advantage on half of trials, and referenced the character with a clear disadvantage on the other half. Participants saw 8 such matchups. The left panel of Figure 2 demonstrates that experienced participants understand the matchups as we had intended. As the matchup under discussion becomes more disadvantageous for the referenced character, experienced participants become less likely to make a generalization about that character excelling in the matchup.

This experiment can be summarized as a 3 (character advantage: +, 0, or −) x 2 (listener type: experienced or inexperienced) design. Character advantage was varied within subjects; listener type was varied between subjects. However, the character advantage conditions are not our unit of analysis because experienced participants might reasonably disagree with each other as to the extent to which a specific character has an advantage. Instead, our modeling approach focuses on describing the process by which expert speakers decide when to endorse a generalization given their prior knowledge. We manipulate character advantage to ensure that the prior knowledge (i.e., the model inputs) experienced participants consider when deciding whether to make a generalization is sufficiently varied for us to capture the bounds of that process.

Results

Comparing Interpretations We first replicated the findings of Coon et al. (2021),[1] which showed that inexperienced listeners interpret generalizations as broadly applicable regardless of the context, whereas experienced listeners can use their prior knowledge to distinguish between broadly- and narrowly-applicable generalizations, as shown in the right frame of Figure 2. These results suggest that there could indeed be miscommunication because there are generalizations that experienced and inexperienced listeners would interpret differently, specifically those that are narrowly applicable. However, there is still the possibility that experienced speakers would not use narrowly-applicable generalizations when speaking to inexperienced audiences, avoiding the generalizations in which their own interpretations would differ markedly from that of their audience.

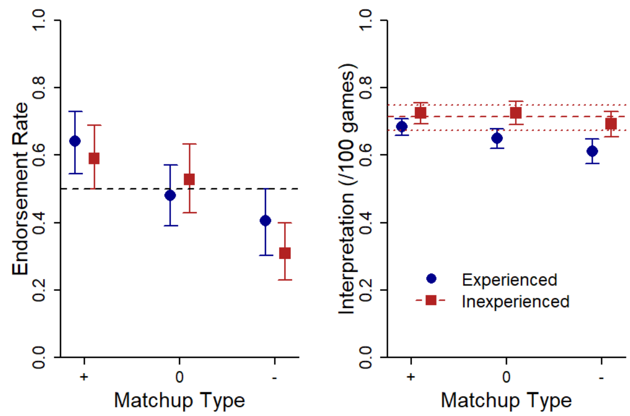

Figure 2: **Left:** Proportion of experienced participants who would endorse the generalization that the referenced character would excel in the given matchup. "Experienced" and "Inexperienced" refers to intended audience. Symbols +, 0, and − designate referenced character's advantage (or lack thereof) in the given matchup. **Right:** Experienced and inexperienced participants' interpretations of generalizations by matchup type. "Experienced" and "Inexperienced" refers to participants' own expertise. All error bars are 95% bootstrapped confidence intervals.

Modeling Approach To examine whether experienced speakers alter their process for deciding when to use generalizations, we first need a working model of that process. Unlike past efforts to model this process (e.g., Tessler & Goodman, 2019a), our model is more descriptive than cognitive because we are looking to summarize people's behavior such that we can look for patterns. To our knowledge, the only previous effort to descriptively quantify the process by which people decide to make a generalization did so by taking the mean of participants' prior estimates for generalizations they endorsed (Cimpian et al., 2010). While this approach was adequate in the context of the specific purpose for which it was used, it neglects information provided by generalizations that a participant rejected.

We instead used a hierarchical model in which a speaker's decision to make a generalization is a binary choice, made with some probability θ. This probability is in turn determined by a function which takes as its input the speaker's estimate of how broadly applicable the generalization is. Our primary interest is in characterizing how this functional relationship differs when the audience is an experienced player versus a new player.

In selecting a specific functional form, we want to capture

[1] Adopting the same model comparison methodology as Coon et al., we found that the best performing model assumed that experienced participants varied their interpretations by condition and that inexperienced participants did not.

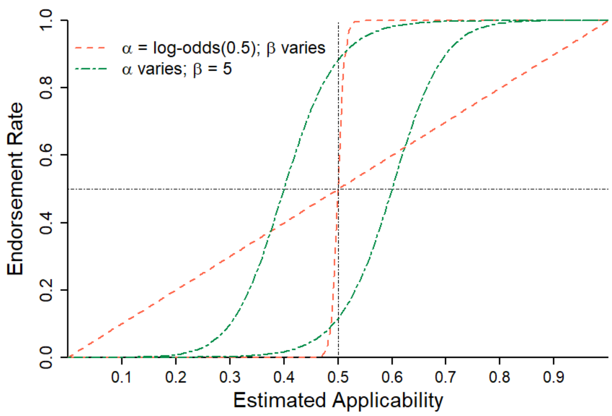

Figure 3: Illustrative example of logistic model varying in terms of shift, $\alpha = \text{log-odds}(0.4, 0.6)$, and scale, $\beta = (1, 45)$.

the important aspects of the relationship without making unnecessary assumptions. First, we assume that the probability of a generalization being endorsed will increase monotonically from 0 to 1 as a function of the speaker's belief of how broadly applicable it is. We also assume the function modeling this process can vary along two dimensions: shift and scale. The shift parameter determines the threshold at which a speaker's beliefs make them more likely than not to endorse the generalization. The scale parameter determines the steepness of the function around the threshold. In other words, it controls how strict that threshold is.

These assumptions together suggest a two-parameter sigmoidal function. We chose a logistic model, which can be specified as

$$\theta = \frac{1}{1 + e^{-\beta(x-\alpha)}}, \tag{1}$$

where θ is the probability of endorsing a generalization, and x is the log-odds of the speaker's belief as to how broadly applicable the generalization is. With this parameterization, α controls the location of the threshold (i.e., shift) and β controls the steepness of the curve (i.e., scale). As demonstrated in Figure 3, the β scale parameter can create a step function, corresponding to a strict threshold. At the other extreme, it can create a consistent and gradual slope.

We allow the functional relationship to differ depending on the audience, with separate group-level parameters α_{new} and β_{new} for the new-player (inexperienced) audience versus α_{exp} and β_{exp} for the experienced audience. We test this assumption in investigating the first question below, by comparing the group-level parameters of each condition.

To account for individual differences, we also allow the functional relationship to vary across participants. In the model, the ith participant's function is specified by offsets from the group-level parameters described above. We denote these offsets as α_i and β_i. We assume that each α_i is independently drawn from either $N(0, \sigma^2_{\alpha exp})$ or $N(0, \sigma^2_{\alpha new})$ depending on the audience condition to which the participant was assigned.[2] We assume that each β_i is independently drawn from $N(0, \sigma^2_{\beta exp})$ or $N(0, \sigma^2_{\beta new})$, again depending on audience condition. We split these σ parameters by condition to account for the possibility that one audience condition leads to more participant-level variation than the other. If participant i was assigned to the new-player audience condition, their functional relationship would be

$$\theta_i = \frac{1}{1 + e^{-(\beta_{new}+\beta_i)(x-\alpha_{new}-\alpha_i)}}. \tag{2}$$

We implemented the model in JAGS (Plummer, 2003). We chose prior distributions that respect the context of this experiment while still being relatively uninformative. We assigned the shift parameters (α_{exp} and α_{new}) prior distributions of $N(0, 2^2)$, which are agnostic about whether α is positive or negative. We assigned the scale parameters (β_{exp} and β_{new}) prior distributions of $N^+(5, 20^2)$.[3] Most of the mass of this prior is between 1 and 45. A β value of less than 1 alters the functional form in a way that is possible but unlikely for this context. A value of 45 is a soft upper bound for β, creating what is essentially a step function. Finally, we had 4 parameters governing the distributions of the participant-level α_is and β_is. For $\sigma_{\alpha exp}$ and $\sigma_{\alpha new}$ we assign priors of $N^+(0, 1^2)$. For $\sigma_{\beta exp}$ and $\sigma_{\alpha new}$ we assign priors of $N^+(0, 10^2)$.

Table 1: Model inferences for key parameters.

Parameter	Posterior Mean	95% Credible Interval
α_{exp}	0.08	(-0.01, 0.17)
α_{new}	0.18	(0.10, 0.25)
β_{exp}	3.43	(2.57, 4.50)
β_{new}	6.99	(4.67, 10.21)
$\sigma_{\alpha exp}$	0.18	(0.03, 0.36)
$\sigma_{\alpha new}$	0.16	(0.05, 0.26)
$\sigma_{\beta exp}$	1.60	(0.70, 2.65)
$\sigma_{\beta new}$	3.49	(1.57, 6.26)

Figure 4 shows the inferred group-level functional relationships alongside summaries of the empirical data. Responses were binary (experienced participants either would or would not make the generalization), but for a more useful comparison to the linking function, we have binned them based on participants' estimates as to how broadly the given generalization would apply. Each point represents the rate at which experienced participants would make a generalization if they believed its applicability fell within that range (0 to .1, .1 to .2, etc.). For example, no participants endorsed generalizations they believed would only apply 10-20% of the time. It is important to remember that we excluded participants who endorsed a generalization which they only expected to apply 0-10% of the time and those who did not endorse a general-

[2] We specify the mean and variance of normal distributions.

[3] N^+ indicates a positively truncated normal distribution.

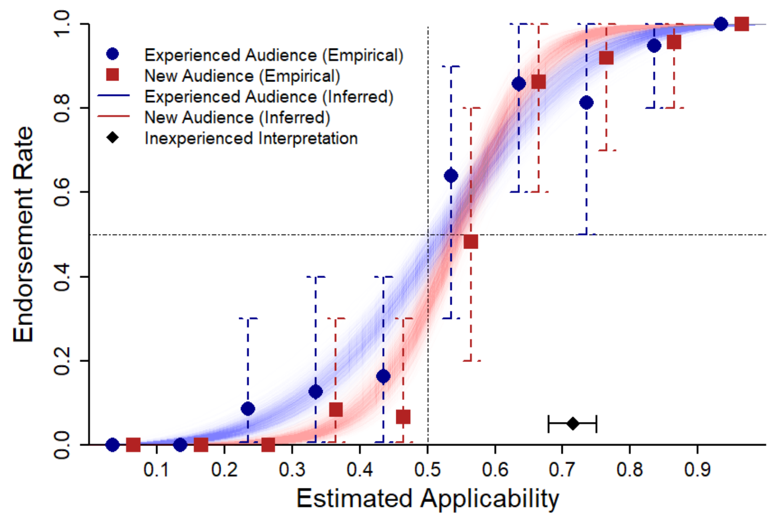

Figure 4: Proportion of experienced participants who would endorse the generalization that the referenced character excels in the given matchup. Split by audience expertise. Red and blue sigmoid lines represent a random sample from each respective condition's posterior distribution. Error bars are 95% bootstrapped confidence intervals.

ization which they expected to apply 90-100% of the time; we interpreted such responses as contaminant behavior.

We present estimates for parameters of interest in Table 1.

Question 1 Our goal in creating this model is to examine whether experienced participants adjust their use of generalizations based on the expertise of their audience. The basic model uses two parameters, shift and scale, to characterize the use of generalizations, so experienced participants can potentially adjust their use of generalizations along two dimensions. There are thus four possibilities for how the curves might compare between audience conditions: the shift changes, the scale changes, both change, or neither change.

Table 2 lists the evidence in favor of each model using Jeffreys weights – an extension of the Bayes factor for situations involving more than two models (Vandekerckhove, Matzke, & Wagenmakers, 2015).[4] If experienced participants use similar thresholds for both audiences, the data would support more parsimonious models which only have a single α shift parameter that applies to both conditions. If experienced participants are similarly strict in following the threshold for both audiences, the data would support models with a single β scale parameter that applies to both conditions.

Our data provide the most support for Model 3, which specifies separate scale parameters for each audience but only a single shift parameter. We also find considerable evidence in favor of Model 1, which specifies a single shift and single scale parameter. The evidence for Model 3 over Model 1 is minor, $BF_{31} = J_3/J_1 = 1.5$.[5] There appears to be a difference in scale when considering Figure 4, but there also remains much uncertainty in the estimates of β_{new} and β_{exp}. We find scant evidence in favor of either of the models which specify separate shift parameters. Overall, our data indicate that experienced participants may make a slight adjustment to their use of generalizations based on their audience, but only in terms of the slope of the threshold at which they become more likely than not to make a generalization, not the location of that threshold.

Table 2: Model summaries and evidence.

Model #	# of αs	# of βs	Jeffreys Weight
1	1	1	0.39
2	2	1	0.01
3	1	2	0.57
4	2	2	0.03

Question 2 Our second question is whether any adjustments participants did make based on the expertise of their audience would be sufficient to prevent miscommunication. For this experiment, we operationally define miscommunication as a speaker and listener failing to align their understanding of how broadly a generalization applies.

In terms of our model, reducing such miscommunication would manifest primarily as $\alpha_{new} > \alpha_{exp}$. In other words, speakers are less willing to use narrowly-applicable generalizations when speaking to an inexperienced audience, since

[4] We derived the Jeffreys weight J_i for models $i = \{1,2,3,4\}$ from the series of Savage-Dickey Bayes Factors for nested comparisons against model 4 (the full model), such that $J_i = BF_{i4}/\sum_j BF_{j4}$.

[5] More informed priors on the β parameters, such as $N^+(5,5)$, would lead to somewhat stronger evidence for Model 3. However, our choice of diffuse priors reflected our limited knowledge of this novel experimental context, and using more informative priors would not have changed our qualitative inferences.

an inexperienced audience would misinterpret those generalizations. It could also be helpful for the slope to get steeper ($\beta_{new} > \beta_{exp}$) as such an adjustment indicates that speakers are more rigid in their determinations of which generalizations are and are not worth making. However, α is the main parameter of interest because adjusting the scale would only help reduce miscommunication if the shift were also adjusted.

Based on our analysis for Question 1, we have strong evidence that the value of α does not change in response to the audience condition. If, for the sake of argument, we assume that there is a difference in the α values and simply test how likely that difference is to be greater for the new-player audience condition, we get a Bayes Factor of approximately 10 in favor of $\alpha_{new} > \alpha_{exp}$. Such a shift is indeed in the direction that would help reduce miscommunication, but the shift is evidently slight, if it exists at all.

Our analysis for Question 1 indicated minor evidence that the value of β changes in response to audience condition. Another directional hypothesis test, this time for β, produces a Bayes Factor of approximately 183 in favor of $\beta_{new} > \beta_{exp}$, which would indicate that experienced speakers are more strict in deciding whether a generalization is worth making if they are speaking to an inexperienced audience. However, since there is hardly any change in α, such an adjustment to β would do little to close the gap between the expectations of the speaker and the listener.

Discussion

In this study, we examined whether speakers adjust their use of generalizations based on the expertise of their audience and whether such adjustments are sufficient to avoid miscommunication. We found that, if experienced speakers made any adjustment, it was so slight that there would still be miscommunication, as we have operationally defined it. For example, Figure 4 shows that if an experienced participant thought a generalization would apply to about 60% of examples, they would be far more likely than not to make such a generalization when speaking to an inexperienced listener. Yet the inexperienced listener would interpret that generalization as applying to approximately 70% of examples.

Interpreting null results is difficult. Perhaps participants did not demonstrate impactful adjustments because our manipulation was ineffective, even though we emphasized the experience (or lack thereof) of the intended audience. In particular, our participants did not have a conversational partner to actively provide cues as to how the generalizations were being interpreted. Past research on how speakers design utterances to fit the expertise of their audience (e.g., Sulik & Lupyan, 2018) indicates that speakers use such cues to adapt over the course of a conversation. Nevertheless, our findings indicate that speakers do not consciously adjust their use of generalizations when explicitly told that the statements will be directed towards a naive listener.

In reaching this conclusion, we are defining miscommunication as the speaker and listener failing to align their understanding of how broadly applicable a given generalization is.The applicability of a generalization is particularly important if that generalization informs multifaceted decisions. For example, someone who is deciding whether they need to wear gloves when handling a duck must weigh the risk of disease (i.e., how broadly applicable is “ducks carry avian flu”) against the inconvenience of wearing gloves. In this particular experiment, the generalizations communicate information about an important strategic consideration, but this consideration is only one of many. A player may need to decide between assisting one of two teammates, both of whom have characters who could be expected to “excel in the laning phase.” To weigh a generalization’s importance relative to other considerations, listeners need a sense of how broadly it applies. We have identified situations in which that aspect of the message would be consistently distorted, leading to what could be termed a miscommunication.

Yet there are other ways of defining miscommunication, stemming from different views about what generalizations are meant to communicate. In particular, a speaker may use a generalization to alter the listener’s behavior regarding a category. For example, if the speaker’s goal in saying that “ducks carry avian flu” is to make the listener avoid ducks, then having the listener overestimate the applicability of the generalization would be more effective than having the listener understand the generalization in the nuanced way that the speaker does. In the context of our experiment, the most pronounced alteration to speakers’ use of generalizations is that their criteria for when to make a generalization may become stricter (i.e., exhibit a steeper boundary between generalizations they would and would not make) when speaking to inexperienced listeners. Such an adjustment could indicate that speakers are focused on inexperienced listeners’ behavior and are therefore reducing the world into information that the inexperienced listener should or should not act on. In doing so, speakers may ignore statistical nuances, assuming that inexperienced listeners are not yet capable of the complex decision-making that would make such nuances valuable.

It may even be adaptive for inexperienced listeners to overestimate the applicability of generalizations from experienced speakers. If an experienced speaker decides that a generalization is worth making, an inexperienced listener can safely assume that the information it contains is worth factoring into their decision-making process. From an error management perspective, erring on the side of over-applying the experienced speaker’s advice is likely less costly than erring on the side of discounting it. Also, if the generalization speaks to the only consideration for a decision (e.g., given the risk of avian flu, do I avoid this duck or not?), overestimating the applicability may help a listener avoid sub-optimal strategies.

To summarize, we hope to have demonstrated that generalizations can lead to a mismatch in how experienced speakers and inexperienced listeners conceptualize associated statistical information. Future research can examine how such a mismatch relates to speakers’ and listeners’ goals.

References

Camerer, C., Loewenstein, G., & Weber, M. (1989). The Curse of Knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*(5), 1232–1254. doi: 10.1086/261651

Carlson, G. N. (1977). *References to kinds in English*. Ph.D dissertation, University of Massachusetts, Amherst.

Carlson, G. N. (2006). Generics, habituals and iteratives. In K. Brown (Ed.), *The Encyclopedia of Language and Linguistics* (2nd edition ed., pp. 18–21). Oxford: Elsevier.

Carlson, G. N., & Pelletier, F. J. (1995). *The Generic Book*. Chicago: University of Chicago Press.

Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science*, *34*(8), 1452–1482. doi: 10.1111/j.1551-6709.2010.01126.x

Coon, J., Etz, A., Scontras, G., & Sarnecka, B. W. (2021). Experts interpret generalizations differently than novices. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43).

Elo, A. E. (2008). *The rating of chess players, past and present*. Ishi Press International.

Gelman, S. A., Star, J. R., & Flukes, J. (2002). Children's Use of Generics in Inductive Inferences. *Journal of Cognition and Development*, *3*(2), 179–199. doi: 10.1207/S15327647JCD0302_3

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria*, *124*.

Sulik, J., & Lupyan, G. (2018). Perspective taking in a novel signaling task: Effects of world knowledge and contextual constraint. *Journal of Experimental Psychology: General*, *147*(11), 1619–1640. doi: 10.1037/xge0000475

Tessler, M. H., & Goodman, N. D. (2019a). The language of generalization. *Psychological Review*, *126*(3). doi: http://dx.doi.org/10.1037/rev0000142

Tessler, M. H., & Goodman, N. D. (2019b). *Learning from generic language* (preprint). PsyArXiv. Retrieved 2019-06-13, from `https://osf.io/hnm8p` doi: 10.31234/osf.io/hnm8p

Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). *Model comparison and the principle of parsimony* (Vol. 1; J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels, Eds.). Oxford University Press. doi: 10.1093/oxfordhb/9780199957996.013.14