Title
Experts Interpret Generalizations Differently Than Novices

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Coon, Jeff
Etz, Alexander
Scontras, Gregory
et al.

Publication Date
2021

Peer reviewed

Experts Interpret Generalizations Differently Than Novices

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

Alexander Etz (aetz@uci.edu)
Department of Cognitive Sciences
2201 Social & Behavioral Sciences Gateway
Irvine, CA 53706 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

Generic statements, such as "mosquitoes fly" and "mosquitoes carry malaria," are remarkable in that they are an intuitive and readily understood means of conveying knowledge, and yet their implied prevalence—the specific quantification they convey—can vary widely. This variability may lead to miscommunication, with speakers using generic statements flexibly and listeners rigidly interpreting them as implying near universal prevalence (Cimpian, Brandone, & Gelman, 2010). However, recent research found that listeners with applicable prior knowledge interpret generic statements flexibly (Tessler & Goodman, 2019b). The evident importance of prior knowledge suggests that expertise may impact how people interpret generic statements. We investigated whether experts and novices systematically differ in the way they interpret generic statements, using the esport *League of Legends* as a cultural microcosm. As hypothesized, experts interpreted generic statements more flexibly than novices did, and novices tended to assume generic statements applied more broadly than experts did. Further research is needed to determine when these differences would lead to miscommunication.

Keywords: Prior Knowledge; Generic Statements; Expertise; Pragmatics; Esports; Bayesian Modeling

Introduction

How would you describe mosquitoes to someone who had never encountered one before? You might start by making generalizations. They are insects. They suck people's blood. They carry malaria. Generic statements such as these are remarkably helpful sources of information about the world because they distill a dizzying array of potential variation into compact facts that are either true or false (Carlson, 1977; Carlson & Pelletier, 1995). But how strong a message do they convey? "Mosquitoes are insects" applies to every category member, but "mosquitoes suck people's blood" only applies to females, and "mosquitoes carry malaria" applies to fewer than 1% of mosquitoes.

Cimpian et al. (2010) showed that this flexibility in the intended strength of generic statements can lead to miscommunication. In their experiment, "speaker" participants were asked to endorse or reject generic statements based on statistical information about novel categories (e.g., "50% of luzeks have purple tail-feathers"), while "listener" participants were asked to interpret the implied prevalence of those generic statements (e.g., "What percentage of luzeks do you think have purple tail feathers?"). Speakers were flexible in endorsing generic statements, even occasionally endorsing them when only 10% of the category had the relevant trait, but listeners consistently interpreted those generic statements as implying near-universal ($\approx$90%) prevalence.

In a recent experiment, Tessler and Goodman (2019b) showed that listeners do not always interpret generic statements so rigidly. In interpreting a generic statement, listeners have two sources of information: the generic statement itself and their own prior knowledge, which can temper any tendency toward assuming near-universal prevalence. Applying this idea to the mosquito example, a listener is unlikely to think the statement "mosquitoes carry malaria" applies to virtually every mosquito if they have some sense of how rare malaria is. So long as the speaker's and listener's prior knowledge are aligned, miscommunication should be minimal.

But what about when generic statements are used to help a less-experienced person learn? By definition, novices and experts have different prior knowledge. In the present study, we examine whether experienced and naive people differ in how rigidly and strongly they interpret generic statements. Our findings shed light on whether differences in expertise can lead to miscommunication in a domain-specific context.

The Present Study

Unlike the research cited above, we focus on habitual statements like "Michael bikes to work," which are considered a subtype of generic statement (e.g., Carlson, 2006). Whereas more conventional generic statements (e.g., "dogs have fur") are generalizations over category members, habituals are generalizations about a single entity over time. In terms of potential miscommunication, the semantics for habituals and classic generic statements are virtually identical; the statement implies some degree of prevalence, but how much is unclear (Tessler & Goodman, 2019a). For example, "Michael bikes to work" clearly indicates that Michael bikes to work in at least some instances, but how often does Michael need to bike to work for the habitual to be true? To avoid confusion, we refer to both habitual and generic statements as "generalizations."

We define expertise in terms of relative experience and knowledge; experienced participants in this study are familiar with the domain, whereas the naive participants are not. We hypothesize that naive and experienced interpretations of generalizations will differ in their rigidity and strength. Naive participants will rigidly (i.e., consistently) interpret generalizations as strong (i.e., broadly applicable), weighting the information from the generalization more heavily than their own less useful prior knowledge. Experienced participants' prior knowledge will give them access to more flexible interpretations, allowing them to interpret generalizations as weaker (i.e., applying more narrowly) when appropriate.

We use the esport League of Legends as the context for this experiment. League of Legends is a competitive online game played between two teams of 5 players each. As they play more, players accumulate experience in a complex cultural environment and are often highly motivated to do so; there are over 100 million active players (Kollar, 2016), as well as a thriving professional scene.

Because League of Legends is a team game in which coordination is invaluable, players often discuss strategy and refine their knowledge of strategy by consulting experts, either via guides and tutorials or via individualized coaching. Generalizations are a fundamental part of such discussions. The primary objective is to destroy the opposing team's base, but there are also numerous secondary objectives that give various forms of assistance towards achieving the primary objective. To prioritize secondary objectives appropriately, players must have some sense of how a given game will progress. Such predictions may find their basis in generalizations about the many variables involved. For example, some characters are adept at obtaining secondary objectives that are particularly helpful early in the game, while other characters are adept at obtaining objectives that are particularly helpful later in the game. By understanding various characters' capabilities, players can get a general sense of how a game will play out and plan accordingly.

For our purposes, we can use the iterative structure of the game, whereby the game environment resets to the same initial state before each new game, to quantify inferences about the intended strength of such generalizations in terms of how often participants think the statement would apply across a set of games.

In the present study, we examine the question of how naive and experienced people differ in their interpretation of generalizations by asking naive and experienced participants to interpret generalizations pertaining to League of Legends. We measure their interpretation of the implied strength of a generalization in terms of how frequently they expect it to be exemplified in a set of iterative games. We hypothesize that (1) experienced participants will have flexible interpretations that take into account prior knowledge, (2) naive participants will have less flexible interpretations, and (3) naive participants will tend to interpret generalizations as more broadly applicable than experienced participants do.

Methods

Participants Experienced participants (n = 49) were recruited from League of Legends online forums. They needed to be ranked in the 27$^{\text{th}}$ percentile (Silver tier) of players or higher to ensure that they were well acquainted with the domain.

Naive participants (n = 33) were recruited from a public research university's undergraduate participant pool. Participants who indicated prior familiarity with the game were removed from the naive group, and those who qualified as experienced participants were reclassified as such. Naive participants were also removed from the dataset if they spent less than 15 seconds reading the background information provided about League of Legends (of the original sample of 50 participants, 17 were eliminated for this reason).

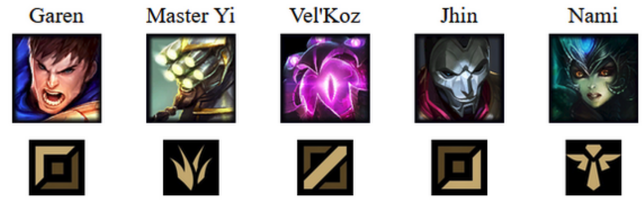

Figure 1: Composition example. Each icon indicates the specific team role being fulfilled by that character.

While these samples were very similar in terms of education and first language, there were considerable gender disparities. The sample of experienced participants was heavily male-dominated (89% male), whereas the sample of naive participants was heavily female-dominated (91% female). Again, naive participants were undergraduates who were recruited at random from a public university's participant pool. While we removed naive participants who indicated that they had prior experience with the game, this criterion only resulted in the removal of 2 females and 1 male.

Materials Participants completed an online survey in which they interpreted what a hypothetical expert player (henceforth the "knowledgeable speaker") is trying to communicate by endorsing a generalization.

The generalizations that participants interpreted pertain to a team's *composition*—the set of 5 characters that make up a team, as illustrated in Figure 1. Specifically, the generalizations are about whether a given composition excels in the early game, meaning that the players should try to build an early advantage, or in the late game, meaning that the players should try to stall the game until they realize their potential. Importantly, while the generalizations can express trends, experienced players should know that contextual factors will impact how things go in actual games. Naive participants were told that compositions could vary in terms of how likely they are to excel in the early and late game. However, they were not explicitly told about the extent to which compositions that *should* excel in the early/late game can be the victim of circumstances. Our results imply that they inferred a probabilistic causal link between team composition and ability to excel in the [early/late] game rather than a deterministic one.

The survey began with some background information. Participants saw broad explanations of how the game works, team compositions, and the phases of a game (i.e., early and late), akin to those provided above.

Following the background information, participants saw a possible composition (see Figure 1 for an example). Participants were first asked whether, in order to explain to a friend how this composition works, they would themselves endorse a series of generalizations (by choosing "yes" or "no" to indicate whether they would say "this team excels in the [early/late] game"). These questions confirmed that the generalizations in this study were ones that real players would make (see Figure 2). Naive participants confirmed their lack of prior knowledge by responding at chance.

Participants were then asked how frequently they would expect the composition to excel in the [early/late] game across a subset of 100 games (by adjusting a slider with endpoints of 0 and 100). While this question is not central to the current study, we do use it below to evaluate competing interpretations.

Next, participants were asked to imagine that, having seen the composition play 100 games against varied opposition, the knowledgeable speaker endorsed the generalization ("this team excels in the [early/late] game"). Participants were then asked to interpret the intended strength of the knowledgeable speaker's generalization in terms of what they thought had happened in the subset of games on which the knowledgeable speaker's decision was based. In other words, in how many of the 100 games seen by the knowledgeable speaker do participants think the composition excelled in the [early/late] game?

Finally, participants were asked to describe their experience with League of Legends in terms of time played, ELO ranking (Elo, 2008), and knowledge of the game's characters. This information was used to verify the classification of participants as either naive or experienced.

Design The present study can be summarized as a 2 (past experience) x 6 (composition type) design. Based on pilot data, we selected a set of 12 compositions that varied in terms of how often experienced players think they would excel in the early or late game.[1]

Specifically, we selected two compositions for each of six types: strong early game, middling early game, weak early game, strong late game, middling late game, and weak late game (abbreviated as E_+, E_0, E_-, L_+, L_0, and L_-). Participants were asked to interpret a generalization about one composition per type (chosen at random, counterbalanced across participants, presented in randomized order) for a total of six trials per participant.

Although each of the 12 compositions was selected with only their early- or late-game ability in mind, a composition that excels in the early game is generally less likely to excel in the late game and vice versa. This assumption is supported by the proportion of experienced participants who would endorse the generalizations themselves (see Figure 2).

We did not vary the two generalizations ("This composition excels in the [early/late] game.") by composition type; experienced participants were occasionally asked to interpret generalizations with which many of them would disagree. In fact, such disagreement happened 42.7% of the time.

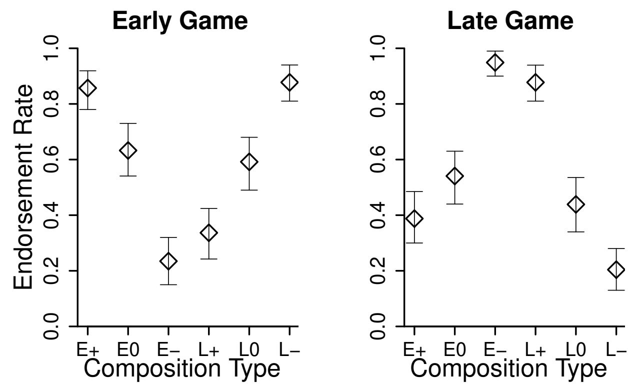

Figure 2: Proportion of experienced participants who would endorse the generalization that the given composition type excels in the early game (left) or late game (right). Error bars are 95% bootstrapped confidence interval.

Results

Modeling Approach

Our dependent variable was participants' estimates of the intended strength of a generalization. We recorded participants' responses as proportions, which represented participants' stated beliefs about how often a given composition had excelled in the [early/late] game across the subset of games on which a given generalization was based.

To analyze the results, we developed the following mixed-effects linear model. For the ith participant viewing the jth composition, we modeled their response, y_{ij}, as the additive result of three influences: an expected response for the given condition, a participant-level offset, and trial-to-trial noise. The expected response in the context of a given composition type is represented by a β parameter. Naive participants have a single β_0 given their lack of domain-specific prior knowledge that may modulate expectations by composition type. In contrast, experienced participants have a β_j specific to each of the j composition types. The ith participant's offset, α_i, is modeled as a random intercept being drawn independently from a normal distribution with a mean of zero and standard deviation τ. A positive value for α_i indicates that the participant tends to provide high responses compared to other participants, with negative α_i indicating the reverse. Finally, independent noise present on a given trial, ε_{ij}, is represented as being independently drawn from a normal distribution with a mean of zero and standard deviation σ.

Thus, the model can be written as

$$y_{ij} = \beta_0 x_0 + \sum_{j=1}^{6} \beta_j x_j + \alpha_i + \varepsilon_{ij}, \qquad (1)$$

[1]Candidate compositions were randomly generated at the rate at which they are used in the game and evaluated by >10 experts.

with $\alpha_i \sim N(0,\tau^2)$ and $\varepsilon_{ij} \sim N(0,\sigma^2)$. Each x is a binary indicator of whether the influence represented by the associated β is relevant for participant i looking at a given composition of type j. For naive participants, $x_0 = 1$, with all other $x_j = 0$; in this case, the model reduces to

$$y_{ij} = \beta_0 + \alpha_i + \varepsilon_{ij}, \tag{2}$$

because naive participants presumably do not distinguish between composition types and thus do not vary their response by composition type. We examine the validity of this assumption later.

The x_1 through x_6 variables indicate, for experienced participants, the composition type for a given trial; only one x_j will ever equal 1 on a given trial. For a trial where an experienced participant is viewing composition type j, the model reduces to

$$y_{ij} = \beta_j + \alpha_i + \varepsilon_{ij}. \tag{3}$$

This same model was applied separately to early game responses vs. late game responses, as it is not obvious *a priori* that these two generalizations are equivalent.

This modeling approach allows for inference on both the group-level (βs) and participant-level (α_is). Each participant only saw 6 generalizations each, so any inferences about an individual α_i are necessarily limited. However, including α_is in the models accounts for potential dependency in a participant's repeated responses and ensures an appropriate level of uncertainty in the estimation of the βs.

The models were implemented in JAGS (Plummer, 2003) and fit to the data using Bayesian parameter estimation. Each β was given an independent $N(0.5, 0.25^2)$ prior distribution, with almost all of its mass between 0 and 1. The mean of 0.5 was chosen because, in the context of this experiment, 0.5 would be a "neutral" response (i.e., the composition exemplified the referenced trait—excelling in the [early/late] game—as often as not). The variation of the participant-level offset, τ, and the trial-to-trial noise, σ, were both given a gamma prior with shape 1.5 and rate 2. The shape and rate parameters were chosen to not give too much credence to unreasonably large values of τ and σ.

Modeling Results: Estimation and Fit

As shown in Figure 3, the models fit the data well, with inferences closely matching the empirical means. There is a U-shaped pattern in the empirical responses for the experienced participants depending on composition type and the content of the generalization (early vs. late). In contrast, the empirical responses of naive participants show no clear pattern for the models to match, but their mean responses by composition type are still close to the model's β_0 estimate. A summary of the posterior distributions for the βs can be found in Table 1.

Variability in responses due to systematic differences between participants (τ) was comparable to variability due to noise (σ). For generalizations about the early game, the posterior estimates of σ and τ are 0.11 (95% CI [0.10, 0.119]) and 0.12 (CI [0.10, 0.14]) respectively. For generalizations about the late game, the posterior estimates of σ and τ are 0.13 (CI [0.12, 0.14]) and 0.13 (CI [0.11, 0.15]). These magnitudes are directly comparable, as they are on the same scale, but can be more conveniently interpreted once they are converted into the familiar correlation coefficient.

For a mixed-effects model using random intercepts, the implied correlation between two responses given by a participant is $\rho = \tau^2/(\tau^2+\sigma^2)$; note that the correlation between any responses from different participants is zero by assumption, after accounting for composition type and expertise. The value of ρ indicates how predictable a participant's responses will tend to be. Performing this transformation using the relevant posterior samples, we obtain posterior estimates for ρ of 0.53 (CI [0.46, 0.63]) and 0.50 (CI [0.42, 0.60]) for the early game and late game generalizations, respectively. These correlations indicate substantial individual differences across participants.

Modeling Results: Hypothesis Testing

Hypothesis 1 We hypothesized that experienced participants would interpret generalizations flexibly, varying how broadly applicable they think a generalization is depending on the context in which the generalization was made. Our specific hypothesis was that experienced participants would interpret generalizations as applying more narrowly when their own prior knowledge conflicted with the generalization (and vice versa). Figure 2 shows how often experienced participants were willing to endorse a given generalization themselves, indicating whether their prior knowledge conflicts with it. For trials in which the generalization referred to a composition's performance in the early game, experienced participants would interpret generalizations about composition type 1 (strong early game) as more broadly applicable than those about composition type 2 (middling early game), and would in turn interpret generalizations about composition type 2 as more broadly applicable than those about composition type 3 (weak early game). We expected this pattern to be inverted for the late game composition types. Furthermore, we expected these two sets of orderings to be reversed when generalizations referenced the late game. Our hypothesis translates into the following simultaneous orderings: $\beta_{E_+} > \beta_{E_0} > \beta_{E_-}$ and $\beta_{L_+} < \beta_{L_0} < \beta_{L_-}$ for early-game generalizations, and $\beta_{E_+} < \beta_{E_0} < \beta_{E_-}$ and $\beta_{L_+} > \beta_{L_0} > \beta_{L_-}$ for late-game generalizations.

As shown in Figure 3, the hypothesized orderings are clearly reflected in the early-game model posteriors and slightly less clearly in the late-game model posteriors. We quantify support for our hypothesis by computing Bayes Factors, comparing the posterior odds for the orderings to the prior odds. Complex hypotheses involving many simultaneous order constraints necessarily have very low prior odds because there are many ways *a priori* for the ordering to be violated. There are 36 ways to simultaneously order two sets of three parameters. Thus, the two hypothesized orderings each have prior odds of 1:35. We calculated the posterior odds of the hypothesized orderings by finding a) the number of joint posterior samples simultaneously obeying the order constraints, finding b) the number of samples violating the constraints, and dividing a by b.

The hypothesized orderings have posterior odds of 8.1:1 in the early-game model and posterior odds of 1.3:1 in the late-game model. Comparing the posterior odds to the prior odds, we obtain Bayes factors for the early game and late game hypotheses of about 280 and 46, respectively. Even in the case of the late-game model where uncertainty remains regarding the ordering of β_5 and β_6, the data provide relatively strong evidence in support of the hypothesized orderings.

Table 1: Inferred descriptive statistics by composition type (β_0 indicates naive participants). Approximate credible intervals for a β can be formed by taking the mean plus or minus 2 times the SD. BF_{0j} is the Savage-Dickey Bayes factor testing equality between β_0 to β_j (>1 supports equality, <1 supports a difference).

Interpretation of Early Game Generalizations

β	0	E_+	E_0	E_-	L_+	L_0	L_-
Mean	.65	.62	.56	.57	.60	.66	.70
SD	.02	.02	.02	.02	.02	.02	.02
BF_{0j}	—	3.3	1/3	1/50	1/50	1/7	6.1

Interpretation of Late Game Generalizations

β	0	E_+	E_0	E_-	L_+	L_0	L_-
Mean	.59	.63	.71	.68	.58	.57	.69
SD	.03	.03	.03	.03	.03	.03	.02
BF_{0j}	—	1/12	1.5	8.6	8.9	1/50	1/25

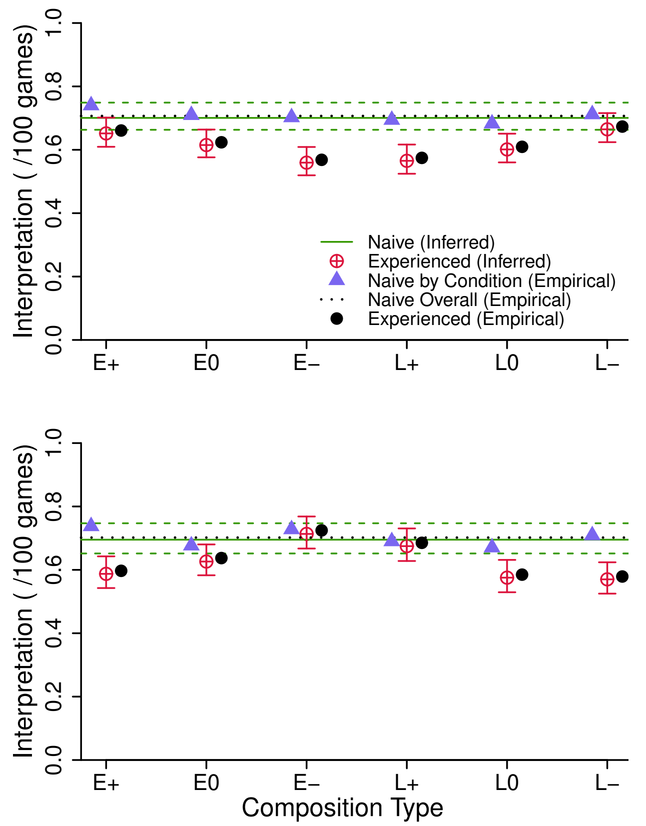

Figure 3: β parameter estimates for early game (top) and late game (bottom) generalizations. Error bars represent 95% credible intervals. Note that the result for β_0 is shown as a stretched horizontal line for easier comparisons.

Hypothesis 2 We hypothesized that naive participants would interpret generalizations rigidly (i.e., consistently), responding similarly regardless of composition type. In fact, this hypothesis is incorporated as an assumption in the models, since only β_0 is available to dictate the expected naive response. This assumption is informally supported by the close correspondence between novices' empirical means (see the blue triangles in Figure 3) and the models' estimates of β_0. To test this hypothesis formally, we created alternative models in which expected responses for naive participants could vary in response to composition type. The priors chosen were the same as those used for the experienced participants' parameters. Model comparisons were done using the Deviance Information Criterion (DIC), a metric that accounts for the inherent trade-off between model fit and complexity (Spiegelhalter et al., 2014). The expanded models do little to improve our account of the data (ΔDIC*s* of 8 and 7 favoring the simpler early and late models, respectively), indicating that naive interpretations are unaffected by composition type.

The above test only considers one aspect of rigidity, namely that of composition irrelevance. Another aspect of rigidity is the relative consistency of naive and experienced responses on a trial-to-trial basis. In other words, are naive participants' responses truly more rigid, or does their variability come in the form of general noise? Whereas composition irrelevance is reflected in the β parameters, trial-to-trial consistency is captured by the distribution of the σ noise parameter.

To test whether trial-to-trial noise differed between naive and experienced participants, we expanded the model so that each ε_{ij} is drawn from one of two normal distributions, with different variances depending on the past experience of a given participant. The priors for the new parameters, which we call σ_{exp} and σ_{nov}, were the same as that of σ in the original model.

For early game generalizations, the posterior estimate of $\sigma_{exp} - \sigma_{nov}$ is -.01 (CI [-.02, .01]). For late game generalizations, the posterior estimate of $\sigma_{exp} - \sigma_{nov}$ is -.02 (CI [-.04, 0.00]). These estimates suggest that any differences that may exist between σ_{exp} and σ_{nov} would be relatively minor. Indeed, a null hypothesis test using the Savage-Dickey method (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010) provides evidence for the equivalence of σ_{exp} and σ_{nov} (early game and late game BF_{01} of about 60 and 5.5, respectively). Since novices and experts exhibit similar trial-to-trial noise, novices not varying their responses by composition type indicates that they do indeed interpret generalizations rigidly when lacking domain-specific prior knowledge.

Hypothesis 3 We hypothesized that naive participants would give higher responses on average. Figure 3 shows that context dictates how experienced participants respond. In some contexts, the average responses for experienced participants are just as high as naive participants'. When the groups do diverge, it is clearly because experienced participants' responses are lower. The question is then not one of direction but of identifying when differences exist between experienced and naive participants.

To test this, we conducted point null hypothesis tests using the Savage-Dickey method to quantify the evidence for equality between β_0 and each β_j for early-game and late-game generalizations. Bayes Factors above 1 indicate evidence of equality between the expected naive and experienced responses, and values below 1 indicate evidence of a difference. The results of these tests are reported in Table 1. For early game generalizations, we find evidence that expected naive and experienced responses differ for Composition Types 2-5, and we find evidence for equality in Composition Types 1 and 6. For late game generalizations, we find evidence for a difference in Composition Types 1, 5, and 6, and we find evidence for equality in Composition Types 3 and 4 (Composition Type 2 is ambiguous). As such, it seems clear that naive participants do indeed tend to have stronger interpretations than experienced participants, though experienced participants' interpretations can be just as strong in certain contexts.

Discussion

Past research suggests that people who are unfamiliar with a domain may misinterpret generalizations (Cimpian et al., 2010). In the present study, we investigated whether the underlying patterns that lead to misinterpretation vary depending on relative expertise. We asked experienced and naive participants to interpret generalizations made by a hypothetical knowledgeable League of Legends player. We found support for our hypotheses that (1) experienced participants would interpret generalizations flexibly whereas (2) naive participants would interpret generalizations more rigidly and (3) as more broadly-applicable than experienced participants do. These findings are consistent with recent cognitive models of the interpretation of habituals (Tessler & Goodman, 2019b).

Regarding hypotheses 1 and 2, it makes sense that experienced participants' interpretations would be particularly flexible. Naive participants were told that the teams can vary in terms of how likely they are to excel in the [early/late] game, but only experienced participants have the prior knowledge necessary to understand *how* compositions vary, so only experienced participants adjust their interpretations accordingly. More interestingly, naive participants dealt with their uncertainty by remaining consistent in their interpretations.

Another possible interpretation of these results is that experienced participants are not flexibly interpreting the generalizations but rather disagreeing with the generalizations at different rates. According to this interpretation, experienced participants assume that their own estimate of how often a team will excel in the [early/late] game is the correct one. Given that the experienced participants know what actually happens, the knowledgeable speaker must have seen that same thing happen and somehow misinterpreted it in making the generalization. Naive participants, on the other hand, do not know what actually happens and just have to trust the knowledgeable speaker in every case, thus explaining the difference between experienced and naive behavior.

Our data do not support this interpretation.[2] If experienced participants differed from naive participants in their willingness to disagree with—and thus disregard—the knowledgeable speaker, there should be a minimal difference between their own expectations and their estimate of what the speaker saw, regardless of whether the experienced participants agree with the generalization or not. Being experienced, these participants know how the compositions play out, and nothing the speaker says will change that. Instead, such a lack of adjustment only happened when experienced participants agreed with the generalization. When experienced participants disagreed with the generalization, they interpreted it as indicating that the speaker saw more positive examples than the experienced participant would have expected. These data suggest that experienced participants are assuming that the speaker has a valid reason for making each generalization.

Our findings also support the third hypothesis, that naive participants tend to assume that generalizations apply more broadly than experienced participants do. However, this effect may be domain-dependent. The present study referenced behavioral traits of a team playing a game, while Cimpian et al. (2010) referenced physical traits of novel biological entities (e.g., "luzeks have purple feathers"). The naive participants in the present study indicated weaker interpretations than participants in Cimpian et al.'s study (i.e., they had lower average estimates of prevalence for generalizations), presumably because the behavior of a team within a game is generally less dependable than the physical features of a given kind of animal. Naive participants may not have any knowledge of the specific game, but they know about games in general.

Additionally, whether naive people assume generalizations apply more broadly than experienced people do depends on how novices update their prior knowledge as they gain experience. In our study, experienced participants tended to have more narrow interpretations, so experience in the domain would presumably teach our naive participants that there are more exceptions to the generalizations than their prior knowledge would suggest. Such a learning trajectory may be common but is surely not universal across domains. There may be cases in which experience teaches formerly-naive people that there are fewer exceptions than they had assumed.

[2] https://osf.io/qr6su/

A key limitation of this study is the gender disparity between the samples being compared. From a purely linguistic perspective, there is no theoretical reason to expect gender to influence our results. However, esports have well-documented issues with gender diversity and sexism (e.g., Kuznekoff & Rose, 2013; Kaye et al., 2018). Interpreting generalizations is an inherently social act, so this disparity should be kept in mind when interpreting our findings.

Our study suggests that interpretations of generalizations may be explained by how much domain-specific prior knowledge an individual has. Naive participants have very few previous observations in the domain, whereas experienced participants have made considerably more. Substantial prior knowledge can act as an anchor, tempering interpretations.

Naive and experienced participants also differ in the quality of their prior knowledge. Experienced participants may know more about the causal link between an underlying trait and observable behavior. In other words, they may have a better sense of the mechanism by which teams "excel in the [early/late] game." This structured knowledge resembles Goodman's (1983) "overhypotheses," where the specific hypotheses individuals consider are restricted and weighted according to higher-level knowledge about the world. Naive participants may have overhypotheses about how games work. Expert participants may have intermediate overhypotheses about how this particular game works and how the traits of excelling in the early or late game work within it, which inform their hypotheses about a given composition.

In the present study, we chose to focus on bare generics because they provide minimal lexical information and thus place minimal restrictions on the range of plausible pragmatic interpretations compared to more specific statements. This flexibility makes bare generics a particularly instructive test case. However, there are other quantificational elements (e.g., "often...","mostly...") that similarly leave room for pragmatic interpretation, and we suspect experienced and naive people would exhibit a similar pattern when interpreting such statements to that found in this study.

Social context may mitigate the extent to which naive people misinterpret generalizations. In the present experiment, the audience's expertise was ambiguous ("Imagine you are explaining this composition to a friend."). If they knew their audience was naive, experienced participants might only make strong generalizations, in which case naive participants would be justified in assuming that such generalizations should be rigidly interpreted as applying broadly. However, experts have a well-documented bias towards overestimating their audience's expertise (the curse of knowledge; Camerer, Loewenstein, & Weber, 1989).

It may even be adaptive for naive people to overestimate the applicability of expert generalizations. Griffiths, Canini, Sanborn, and Navarro (2019) point out that such overestimation can be viewed as a rational way for naive people to make the most of what little information they have. If all you know about mosquitoes is that they carry malaria, it is reasonable to avoid mosquitoes whenever possible; better to be overly cautious than unaware of potential danger. This perspective on rationality may point to a general learning trajectory by which people apply new knowledge as broadly as they reasonably can, refining their understanding as they learn more.

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.

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., p. 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.

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

Goodman, N. (1983). *Fact, fiction, and forecast*. Harvard University Press.

Griffiths, T., Canini, K., Sanborn, A. N., & Navarro, D. (2019). *Unifying rational models of categorization via the hierarchical Dirichlet process* (preprint). PsyArXiv. Retrieved from `https://osf.io/ketw3`

Kaye, L. K., Pennington, C. R., & McCann, J. J. (2018). Do casual gaming environments evoke stereotype threat? Examining the effects of explicit priming and avatar gender. *Computers in Human Behavior*, *78*, 142–150.

Kollar, P. (2016). The past, present and future of League of Legends studio Riot Games. *Polygon*.

Kuznekoff, J. H., & Rose, L. M. (2013). Communication in multiplayer gaming: Examining player responses to gender cues. *New Media & Society*, *15*(4), 541–556.

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*.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014, June). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(3), 485–493.

Tessler, M. H., & Goodman, N. D. (2019a, April). The language of generalization. *Psychological Review*, *126*(3).

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

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189.