Minding the Politeness Gap in Cross-cultural Communication

Yuka Machino[1], **Matthias Hofer**[1], **Max Siegel**[1], **Joshua B. Tenenbaum**[1], **and Robert D. Hawkins**[2]
[1]Department of Brain and Cognitive Science, MIT
[2]Department of Linguistics, Stanford University

Abstract

Misunderstandings in cross-cultural communication often arise from subtle differences in interpretation, but it is unclear whether these differences arise from the literal meanings assigned to words or from more general pragmatic factors such as norms around politeness and brevity. In this paper, we report three experiments examining how speakers of British and American English interpret intensifiers like "quite" and "very," finding support for a combination of semantic and pragmatic factors. To better understand these differences, we developed a computational cognitive model where listeners recursively reason about speakers who balance informativity, politeness, and utterance cost. A series of model comparisons suggest that cross-cultural differences in intensifier interpretation stem from (1) different literal meanings, (2) different weights on utterance cost. These findings challenge accounts based purely on semantic variation or politeness norms, demonstrating that cross-cultural differences in interpretation emerge from an intricate interplay between the two.

Introduction

One of the fundamental maxims of cooperative communication is that speakers aim to be *informative*—providing relevant and useful information to their listeners (Grice, 1975). However, speakers must balance many other considerations when choosing what to say, including politeness, efficiency, and clarity. These alternative considerations may come into tension with informativity, leading to indirectness, vagueness, hedging, and other forms of pragmatic ambiguity that listeners must resolve. For example, when interpreting the strength of an utterance containing amplifiers like "very" or downtoners like "kind of", listeners may wonder whether the speaker is literally trying to convey a stronger or weaker value on the given scale, or whether they're just trying to be polite. Consider the phrase "That was very helpful"—does the speaker mean the help was exceptionally good, or are they simply being gracious? Similarly, "I'm kind of tired" could indicate mild fatigue or serve as a polite understatement (Brown & Levinson, 1987).

How do listeners navigate this ambiguity? Recent work in computational pragmatics (Yoon et al., 2020) has formalized these tradeoffs in the Rational Speech Act (RSA) framework, where speakers are assumed to select utterances by maximizing a utility function that weighs epistemic goals against other social goals. Listeners, in turn, must reason about speakers' motivations to correctly interpret what is said. While this framework offers valuable insights into the pragmatics of polite language use, an important factor has received limited attention: the weights speakers assign to these competing goals—and listeners' expectations about these weights—may differ systematically across cultures (Goddard & Wierzbicka, 2004, 2013; White, Pandey, & Pan, 2024).

Empirical work has extensively documented such cross-cultural differences in how speakers use and interpret modifiers (Su, 2016; Stratton, 2021; Romero, 2012). For instance, the intensifier 'quite' is known to carry different meanings in British versus American English, functioning as an intensifier in American English but often as a downtoner in British English. These differences extend beyond individual lexical items to broader patterns of how modifiers function pragmatically–as devices for hedging, emphasis, or managing interpersonal dynamics (Ruzaitė, 2007; Haugh & Schneider, 2012; Schneider, 2012; Waters, 2012). Understanding these differences is crucial for successful cross-cultural communication, as misaligned expectations can lead to misinterpretation and stereotyping (Goddard, 2012; Haugh & Bousfield, 2012; Thomas, 1983). Do British and American speakers simply assign different literal meanings to these modifiers, or do they reason differently about speakers' pragmatic intentions when interpreting modified utterances?

In this study, we aim to disentangle how cultural conventions shape modifier interpretation, extending the RSA framework to account for cross-cultural variation in pragmatic inference. We focus on British and American English, designing a series of three experiments to test whether differences in modifier interpretation stem from semantic differences (literal meanings) or pragmatic differences (how speakers weigh communicative goals). We then develop a computational model building on Yoon et al. (2020) to disentangle these factors. By fitting the model to experimental data from both cultures, we infer culture-specific parameters for literal meanings, informativity weights, and utterance costs.

Our results reveal that modifier interpretation differs between cultures through both semantic variation (different thresholds for modifiers like 'quite' and 'very') and pragmatic variation (different weights on informativity and different perceived costs of modification). These findings demonstrate that cross-cultural differences in interpretation arise from a complex interplay of semantic and pragmatic factors, and highlight the importance of incorporating cultural variation into computational models of language understanding.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

Experiments

We conducted three experiments, each designed to isolate distinct explanations for differences in intensifier interpretation between English speakers in the US and UK. By systematically varying conversational context and speaker intent, our experiments aimed to disentangle the relative contributions of semantic and pragmatic factors.

Experiment 1: Dialogue Version

Before we explore potential explanations of cross-cultural variation in modifier interpretation, we must establish whether such variation exists in the first place. Experiment 1 serves as a baseline for detecting interpretation differences between UK and US participants in modifier use.

Participants We recruited 49 participants via Prolific (23 from the UK, 26 from the US) to interpret a series of utterances with and without intensifiers.

Stimuli & Design We designed stimuli consisting of seven predicates ("exhausted", "boring", "difficult", "concerned", "understandable". "impressive", "helpful") crossed with five modifiers ("slightly", "kind of", "quite", "very", "extremely"). Each predicate appeared in two versions in a within-subject design: UNMODIFIED (e.g. "You've been helpful") and MODIFIED (e.g., "You've been very helpful"). These predicate-modifier combinations were embedded in short vignettes designed to contextualize the utterance in a natural conversational setting (see Table 1). The vignettes varied in non-linguistic social features such as speaker-listener relationships, social hierarchy, and formality of interaction, ensuring that interpretation differences could not be attributed solely to these extraneous factors.

On each trial, participants were provided with a continuous slider and asked to interpret the relevant predicate (e.g. "how helpful?") on a scale from 'minimally {predicate}', and 'maximally {predicate}' (e.g. 'minimally helpful', and 'maximally helpful'). This approach provided graded responses, capturing the influence of modifiers on interpretation. Each participant provided ratings for 70 scenarios (35 modified sentences and 35 unmodified baseline sentences), allowing for a fully within-subject comparison of modifier effects. Scenarios were presented in a randomized order.

Table 1: Scenario variants and corresponding questions.

Version	Scenario	Question
Exp.1: Dialogue	Amy asks her friend, Lisa, for some tips on baking bread: *"Thank you for those tips. You've been very helpful."*	How helpful did Amy find Lisa's tips?
Exp.2: Narrator	Amy asks her friend, Lisa, for some tips on baking bread. The tips are very helpful, and Amy thanks Lisa for it.	How helpful did Amy find Lisa's tips?
Exp.3: Politeness	Amy asks her friend, Lisa, for some tips on baking bread: *"Thank you for those tips. You've been very helpful."*	How polite is Amy being?

Results To determine whether British and American participants systematically differed in their modifier interpretations, we fit a linear mixed-effects model with random effects for participant and scenario. We predicted the within-participant difference between the BASELINE and MODIFIED conditions for each scenario. To control for individual differences in scale use, we first z-scored all responses within each participant. We then computed a modifier effect score by subtracting the z-scored response of the unmodified version from the z-scored response of the modified version of each scenario thereby isolating the effect of the modifier. We compared a simpler model including only fixed effects of country (US vs. UK) and modifier against a more complex model including a country $\times$ modifier interaction. A nested likelihood ratio test revealed that the model allowing for different effects for each modifier provided a significantly better fit ($\chi^2(4) = 14.48, p = .006$), confirming systematic cross-cultural differences in modifier interpretation.

Post-hoc comparisons for individual modifiers (see leftmost graph in Figure 1) revealed that these differences were specific rather than general: "quite" was interpreted as weaker by British participants compared to Americans ($\beta = -0.28, SE = 0.12, p = .023$); "very" was interpreted as stronger by British participants ($\beta = 0.28, SE = 0.13, p = .032$), while no reliable differences emerged for the rest of the modifiers. Across both groups, the modifiers followed a consistent strength hierarchy: "slightly" $<$ "kind of" $<$ "quite" $<$ "very" $<$ "extremely" (all pairwise comparisons $p < .001$). To further contextualize influences, we tested whether modifier effects varied by predicate valence. We categorized predicates as positive (understandable, impressive, helpful) or negative (exhausted, boring, difficult, concerned) based on their typical evaluative connotations. A mixed-effects model including the additional effect of valence suggested that modifier interpretation varied systematically with valence ($\chi^2(4) = 20.83, p = .0003$), particularly for *downtoners*: both "slightly" ($p = .017$) and "kind of" ($p = .042$) showed stronger weakening effects with positive predicates compared to negative ones.

British participants' weaker interpretation of "quite" occurred only for negative predicates ($p = .022$), while their stronger interpretation of "very" was most pronounced with positive predicates ($p = .053$).

Experiment 2: Narrator Version

Our findings in Experiment 1 support the observation that intensifiers have different effects across US and UK cultures, but leave open whether these differences stem from semantic or pragmatic factors. These scenarios embedded utterances in a dialogue, where social and politeness considerations naturally arise. In Experiment 2 ("Narrator Version"), we designed variants of the same scenarios to minimize explicit politeness considerations by re-framing the final sentence as an omniscient narrator's description, akin to an inner monologue (see Table 1). In principle, this manipulation should remove the social pressure to be polite. Without a conver-

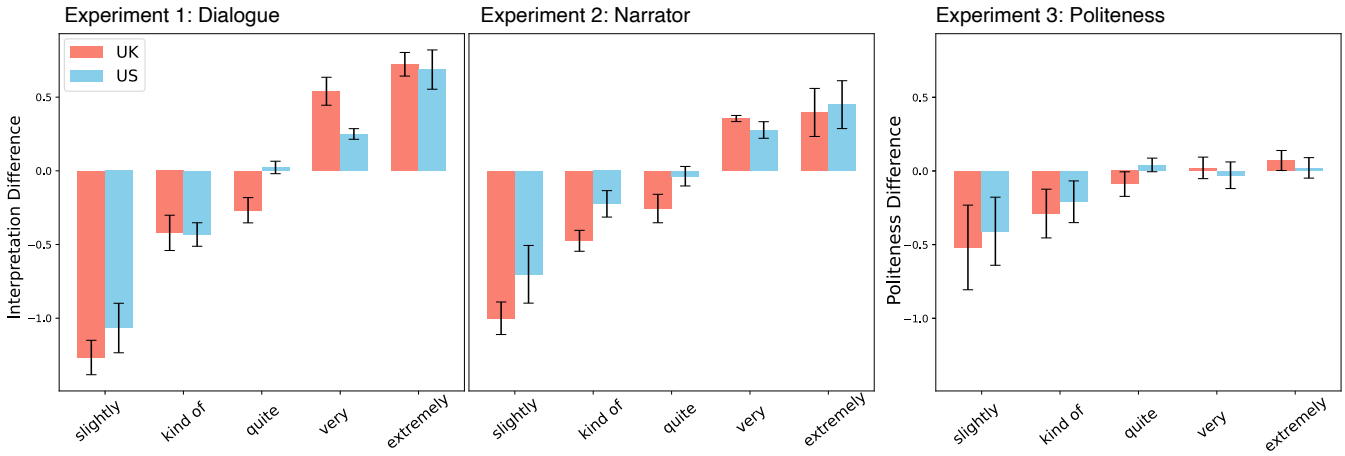

Figure 1: Results from Experiment 1, 2 and 3 reveal robust interpretation differences between UK and US participants

sational partner, there is no overt reason to hedge, intensify, or soften statements for politeness. Thus, if the differences in modifier interpretation we observed between US and UK participants in Experiment 1 persist in a non-communicative setting, it would suggest that these differences may be more due to differences in the literal meaning of intensifiers. However,if the differences disappear or significantly reduce, this would suggest that pragmatic considerations –such as politeness norms– play a more significant role in shaping cross-cultural differences.

Results We recruited 39 participants, 19 from the UK and 20 from the US; aside from using the modified variants of the stimuli, all other elements of the design were the same. As in Experiment 1, we examined difference scores for each modifier. To quantify how much of the differences in Experiment 1 still persist in the absence of the need of the speaker to be polite, we added per-item averaged responses from Experiment 2 as an additional predictor in the same linear mixed-effects model we used for Experiment 1. Including the effect of framing (dialogue vs. narrator) significantly improved model fit in a likelihood ratio test ($\chi^2(1) = 7.31$, $p = .007$), indicating some effect of the narrator manipulation. Critically, we still observed persistent differences in modifier interpretation between UK and US participants (see Figure 1, middle), suggesting that perceived politeness pressures alone do not fully explain the interpretation differences found in Experiment 1. Notably, while the cross-cultural gap for "very" closed, the modifier "quite" still showed significant variation. In US English, "quite" functioned (if anything) as an intensifier, while in UK English, it functioned as a downtoner.

Our results also suggested that participants consider pragmatic factors: if the only difference between the narrator and dialogue scenario were politeness, we would expect *stronger* interpretations of modifiers in the narrator condition, since modifiers should now purely convey information without ambiguity about social constraints. Instead, we found weak statistical evidence (p=0.13) suggesting the opposite effect, that modifiers actually influenced interpretation less in the narrator condition, meaning that listeners inferred additional meaning from overt speech compared to inner monologue. One possible explanation is utterance cost: in a communicative setting, the mere act of choosing to modify an utterance may signal additional meaning beyond the literal semantics of the intensifier. In contrast, in a (non-communicative) narrator setting where there is no overt utterance production, such cost-based inferences may not apply, leading to weaker modifier interpretations. Importantly, the extent of this reduction varied across cultures, suggesting that the perceived "cost" of an utterance, and the inferences listeners draw from modification may be culturally specific.

In summary, Experiment 2 provides indirect evidence that at least some of the observed variation is due to different literal meanings between the UK and the US. However, we could not fully rule out pragmatic influences at play in the narrator condition itself. As we have observed with the reduction of modifier effect in the narrator condition, multiple pragmatic factors changed between the narrator and dialogue condition. In addition, participants may have implicitly interpreted the narrated inner monologue as a kind of speech act, introducing residual pragmatic considerations into their judgments. To address this, Experiment 3 directly measured perceived politeness for each predicate-modifier combination, allowing us to explicitly quantify the role of politeness in modifier interpretation.

Experiment 3: Politeness Version

To explicitly quantify the role of politeness norms, Experiment 3 directly measured perceived politeness for the same predicate-modifier combinations used in Experiment 1. Instead of asking participants to interpret the meaning, we asked them to rate the politeness of each utterance. We recruited 40 participants from the UK and US, ensuring that each participant saw the same set of scenarios as in previous experiments.

Results Our results partially support our hypothesis, that intensifying a negative predicate is perceived as impolite and intensifying a positive predicate is polite, whereas downtoning has the opposite effect. Interestingly the modifiers which exhibited the highest sensitivity to valence were 'kind of' and 'slightly', exactly the same as in Experiment 1. For both of these modifiers, modifying (downtoning) positive predicates were perceived as impolite ($p<0.01$). The other modifier-valence pairs did not have a significant effect on politeness.

In order to compare main effects of the modifiers, we flipped the sign of the politeness difference according to the valence (see 2 right for the mean and standard error across predicates). Intuitively, the more negative this value is, the more polite they are in negative contexts and impolite in positive contexts, hence exhibiting more downtoner like characteristics according to our hypothesis. We found significant variation between modifiers, and the order generally followed this predicted trend, supprting the view that the change in degree and the politeness interpretation are highly related (the relations "slightly" $<$ "kind of" $<$ "quite" were both statistically significant with $p< 0.01$, while comparisons between "quite" ,"very" , and "extremely" did not reach statistical significance).

To test whether cross-cultural differences in modifier interpretation are driven by politeness-based pragmatic inferences, we fit a linear mixed-effects model including perceived politeness judgments as a predictor. Politeness was included as an per-item averaged additional measure as predictor into the linear model. We found that incorporating perceived politeness as a predictor significantly improved the fit of the dialogue data ($\chi^2(1) = 27.90,\ p < 0.001$), suggesting that politeness perceptions capture relevant contextual information which affect modifier interpretation. Furthermore, while politeness ratings were highly correlated across UK and US participants, we observed systematic differences between the two groups. In particular, modifiers that showed interpretation differences in the narrator condition (e.g., "quite") also showed differences in politeness perception, suggesting that some semantic differences between UK and US English contribute to the effects of its role as modifiers.

To test whether difference in politeness perception explains cross-cultural differences in modifier interpretations, we fit a model which included modifier-predicate interactions (and hence encoded valence and broader contextual information), and investigated whether adding the country dependent per-item averaged politeness perception increased fit to the modifier interpretations from Experiment . We found that the improvement was not significant ($p=0.60$), suggesting that cross-cultural differences in politeness perception does not linearly explain interpretation differences. Taken together, while these findings provide strong evidence that politeness norms systematically influence modifier interpretation, the difference in politeness norms as measured in Experiment did not directly translate into differences in pragmatic interpretation of modifiers across cultures.

Computational Modeling

Our behavioral experiments established that modifier interpretation varies across cultures. Additionally, we observe that politeness perception and semantics of modifiers also vary, and our statistical analyses showed that contextual factors have a significant effect on modifier interpretations. While the statistical analysis in Experiment 3 shows that politeness differences between the UK and the US do not explain interpretation differences directly, differences in politeness perceptions may be affecting utterance interpretations in more nuanced ways. Listeners may be interpreting polite utterances as uttered for the purpose of being polite rather than informative, discounting informative content to infer the true state. Furthermore, the extent to which listeners discount the informative content may be culture-dependent, which may be causing the difference in interpretation. In order to test this hypothesis, as well as quantitatively investigate how semantic and pragmatic differences contribute to interpretation differences between cultures, we formalize a model of pragmatic interpretation by incorporating cultural differences into a Rational Speech Act (RSA) framework.

Modeling framework Following Yoon et al. (2020) and Lumer and Buschmeier (2022), we model listeners as interpreting an utterance by recursively reasoning about a speaker who balances three competing objectives: *informativity*, *conciseness*, and *kindness*. We formulate this through a utility function that the speaker is maximizing:

$$U_S(w|s,\phi_i,\phi_s) = \phi_i \cdot U_i(w|s) + \phi_s \cdot U_s(w) - C(w)$$

Here we denote by w be the speaker's utterance, and s the true underlying state that the speaker wants to communicate. U_i, U_s and $C(w)$ represent the speaker's utility to be informative, kind and concise, respectively, which we formalize mathematically in the following way.

We define the informativity, $U_i(w|s)$ to be the log-likelihood that a literal listener L_0 correctly interprets the true state s which the listener wants to convey: $U_i(w|s) = \ln(P_{L_0}(s|w))$, where the literal listener's interpretation can be expressed as $P_{L_0}(s|w) \propto [\![w]\!](s) \cdot P(s)$. The semantic denotation $[\![w]\!](s)$ is 1 if the literal meaning of the utterance w is consistent with the state s, and $\varepsilon << 1$ when they are inconsistent. We define the social utility, $U_s(w)$ as how kind the speaker is perceived as being. Finally, we define the utterance cost, $C(w)$ as the cost required to produce an utterance, penalizing longer utterances. This cost captures our intuition that as listeners, we infer that speakers wanted to convey some extra meaning by adding a modifier to the utterance. The higher the cost of the utterance, the more the listener infers that the addition of the utterance was "worth it".

Using this utility, we model the speaker's utterance probability as the softmax of the combined utility marginalized over each state, i.e.

$$P_{S_1}(w|s,\phi) \propto \exp[U_{S_1}(w|s,\phi)]$$

where $\phi = (\phi_s, \phi_i)$ expresses how much the speaker values each of social utility, ϕ_s, and informational utility, ϕ_i. The pragmatic listener's interpretation can then be expressed as

$$P_{L_1}(s|w,\phi) \propto P_{S_1}(w|s,\phi) \cdot P(s).$$

Culturally dependent parameters Equipped with this modelling framework, we investigate two related hypotheses about how pragmatic factors may affect modifier interpretations. First, we hypothesize that differences in politeness perception across cultures explain differences in interpretation. Second, we hypothesize that differences in the way cultures trade off pressures to be polite, informative and concise lead to differences in interpretation. If the first hypothesis holds, politeness perception will affect modifier interpretation, hence the best fitting model would have a non-zero coefficient on social utility ϕ_s. If the second hypothesis is true, we would expect to see different values for ϕ_s, ϕ_i and $C(w)$ between cultures. To understand how cross-cultural differences in semantic and pragmatic differences affect modifier interpretation, we thus consider different models with different combinations of culturally-dependent and culturally-shared parameters, and compare how well they capture our behavioral data.

Our model has a total of fifteen parameters. Twelve of these parameters control the literal semantics of each modifier. The literal meaning, which corresponds to the denotation function $[\![w]\!](s)$ in the model, is modeled with a (smooth) double threshold, in which the function evaluates to 1 between the two thresholds, and 0 outside that interval. Therefore, the literal meaning of each modifier is described with two parameters (the lower and upper threshold), giving a total of 5 $\times$2 = 10 parameters to describe all five modifiers. We further have two additional thresholds which captures the baseline range of values which are acceptable without any modifier, bringing us to a total of 10+2=12 parameters. The remaining three parameters capture possible differences in pragmatic factors: the utterance cost, $C(w)$, and the weight on social and informational utility $\phi = (\phi_s, \phi_i)$. We allow different combinations of these parameters to differ or to be the same between cultures (see Table 2). We compare model fit by summing the log loss $\log P_{L_1}(s|w)$ across all participants and scenarios. By comparing a model which allows for variation with a model which does not, we investigate whether modeling these parameters as culturally specific better explains the data.

Model results

General interpretation of the model comparisons First, in support of differences in literal meaning, we found that allowing the literal meaning of modifiers to vary between the US and the UK improved the fit significantly (M8 in Table 2). In agreement with our earlier statistical analyses, which found that 'quite' and 'very' were more different across cultures than other modifiers, we found that a model only allowing the literal meaning of "quite" to vary cross-culturally (M6 in Table 2) accounts for the largest improvement (allowing thresholds of other modifiers to differ across cultures did not improve log loss significantly). In fact, the model which only allows the 'quite' threshold to be culturally specific had the best fit in terms of BIC. This agrees with the observations from our statistical analysis, in which we found that 'quite' and 'very' were more different across cultures than other modifiers.

Table 2: Model Comparisons

	Culturally Diff. Params	df	Log Loss	AIC	BIC
M1	none	15	11250	22530	22622
M2	soc	16	11250	22532	22629
M3	inf	16	11244	22520	22618
M4	cost	16	11248	22529	22627
M5	cost + inf	17	11236	22507	22611
M6	'quite' only	17	11226	22485	**22590**
M7	'very' only	17	11227	22489	22593
M8	thresholds	27	11219	22491	22656
M9	all	30	**11194**	**22448**	22631

Furthermore, we also see that there is still a big gap in log loss between the model accounting for all semantic meaning (M8) and the most parsimonious model which also allows pragmatic factors to vary (M9). This suggests that while the remaining three pragmatic factors each do not explain much of the variance individually (M2, M3, and M4 all have high log loss), considering these factors in conjunction with the cultural difference in literal semantic increases fit significantly. This suggests that cultural differences arise from a combination of differences in pragmatic and literal meaning interpretation. Furthermore, as M5 suggests, combining considerations for cultural differences in utterance cost, and utility of informativity increases model fit significantly compared to only allowing each one of the two parameters to vary independently. This suggests that cultural differences in interpretation cannot be understood in isolation of different pragmatic factors but needs a nuanced interpretation considering the combined effect of these factors.

In addition to the model comparisons, in order to understand the significance of modeling pragmatic factors, ϕ_{soc}, ϕ_{inf}, and $C(w)$, as culturally dependent, we minimally modified the best fitting models for the UK and the US so that they share the same relevant parameters. Forcing utterance cost, $C(w)$, or the coefficient of informativity, ϕ_{inf} to be the same across models (while keeping the other parameters the same as their respective best fitting models) increased the log loss significantly from 1994 to 11222 and 11234 respectively, further supporting the interpretation that pragmatic dif-

ferences combine with semantic differences to play a significant role in shaping utterance interpretation. Analyzing the

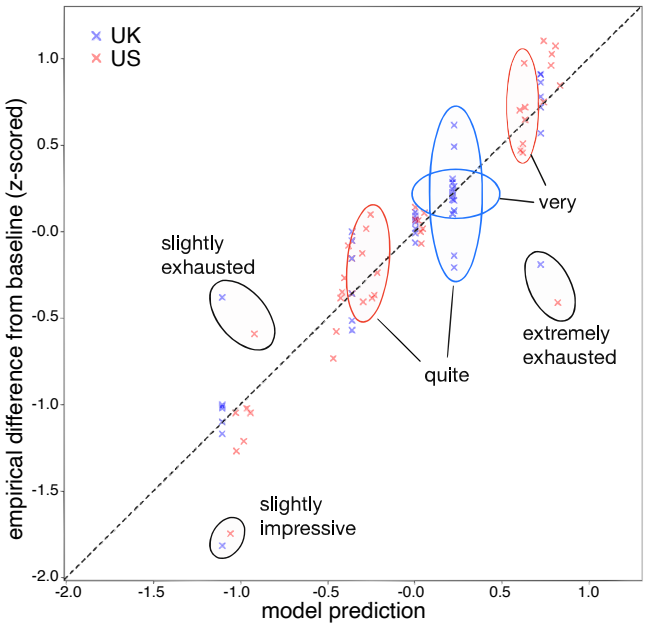

Figure 2: Comparison between model predictions and behavioral response. The predictions generated are from the model parameters specifically optimized to fit each country's data separately. Each cross represents a modifier-predicate pair.

models in more depth, in Figure 2 we compare the best fitting model for each of the UK and US with behavioral data. Consistent with our model comparisons, the U_{soc} term has little effect in fitting the model, meaning that regardless of the predicate, model predictions are roughly the same within each modifier. This lack of sensitivity to predicate variation means that modifier-predicate pairs that deviate from the average modifier values are significantly off-diagonal. These include modifier-predicate pairs such as 'extremely exhausted', whose value is much lower than the average interpretation of the modifier 'extremely' across predicates. These off-diagonal points including 'extremely exhausted' are mostly infrequent modifier-predicate pairs, which suggests that the infrequent nature may be the reason for the deviation from the rest of the data. More specifically, a listener may interpret the infrequent combination as a way to convey additional meaning, such as exaggeration or sarcasm. For example, the listener may interpret the utterance 'extremely exhausted' as sarcastic which will also lead them to discount its informativity, as it is unlikely to modify the predicate 'exhausted' which by itself already indicates high degree of tiredness and is typically considered as non-gradable.

Relevance of the social utility term Social utility does not have a significant effect in any of these models. If we ignore the social utility component (by setting the coefficient to be 0), this only increases the logloss by at most two points, indicating that the additional complexity introduced by adding the politeness component in our model is not justified. This seemingly conflicts with findings from statistical analyses from our behavioral data, which suggest that politeness perceptions capture relevant contextual information for modifier interpretation. This suggests that while politeness perceptions that we measured were relevant contextual information in predicting modifier interpretation (as shown by our statistical analyses), it did not affect interpretation in the way as expected by our computational model. As we identified from analyzing Figure 2, this is likely due to the fact that the 'politeness perception' measured from Experiment 3 encodes other social factors such as sarcasm and exaggeration, hence does not affect interpretation in the way it is expected by our model which only captures the tradeoff between informativity and politeness, and ignores the possibility of other factors.

Discussion

This study proposed a novel way to investigate how pragmatic communication may vary between cultures. Through behavioral experiments, we found systematic differences in modifier interpretations across the UK and the US and found evidence that these differences were due to both semantic and pragmatic differences between the two cultures. Through computational analysis, we found further evidence that differences in pragmatic and semantic factors in combination explain differences in interpretation across cultures. The fact that a combination of factors such as informativity and utterance cost affect interpretation differently across cultures points to how difficult it might be to interpret an utterance of someone from a different culture, reiterating the difficulty of cross-cultural communication.

These results offer only a partial view of how politeness affects interpretation across cultures. Significant variation exists within the UK and US due to diverse speech communities, which could be better accounted for thorough a hierarchical approach, modeling within-cultural as well as cross-cultural variation. Furthermore, our stimuli included a range of different scenarios and contexts which made it difficult to capture the pure effect of the difference in politeness perception on interpretation. Analyzing data on a larger scale with more targeted experiments may help better understand how politeness of modifiers differ across cultures.

Zooming out, our findings that different cultures have different conversational norms on politeness pose new questions about cross cultural research more broadly. How does difference in interpretation of modifiers extend to broader cross cultural differences in expectations of communication? How do people navigate these differences when communicating cross culturally?

All code and materials available at:
https://github.com/yukam997/PolitenessAcrossCultures

Acknowledgments

The authors thank the anonymous reviewers for their helpful feedback, as well as members of the Computational Cognitive Science Lab at MIT, and members of the Social Interaction Lab at Stanford for valuable discussions. Yuka Machino was supported by the Ezoe Recruit Memorial Foundation.

References

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge university press.

Goddard, C. (2012). 'Early interactions' in Australian English, American English, and English English: Cultural differences and cultural scripts. *Journal of Pragmatics*, *44*, 1038-1050.

Goddard, C., & Wierzbicka, A. (2004). Cultural scripts: What are they and what are they good for? *Intercultural Pragmatics*, *1*(2), 153–166.

Goddard, C., & Wierzbicka, A. (2013). *Words and meanings: Lexical semantics across domains, languages, and cultures*. OUP Oxford.

Grice, H. P. (1975). Logic and conversation. *Syntax and Semantics*, *3*, 41-58.

Haugh, M., & Bousfield, D. (2012). Mock impoliteness, jocular mockery and jocular abuse in Australian and British English. *Journal of Pragmatics*, *44*(9), 1099-1114.

Haugh, M., & Schneider, K. P. (2012). Im/politeness across englishes. *Journal of Pragmatics*, *44*(9), 1017-1021.

Lumer, E., & Buschmeier, H. (2022). Modeling social influences on indirectness in a rational speech act approach to politeness. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Romero, S. (2012). *This is so cool! a comparative corpus study on intensifiers in british and american english.* (Unpublished manuscript, University of Tampere)

Ruzaitė, J. (2007). Vague references to quantities as a face-saving strategy in teacher-student interaction. *Lodz Papers in Pragmatics*, *3*(-1), 157–178.

Schneider, K. P. (2012). Appropriate behaviour across varieties of english. *Journal of Pragmatics*, *44*(9), 1022-1037.

Stratton, J. M. (2021). 'That's proper cool': The emerging intensifier proper in British English. *English Today*, *37*(4), 206–213.

Su, Y. (2016). Corpus-based comparative study of intensifiers: quite, pretty, rather and fairly. *Journal of World Languages*, *3*, 224-236.

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied linguistics*, *4*(2), 91–112.

Waters, S. (2012). "it's rude to vp": The cultural semantics of rudeness. *Journal of Pragmatics*, *44*(9), 1051-1062. (Im/politeness across Englishes) doi: https://doi.org/10.1016/j.pragma.2012.02.002

White, I., Pandey, S., & Pan, M. (2024). Communicate to play: Pragmatic reasoning for efficient cross-cultural communication. In *Findings of the association for computational linguistics: Emnlp 2024* (pp. 12201–12216).

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2020). Polite speech emerges from competing social goals. *Open Mind : Discoveries in Cognitive Science*, *4*, 71 - 87.