Semantics & Pragmatics Volume 17, Article 10: 1–34, 2024
https://doi.org/10.3765/sp.17.10

Social stereotypes affect imprecision resolution across different tasks*

Andrea Beltrama
University of Pennsylvania

Florian Schwarz
University of Pennsylvania

Submitted 2024-01-08 / First decision 2024-04-10 / Revision received 2024-05-24 / Accepted 2024-07-14 / Published 2024-07-26 / Final typesetting 2025-02-05

Abstract In two experiments, we investigate how social information about the speaker affects pragmatic reasoning in numeral interpretation. Results from a picture selection task show that comprehenders interpret numerals more precisely when uttered by Nerdy speakers — described as studious, introverted and uptight — as opposed to Chill ones — described as extroverted, sociable, and laid-back (Experiment 1). Data from a Truth-Value Judgment task (Experiment 2) confirm this pattern: comprehenders exhibit more tolerance towards accepting imprecise statements uttered by Chill speakers than Nerdy ones. These findings highlight the importance of incorporating the interplay of social and descriptive meaning into our understanding of pragmatic reasoning and outline several directions of inquiry to take this enterprise further.

Keywords: Social meaning, imprecision, stereotypes, numerals, experimental pragmatics, truth-value judgment task

* We thank editor Elin McCready and the anonymous referees who reviewed the manuscript for providing constructive and detailed comments. We are likewise grateful to Alexandros Kalomoiros, Chris Kennedy, Anna Papafragou, Rista Plate, Hannah Read, Gareth Roberts, Meredith Tamminga, Lisa Titus, and Lacey Wade for sharing extensive feedback on prior versions of this manuscript. We are likewise grateful to: members of the Language & Cognition Lab; Meaning Lab; Variation & Cognition Lab; and the MIRA group at the University of Pennsylvania; to audiences from SALT 31 & 32; NWAV 49 and 50; ELM 1 & 2; HSP 1; the 'Gradability and measurement across domains' workshop at Rutgers University; and to colleagues and friends from the University of Delaware, the University of Dusseldorf, Tel Aviv University and the University of Manchester who kindly invited us to present this work and shared helpful comments and suggestions. This material is based upon work supported by the National Science Foundation under Grant BCS-2140765. All errors are our own.

1 Introduction

Besides allowing us to describe reality, linguistic expressions carry a *socio-indexical* meaning, which conveys information about speakers' social identity and personality (Labov 1966, Silverstein 2003, Agha 2003, Eckert 2008, Podesva 2011, Hall-Lew, Moore & Podesva 2021). In some circumstances, this content can be conceived of as a cluster of attributes: for example, in North America, alveolar realizations of (ING) — e.g., *goin' fishin'* — typically evoke, among others, traits such as "working class", "Southerner", "friendly", "folksy" (Campbell-Kibler 2011). In other cases, social meanings come to represent complex, stereotypical *types* of speakers — or *personae*. For example, in the context of US high schools, clusters of morpho-syntactic (e.g., negative concord) and phonological features (e.g., vowel backing/raising) conspire to invoke a "Burnout" persona, opposed to a "Jock" one (Eckert 2008); similar associations have been discussed in relation to stereotypical figures such as "Valley Girls" (D'Onofrio 2015), "Nerds" (Bucholtz 2001), or "Frat Boys" (Kiesling 2018; see D'Onofrio 2020 for an overview). Furthermore, it's been shown that the link between social information and linguistic behavior is often bi-directional: comprehenders not only infer social information from the linguistic signal; they also adjust their perception of the linguistic signal depending on what information is available (Niedzielski 1999, Hay & Drager 2010, Babel 2012, Drager 2015, D'Onofrio 2015, 2018, 2019). Notably, a similarly tight connection between social information and human behavior is also observed in other domains besides language, including decision making (Fiske 2018, Jenkins et al. 2018); social categorization: (Stolier, Hehman & Freeman 2020, Kinzler 2021); and credibility ascription (Fricker 2007, Davis 2016). As a result, the study of the relation between social information and behavior continues to draw considerable attention — both in linguistics and the cognitive science more broadly (see Wade 2020, D'Onofrio 2020, Kinzler 2021 for overviews).

In contrast with these works, social meaning has received less attention in semantics and pragmatics. Much work in this area explored — and modeled — how comprehenders engage in inferential work to interpret linguistic expressions (Grice 1975, Horn 1984, Gazdar 1980, Levinson 2000, Roberts 1996 i.a.). Yet, efforts to formalize this reasoning, especially in Gricean and neo-Gricean frameworks, have not been particularly concerned with capturing the role of sociolinguistic considerations in the process. Most work in this tradition, in fact, has effectively endorsed a view of conversation on

which conversational partners are treated as idealized, socially undifferentiated rational agents — an outlook that relegated social distinctions between speakers as tangential, at best, to the study of pragmatic reasoning.

Crucially, the backgrounding of sociolinguistic matters in this area isn't so much grounded in the substance of pragmatic theories, as it is in the particular epistemological and methodological approach adopted by scholars working in this tradition. In fact, Grice himself observes that "there are all sorts of other maxims (aesthetic, social or moral in character) [...] that could also generate inferences" (Grice 1975: 47). All in all, these premises fully align with the idea, central to sociolinguistic theory, that speakers differ — and are expected to differ — on how they like to use language, just as they differ in their epistemic and emotive state. Accordingly, recent work in pragmatics has placed social considerations front and center, showing that comprehenders infer social meaning(s) from the semantic and pragmatic properties of utterances (Acton & Potts 2014, Beltrama & Staum Casasanto 2021, Acton 2019, Glass 2015, Jeong 2021); and that semantic interpretation is shaped by social factors such as politeness (Bonnefon, Feeney & Villejoubert 2009, Yoon et al. 2020, Mazzarella et al. 2018); affect (Kao et al. 2014, Bergen 2016); and speaker-specific information like linguistic nativeness (Fairchild & Papafragou 2018) or political orientation (Henderson & McCready 2019, Mahler 2022; see Beltrama 2020 for an overview).

Despite these advances, however, the investigation of the socio-indexical vs. semantic/descriptive dimensions of meaning has continued to proceed on parallel tracks, leaving one questions largely unexplored: how do comprehenders recruit the wealth of socio-indexical information typically available in a conversation to compute a speaker's intended message?

In this paper, we address this question by asking how comprehenders utilize stereotypical social information in two tasks involved in interpretation: inferring the intended extension of an expression; and adjudicating a description's suitability to represents a given fact. As our testbed, we focus on *(im)precision* with numerals — a phenomenon that presents itself as especially likely to show a principled effect of socio-indexical information on pragmatic reasoning (see Section 2). We thus see this case as one that maximizes our chances of observing the effect of interest taking a meaningful first step towards filling the gap — with an understanding that the question of interest will have to be tested against a broader range of phenomena (See Section 7).

2 Imprecision and numerals: A case study

Speakers often use quantity expressions in a way that isn't fully adherent to the relevant facts. For example, someone could reasonably utter (1) when the time is 6:03; or (2) when the price is $207.

(1) It's 6 o'clock.

(2) The ticket costs $200.

This phenomenon, known as *imprecision*, has been extensively investigated in semantics and pragmatics, both formally (Lewis 1979, Pinkal 1995, Lasersohn 1999, Sauerland & Stateva 2011, Solt 2014, Kao et al. 2014, Klecha 2018, Beltrama 2021) and experimentally (Van Der Henst, Carles & Sperber 2002, Syrett, Kennedy & Lidz 2009, Cummins, Sauerland & Solt 2012, Leffel, Xiang & Kennedy 2016, Aparicio 2017). Two distinctive properties of (im)precision highlight this phenomenon as one in which socio-indexical information is likely to affect the interpretation process.

One property is that the precision level with which numerals are used serves as a cue for listeners to infer social meanings. In particular, speakers using sharp vs. round numbers (e.g., $207 vs. $200) — respectively signaling a higher vs. lower precision (Krifka 2009) — evoke distinct clusters of social qualities (Beltrama 2018, Beltrama, Solt & Burnett 2022), which aligns with the two fundamental social dimensions of *Competence* and *Warmth* that models of social cognition — and particularly the *Stereotype Content Model* (Fiske et al. 1999, Fiske, Cuddy & Glick 2007, Fiske 2018) — see as central to person perception. Specifically, precise speakers are ascribed traits that foreground *enhanced intellectual status*, or *Competence*, and *reduced sociability*, or *Warmth*: for instance, they are perceived as articulate, educated, and intelligent; but also pedantic and obsessive. Less precise speakers are instead associated with qualities that, conversely, foreground *Warmth* over *Competence*: for instance, they are perceived as friendly, extroverted and laid-back, though less intelligent and educated than precise speakers. Notably, similar indexical links have been highlighted in connection to other linguistic manifestations of detail-orientedness, and particularly *hyper-articulation* (Bucholtz 1999, 2001) — a style characterized by phonetic forms displaying a high degree of detail along one or more dimensions (see Lindblom 1990). Specifically, hyper-articulated speech has been claimed to index high-Competence qualities parallel to those evoked by precise numerals (*articulateness*, Podesva et al. 2015; *learnedness*, Benor 2004; *effortfulness*, Eckert 2008), as well as stereotypi-

cal personae that centrally epitomize these traits (e.g., *Nerds*; see Bucholtz 1999, 2001). Similarly, hypo-articulation has been linked to stereotypical personae such as "Surfer dudes", "Skaters", and "Frat Boys", which incarnate high-Warmth traits — e.g., laid-backness, chillness, sociability (Kiesling 2018) reminiscent of those indexed by imprecise numerals.

The second relevant property is that the possibility of speaking imprecisely requires comprehenders to perform inferential work in interpreting numerals, as can be observed in two types of communicative situations. First, when a comprehender is inferring what facts are being described by an utterance, they have to determine what range of values the numeral can be taken to represent: the sentence "The ticket costs $200", for instance, can reasonably be taken as describing a value that falls within an interval of varying range — e.g., exactly $200; $195-$205; $190-$210; a comprehender has to reason about the conversational setting — and ultimately settle on a level of precision — to determine what range is contextually appropriate. But comprehenders may also have to engage in inferential work in another type of situation: one in which they have access to the relevant facts, and have to evaluate whether a particular numerical description can be treated as adequately representing them. Here, the reasoning crucially involves the comprehender calibrating how much *leniency* they are willing to extend toward a speaker describing facts imprecisely. For example, given a known price of $207, a comprehender may need to adjudicate whether a statement like "the price is $200" is a close enough representation of this price in the context, or should instead be rejected as a mischaracterization.

In out studies, we explore the role of social information in these processes by comparing comprehenders' interpretation of numerals produced by a *Nerdy* vs. a *Chill* speaker — two stereotypes embodying the distinctive social traits linked to high vs. low precision respectively.

3 Speaker properties and imprecision resolution: Hypotheses

Concerning the first domain of interpretation — i.e., the determination of the contextual extension of the numeral — we predict that numerals uttered by someone who features the typical high Competence social profile of a precise speaker should be interpreted more strictly than numerals uttered by someone who features the High Warmth profile of a less precise speaker. We call this Hypothesis 1. This prediction is grounded in the finding that the relationship between social information and language understanding is typically

bi-directional (see Section 1): if comprehenders consistently associate these clusters of social traits with speakers using different levels of precision, they should thus conversely rely on these clusters to adjust the precision level in the context. Note that, while intuitive, this prediction is not a trivial one. First, it's been shown that, in certain cases, the link between social information and language understanding is merely *uni*--directional; and when this is the case, linguistic information is more constraining of sociolinguistic perception perception than vice versa (see Squires 2013 on agreement mismatches). As a result, it is not guaranteed that, just because (im)precision informs social perception, the link between the two necessarily works both ways. Second, even if a difference in interpretation between speakers embodying these identities were found, the question would still remain of whether the effect is equally driven by Competence- vs. Warmth-based clusters of traits; or whether it is only one of these two stereotypes that comprehenders consider when computing the extension of the numeral.

Concerning the second type of situation — i.e., the adjudication of imprecise statements *vis-à-vis* a known fact — there are two plausible outcomes one could imagine. One is that, similar to what predicted by Hypothesis 1, imprecise descriptions produced by a speaker embodying the social traits linked to high precision should be rejected *more* often, by virtue of being associated with a narrower range of values. We call this Hypothesis 2A.

An alternative possibility is that comprehenders might determine whether to accept or reject an imprecise statement by recruiting social information to assess the epistemic authority that they are willing to extend to the speaker. This possibility can be grounded in the idea that meaning interpretation doesn't merely require the computation of the intended extension of the expressions, but is also shaped by how credible a comprehender takes the speaker to be (see in particular Sperber et al. 2010, Breheny, Ferguson & Katsos 2013 on the Principle of Epistemic Vigilance). Crucially, it's been argued that credibility assessments of this sort are often affected by the speaker's social identity. For example, work on *testimonial injustice* in philosophy (Fricker 2007) has highlighted the role of social stereotypes as a driving force behind the ascription of epistemic authority to an agent, often in a discriminatory fashion — e.g., when a woman is seen as less credible than a man on a given topic (see McCready & Winterstein 2017; Section 7 for further discussion). Accordingly, we consider an alternative hypothesis (Hypothesis 2B): comprehenders could see an agent with the social profile of a precise speaker as especially invested in descriptive accuracy, and thus more deserving to

being trusted in the presence of an imprecise description. This should lead to a *lower* rate of rejection of imprecise descriptions for people embodying the social profile of a precise speaker — generating the opposite pattern of results to that predicted under Hypothesis 2A.

4 Experiment 1: Inferring facts from statements

In Experiment 1 we test Hypothesis 1 utilizing a COVERED SCREEN task (see Huang, Spelke & Snedeker 2013, Schwarz, Romoli & Bill 2016 for predecessors).

4.1 Methods & design

Our stimuli consisted of written dialogues between two characters, Arthur and Rachel. In each dialogue, one character asked a question, and the other respond after looking at their phone, uttering a quantity expression in the form of a round number (e.g., "It's $200."). After seeing the dialogue, participants were asked to determine which phone the speaker was looking at by selecting one of two images of a phone. In one image, the phone was turned face down, making the content of the screen invisible (COVERED screen). In the other, the phone' display was fully visible (VISIBLE screen).

Two factors were manipulated. The "Screen Fit" manipulation involved the relation of the number displayed on the visible phone to the one in the utterance, varied across three levels. Two levels were controls: *Match*, with displayed and uttered numbers identical; *Mismatch*, with the two largely divergent. The third level, *Imprecise*, was our critical one: it featured only a slight divergence between the uttered and the displayed number, such that, depending on participants' reasoning, the display might or might not be seen as having prompted the utterance (see below for details on how this margin was determined). The Screen Fit manipulation is illustrated in Figure 1.

The "Persona" manipulation varied the social identity of the interlocutors. To implement this manipulation, we draw on D'Onofrio (2018)'s work investigating the effect of persona-based information on phonological categorization: in this paradigm, participants were provided social information textually, via a sentence stating that a given speaker "had been described as X", where X is a succinct label — typically, the term for a stereotypical persona — for a persona that embodies to evoke a specific cluster of social

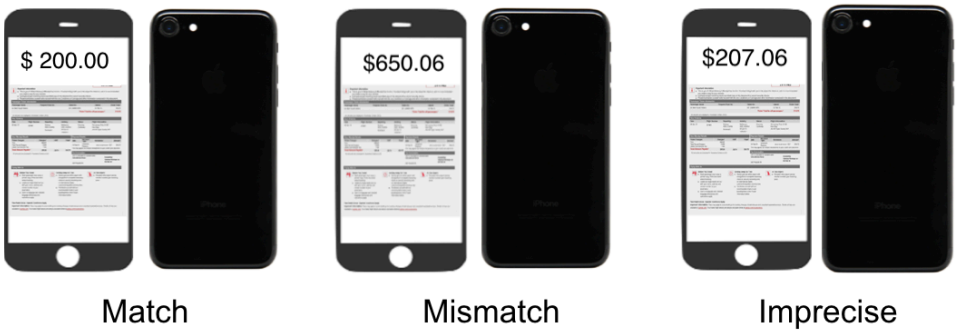

Figure 1 Screen Fit Manipulation. Utterance: "The price is $200"

traits of interest to a particular phenomenon and research question.[1] In one condition, the interlocutors were described as "Nerdy", a persona that aptly foregrounds the high Competence qualities associated with high precision; in another condition they were described as "Chill", a persona that instead incarnates the high Warmth qualities associated with lower precision. In each of these conditions participants were exposed to a trial at the beginning of the experiment, which made explicit the key social qualities that were intended to be associated with each persona. Respondents were told to consider the term "Nerdy" to describe someone who is very "studious, articulate, introverted, and uptight"; and "Chill" to describe someone who is very "laid-back, sociable, extroverted, and care-free". These clusters of traits were selected to provide as exhaustive a cover as possible of the range of associations linked to high vs. low precision, based on prior work focused on the social perception of detail-orientedness with respect to either numerals or phonetic detail (see Section 2).[2] Finally, in a third condition, no information about the speaker was provided. This condition served baseline for assessing the effect on imprecision resolution of the Nerdy and Chill personae independent of one another (see below).

1 In D'Onofrio's study, for instance, the labels of choice were "Valley Girl", "Business Professional", and "Chicago Bear Fan".

2 A question raised by this approach revolves around how Nerdy vs. Chill personae relate to each other — above and beyond the fact that they are distinct. We submit that these two stereotypes are best thought of as occupying opposite quadrants (High Competence/Low Warmth vs. Low Competence/High Warmth) of the social evaluation space — and thus as in opposition to each other. This idea is crucially supported by the observation that, when participants are asked to express the degree to which they see themselves as embodying these personae, the two turn out to be negatively correlated (see also Footnote 8).

In each trial, participants were instructed to select the VISIBLE screen if they thought that the person giving the answer would have said what they did if they had been looking at the screen corresponding to the visible phone; and the COVERED screen if they didn't believe so. We expected the Match and Mismatch conditions to evoke COVERED responses at floor vs. ceiling levels respectively, without significant variation by speaker persona. By contrast, responses in the Imprecise condition should depend on the (im)precision standards employed by participants on a given trial. In particular, a strict, precise interpretation — one which excludes the value displayed on the visible screen from the numeral's extension — should lead to a COVERED screen response; a liberal, less precise interpretation — one with a lower level of precision, which instead includes the value displayed on the visible screen in the numeral's extension — should yield a VISIBLE screen choice.

Given this setup, the crucial requirement for items in the Imprecise condition was that they feature a margin of deviation from the uttered number that leaves participants somewhat on the fence: this margin should indeed be large enough to make the displayed number clearly different from the uttered one; but close enough to lead respondents to at least consider the possibility that the respondent might have based their answer on that number. Based on our own initial intuitions, as well as pilot results, we implemented an imprecision range deviating between 5 and 18% from the uttered number (relative to the largest relevant count unit in our stimuli; e.g., 100 for dollar amounts). In practical terms, in the Imprecise condition there should be genuine variation between participants as to how a numeral is interpreted strictly vs. liberally; as a consequence, if our approach to implementing the Imprecise condition is correct, we should observe a rate of COVERED responses in this condition in-between that of the Mismatch and Match condition respectively — and ideally hovering around 50%.

4.2 Materials

24 experimental items were created, each varied across 9 different conditions resulting from the 3×3 manipulation of the factors described above: Nerdy, Chill and neutral for the persona manipulation; Match, Mismatch, and Imprecise for the Screen Fit manipulation. The Persona factor was administered between-subjects to avoid raising participants' awareness of this manipulation: respondents would thus see either only dialogues between Nerdy, Chill, or neutral, unspecified interlocutors. Moreover, the interlocutors' genders

were counterbalanced (within-subjects) by alternating dialogues in which the Rachel asked the question and Arthur responded with a numeral; and dialogues in with reversed order.

The Screen Fit manipulation was administered within-subjects: each participant saw 6 items in Match and Mismatch each, and 12 items in Imprecise (with item-condition pairings counterbalanced in a Latin Square Design). 8 items contained utterances describing prices, expressed in dollars (as in Figure 1); 8 describing distances (in miles); and 8 times (in hours:minutes).

The experiment also included 24 filler dialogues between two separate characters — named Becky and Tyler and described as "college students". The fillers were alternated with the experimental items, so that participants would never see two consecutive items of the same type.[3]

4.3 Procedure

The study was implemented and administered online on the PCIbex platform (Schwarz & Zehr 2021).[4] For each item, the context sentence was introduced first on the top left of the screen. This sentence, besides reporting on the interlocutors' goals in the conversation, also provided the key social information from the Persona manipulation. Sequentially, two separate speech bubbles would appear, one containing the question asked by the first character and the other the answer from the other character. Next, the question task-prompt for participants as well as the two pictures of the phone appeared on the right-hand side of the screen. Participants entered their responses by pressing the key matching the letter displayed under the respective phone pictures on the keyboard. The experimental items were preceded by three practice items. Figure 2 provides a full illustration of the display that participants would see once all the elements appeared.[5]

Following the last trial, participants were asked to complete an exit questionnaire by indicating the degree to which they think they would be described as "Nerdy" and "Chill" on a 1(min) to 10(max) scale. Before responding, participants were reminded of the set of qualities that were taken to be distinctive of each persona throughout the experiment, so as to ensure that their self-assessment in the questionnaire would target the same social traits. The questionnaire's aim was to provide a measure to track the de-

3 See Supplementary Materials for full experimental/filler items lists.

4 https://www.pcibex.net

5 Here and in Figure 5, "F" and "J" indicate the keys to press in order to choose each option.

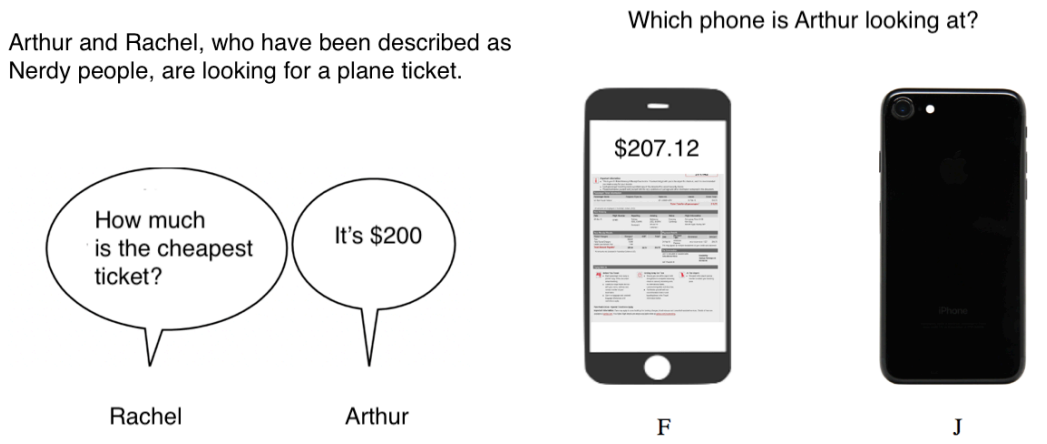

Figure 2 Experiment 1: Display before making the choice (Condition: Nerdy, Imprecise)

gree to which respondents saw themselves as sharing key qualities of the speakers in the experiment — and thus allow us to explore how participants' responses were modulated by their own social identity.

4.4 Participants

282 self-declared native speakers of English (Age median=33; female=166; male=114; other=2) were recruited on Prolific and compensated $2 ($8/hour).

4.5 Results

Figure 3 plots the proportions of COVERED choice rates by Persona and Screen Fit. Two descriptive patterns can be observed. First, there is a step-wise effect of Screen Fit Manipulation, with ceiling and floor-level rates for the controls, and the Imprecise condition in the middle. This validates the viability of our chosen imprecision range. Second, the rate of COVERED choices in the critical Imprecise condition appears to be considerably higher for Nerdy than for Chill speakers, with the baseline condition patterning in-between.

To test this pattern, we first re-coded Screen Fit as a binary factor with "Control" (Match and Mismatch collapsed) and "Imprecise" as levels. We then fit a maximal converging mixed-effects logistic regression on COVERED choice

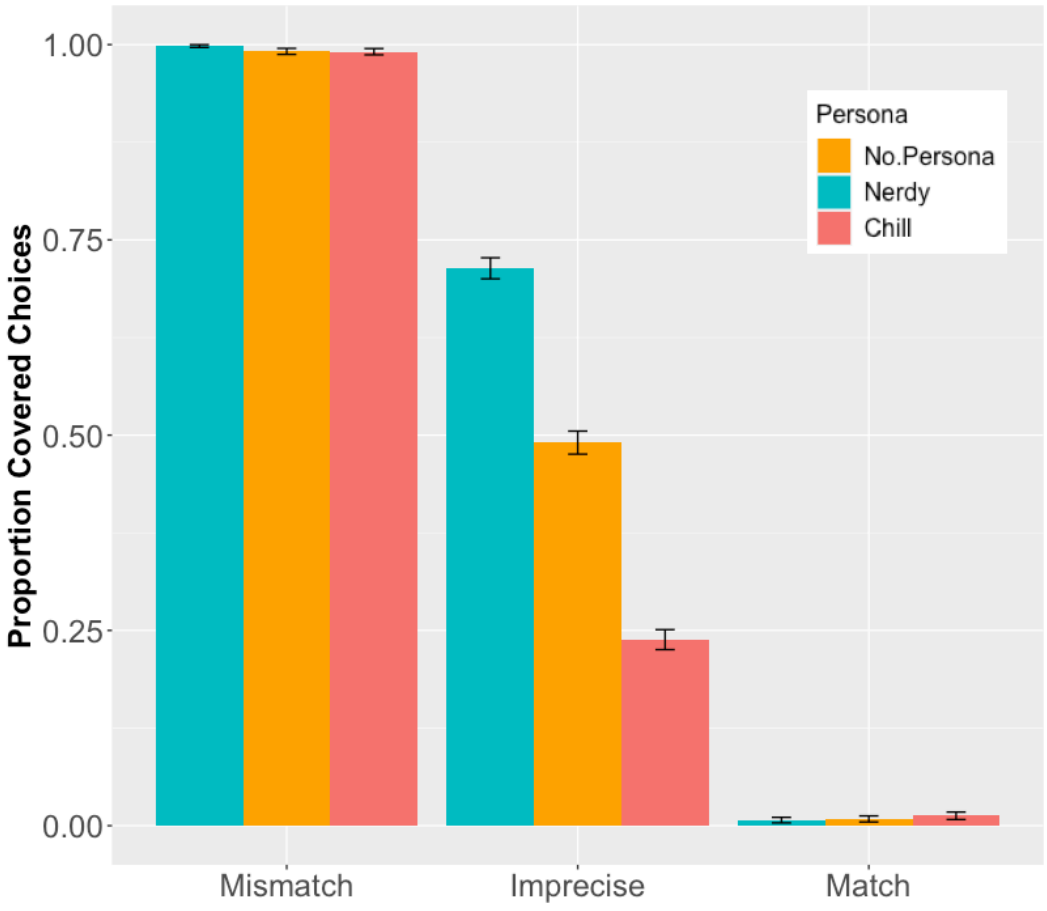

Figure 3 COVERED choices across Screen Fit, split by Persona

rates with Persona and Screen Fit as interacting predictors, random intercepts for participants and items, and random slopes for Screen Fit for participants and items. Persona was treatment coded with "No.Persona" as reference, so as to be able to infer the effect of both Nerdy and Chill with respect to the baseline; Screen Fit was sum coded.

The model (as all other models reported in the manuscript) was fit with the *lmer_alt* function from the *Afex* package in R.[6] The model revealed significant effects of Persona in opposite directions (Chill: $\beta = -0.67$; $SE = 0.13$; $z = 4.86$, $p < 0.0001$; Nerdy: $\beta = 0.77$; $SE = 0.14$; $z = 5.54$, $p < 0.0001$),

6 This makes it possible to excluded random correlations between intercept and slopes by items through the || syntax, and thus minimize convergence and singularity issues.

reflecting the fact that COVERED choice rates were respectively higher and lower than the No.Persona condition for Nerdy and Chill. But both effects are crucially qualified by interactions between Persona and Screen Fit involving both Nerdy ($\beta = -0.78$; $SE = 0.13$; $z = 5.7$, $p < 0.0001$) and Chill ($\beta = 0.66$; $SE = 0.13$; $z = 4.86$, $p < 0.0001$). The interaction reflected the fact that, in the Imprecise condition, COVERED choices rates were *higher* for Nerdy speakers than speakers in the No.Persona condition ($SE = 0.20$; $z = 6.62$, $p < 0.0001$); and *lower* for Chill speakers than speakers in the No.Persona condition ($SE = 0.20$; $z = 7.61$, $p < 0.0001$); no significant Persona contrasts were found in the Control condition (Nerdy vs. No.Persona: $SE = 0.18$; $z = 0.07$, $p = 0.99$; Chill vs. No.Persona: $SE = 0.18$; $z = 0.05$, $p = 0.99$).[7]

Finally, no main effect of Screen Fit was found ($\beta = 0.00$; $SE = 0.10$; $z = 0.09$, $p = 0.92$), reflecting the fact that COVERED screen rates do not differ across the Imprecise and Control conditions.[8]

To gain a more fine-grained perspective on the link between persona and imprecision resolution, we next explored how the Persona effect in the Imprecise condition was modulated by two further social factors: the gender of the speaker (which had been counterbalanced across items); and the self-assessed similarity between participants' social identity and the speaker. The latter was operationalized via a Similarity index — obtained by combining self-ascribed nerdiness ratings for participants seeing the Nerdy speakers and self-ascribed chillness ratings for participants seeing the Chill ones.[9] We fit another maximally complex converging mixed-effects logistic regression on COVERED choices for the Imprecise condition data, with Persona, Speaker Gender (both sum-coded) and Similarity (re-scaled around the mean via the function *scale*() in R) as interacting predictors; random intercepts and random slopes for Speaker Gender for items and participants. The choice patterns predicted by the model are shown in Figure 4.[10]

7 Contrasts were extracted with the *emmeans* package in R, with automatically adjusted α levels for multiple comparisons.

8 The Control proportion resulted from averaging floor- and ceiling-level rates for the two original Control conditions. Note that, in parallel analyses without the control conditions collapsed, the NoPersona condition differs from both Match and Mismatch as clearly suggested by the graph.

9 Post-hoc analyses show that self-ascribed Nerdy and Chill traits were negatively correlated, confirming the assumption that these personae should be thought of as occupying opposite quadrants of the social space. See Supplementary materials for details on the density distribution and correlations relative to these ratings.

10 The graphs were obtained via the function *ggpredict* from the *ggeffects* package in R.

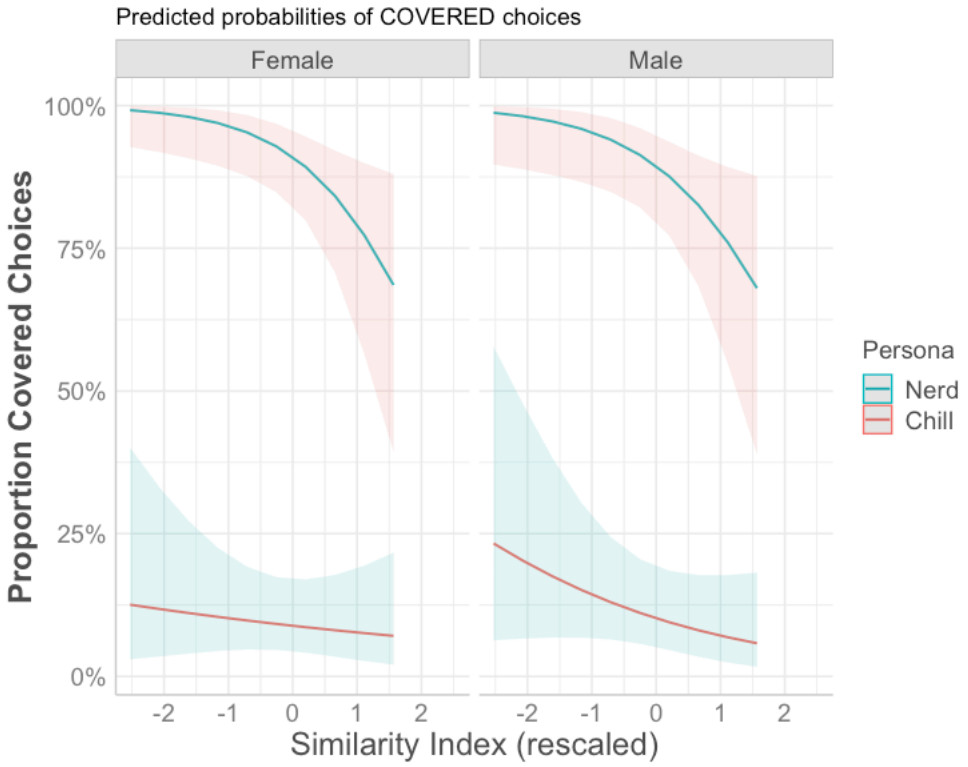

Figure 4 COVERED choices by Similarity and Speaker Gender (Imprecise only)

Consistent with the findings from the first model, we found a main effect of Persona (β=2.23; $SE = 0.24$; $z = 9.12$, $p < 0.0001$), with higher rates of COVERED responses for Nerdy speakers. The model additionally showed a main effect of Similarity, such that COVERED choice rates decreased as Similarity increased (β=0.60; $SE = 0.23$; $z = 2.59$, $p < 0.01$). No main effect of Speaker Gender was found (β=0.01; $SE = 0.27$; $z = 0.05$, $p = 0.97$); and no interaction Persona*Similarity (β=0.65; $SE = 0.45$; $z = 1.43$, $p = 0.15$), Persona*Speaker Gender (β=0.33; $SE = 0.28$; $z = 0.18$, $p = 0.23$), Speaker Gender*Similarity (β=0.06; $SE = 0.14$; $z = 0.46$, $p = 0.64$), and Persona*Speaker Gender*Similarity (β=0.33; $SE = 0.29$; $z = 1.13$, $p = 0.25$) were found. To explore whether the Similarity effect obtained with both Nerdy vs. Chill speakers, we re-ran the model with Persona treatment-coded with Nerdy and Chill respectively set as reference. The model revealed a Similarity effect when the

speaker was Nerdy (β=0.92; $SE = 0.36$; $z = 2.56$, $p = 0.01$) but not with Chill (β=0.27; $SE = 0.28$; $z = 0.95$, $p = 0.33$).

4.6 Discussion

The results from Experiment 1 show that numerals uttered by Nerdy speakers are associated with a *narrower* range of values than numerals uttered by Chill speakers, supporting Hypothesis 1. Moreover, as suggested by the comparison with the "No.Persona" baseline condition, this effect is driven by both clusters of traits associated with the Nerdy and Chill persona. While no interaction between Persona and Similarity was found, we found that, limited to the Nerdy persona, precision in interpretation decreased as participants' self-ascribed nerdiness increased.

Taken together, these findings suggest that the persona embodied by the speaker affects the outcome of pragmatic reasoning; however, the question remains open of how comprehenders engage in the reverse reasoning — i.e., how they assess the appropriateness of a numerical description in light of a known value. Experiment 2 uses a Truth-Value judgment task to evaluate these possibilities.

5 Experiment 2: Assessing statements about given facts

In Section 3 we outlined two alternative hypotheses about how our Persona manipulation might affect comprehenders' adjudication of imprecise statements *vis-à-vis* an established fact. If comprehenders base their assessment on the range of values expected to fall under the intended extension of the numeral given the speaker's identity — i.e., a narrower one for Nerdy speakers — imprecise descriptions from Nerdy speakers should be rejected as 'wrong' more often than those coming from Chill speakers (Hypothesis 2A). But if comprehenders perceive Nerdy speakers as agents that are more invested in descriptive adequacy than Chill ones, they could see them as more trustworthy in the presence of an imprecise statement, and thus *accept* their imprecise descriptions more often (Hypothesis 2B).

5.1 Methods, design, materials & procedure

Experiment 2 adapts the setup of Experiment 1 to a paradigm in which participants are provided with the actual state of affairs in question, and are asked

to assess whether a statement correctly describes it. This setup corresponds to the *Truth Value Judgment* task, widely utilized in experimental semantics and pragmatics (Crain & McKee 1985, Noveck 2001, Papafragou & Musolino 2003, Doran et al. 2012).[11]

We implemented this setup by leaving the logic of Experiment 1 unchanged except for one crucial modification: participants were only shown the picture of one phone, with a number visible on the screen. They were told that the speaker was indeed looking at this phone, and asked to indicate whether the character's response was RIGHT or WRONG. The full display of an item is reproduced in Figure 5.

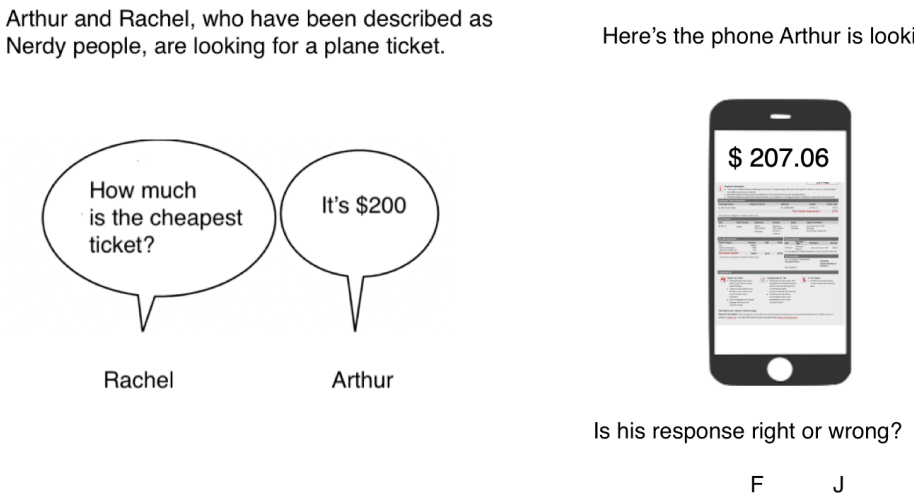

Figure 5 Experiment 2: Display before making the choice (Condition: Nerdy, Imprecise)

As before, the Screen Fit manipulation (within-subjects) varied the fit between the number on the phone and that uttered by the speaker (Match, Mismatch, Imprecise). The Persona manipulation (between-subjects) varied

11 Despite the term *Truth* in the name of the task, a number of different wordings have been used in the prompt, including whether participants thought the sentence was *right/wrong* (Crain & McKee 1985, Jasbi, Waldon & Degen 2019, Waldon & Degen 2020), or whether they *agreed/disagreed* with the statement (Noveck 2001). A recent study by Cremers, Fricke & Onea (2023) finds different formulations to yield comparable results.

the speaker persona, with Nerdy, Chill and No.Persona as levels. Once again, the speaker gender was counterbalanced. The experimental and filler items were the same as those used in Experiment 1. At the end of the experiment, participants gave themselves Nerdiness and Chillness rating in the same exit questionnaire deployed in Experiment 1.

5.2 Participants

244 participants were recruited on Prolific and compensated $8/hour (Age range: 18-82; Age Median: 33; female=135; male=102; other=7).

5.3 Results

Figure 6 plots the rate of rejections (="wrong" responses) by Persona and Screen Fit. As in Experiment 1, rejections were at floor and ceiling in Mismatch and Match respectively, with intermediate rates in the Imprecise condition. Second, rejection rates in the Imprecise condition differed depending on the speaker persona. However, contrary to Experiment 1, the difference seems to involve only the Chill speakers: while the rate of rejections for Chill speakers appears to be lower than the No.Persona condition, no difference can be observed between the Nerdy and the No.Persona condition.

For statistical analysis, we adopted the same procedure as in Experiment 1. First, we re-coded Screen Fit as a binary factor with "Control" (Match/Mismatch) and "Imprecise" as levels; second, we fit a (maximally converging) mixed-effects logistic regression on the WRONG choices with Persona (reference: NoPersona) and Screen Fit (sum-coded) as interacting predictors, and random intercepts and slopes for Screen Fit for items and participants. The model revealed a significant effect of Persona for Chill ($\beta = 1.04$; $SE = 0.16$; $z = 6.29$, $p < 0.0001$), reflecting the fact that WRONG choice rates for Chill speakers were lower than those in the No.Persona condition. But this effect was further qualified by an interaction with Screen Fit ($\beta = 1.00$; $SE = 0.16$; $z = 6.14$, $p < 0.0001$): in the Imprecise condition, WRONG choice rates were *lower* for Chill speakers than speakers in the No.Persona condition ($SE = 0.24$; $z = 8.43$, $p < 0.0001$), while no difference was observed in the control condition (Nerdy vs. No.Persona: $SE = 0.24$; $z = 0.05$, $p = 0.97$). We found no effect for Nerd ($\beta = 0.02$; $SE = 0.15$; $z = 0.15$, $p = 0.87$), and no interaction involving Nerd and Screen Fit ($\beta = 0.03$, $SE = 0.18$; $z = 0.18$, $p = 0.85$).

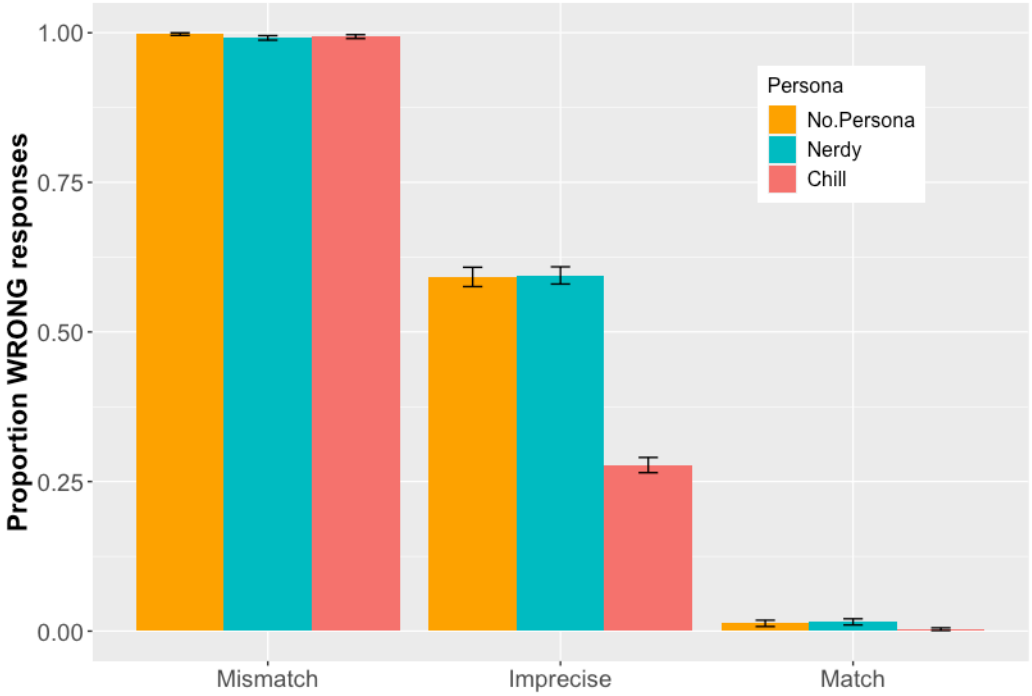

Figure 6 WRONG choices across Screen Fit, split by Persona

As with Experiment 1, we then explored whether the Persona effect in the Imprecise condition was modulated by the gender of the speaker and the self-assessed Nerdy/Chill ratings provided by participants. We fit another mixed-effects logistic regression on WRONG choices for the Imprecise condition data, with Persona, Speaker Gender (both sum-coded) and Similarity (re-scaled as in Experiment 1) as interacting predictors and random intercepts for items and participants. The choice patterns predicted by the model are shown in Figure 7.

As expected, we found a main effect of Persona (β=1.57; SE = 0.27; z = 5.68, $p < 0.0001$), with higher rates of WRONG responses for Nerdy speakers. Yet, despite the impressionistic increase in WRONG responses for Nerds for higher values of Similarity, no main effect of Similarity (β=0.24; SE = 0.27; z = 0.89, p = 0.36) was found. In addition, no effect of Speaker Gender (β=0.13; SE = 0.11; z = 1.17, p = 0.23) were found; and No.Persona*Similarity (β=0.27; SE = 0.27; z = 1.01, p = 0.30), Persona*Speaker Gender (β=0.07; SE = 0.07; z = 1.02, p = 0.30), Speaker Gender*Similarity (β=0.04; SE =

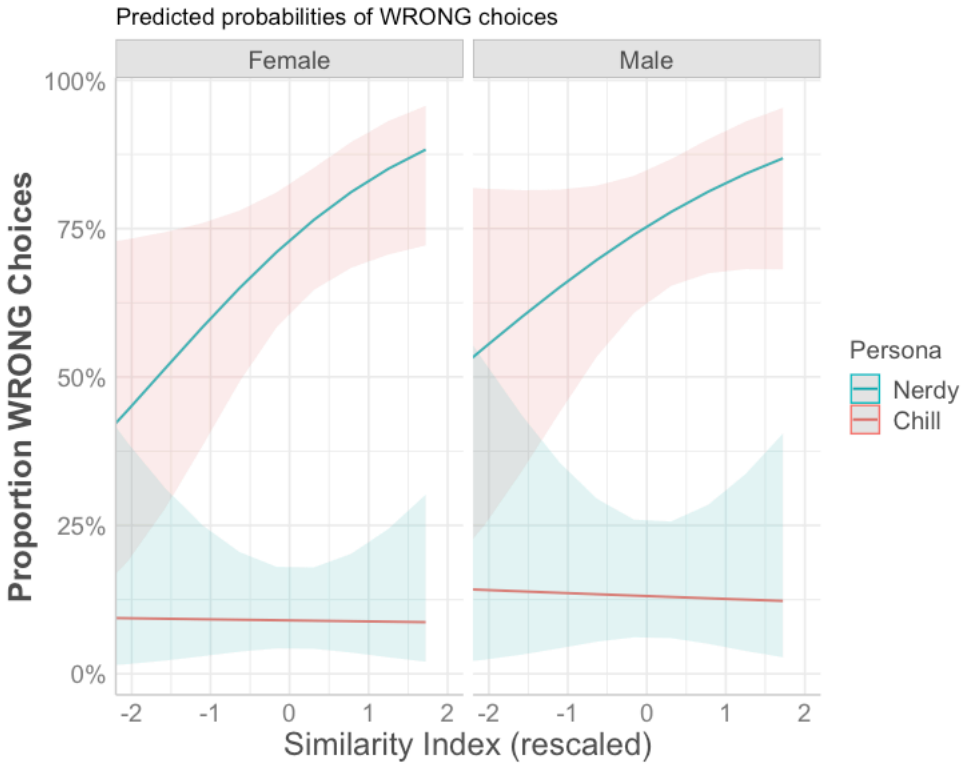

Figure 7 WRONG choices by Similarity and Speaker Gender (Imprecise only)

0.07; $z = 0.54$, $p = 0.58$), and Persona*Speaker Gender*Similarity (β=0.03; $SE = 0.07$; $z = 0.41$, $p = 0.67$) interactions were found.

5.4 Discussion

The findings from Experiment 2 partially support Hypothesis 2A: comprehenders show more leniency towards accepting the imprecise descriptions of Chill speakers, consistent with the idea, suggested by Experiment 1, that they associate numerals uttered by this type of speaker with a broader range of values. This is not observed, however, for the Nerdy speakers: even though their numerals are associated with an especially narrow range of values (see Experiment 1), this doesn't translate in comprehenders being stricter in the adjudication of imprecise utterances — as suggested by the fact that WRONG responses rates do not different across the Nerdy and the No.Persona condition. In addition, we didn't observe an effect of Similarity for either Nerdy or Chill speakers.

6 Combined analysis

Given the partially diverging results in Experiments 1 and 2 and the fact that they only differed minimally in the specifics of the task, we carried out a combined analysis pooling the data from the two studies. Figure 8 illustrates the rejection rates in the Imprecise condition across the two tasks.

Figure 8 Rejection rates across tasks: Covered Screen (Experiment 1) vs. Truth Value Judgement (Experiment 2)

To explore this pattern, we fit a mixed effect logistic regression with Task and Persona as interacting fixed effect factors, and random intercepts for Items and Participants.[12] Task was sum coded; Persona, consistent with the approach followed so far, was treatment-coded, with No.Persona as reference levels.

Confirming the results of the separate analyses for Experiment 1 and 2, we found an interaction between Persona and Task involving Nerdy speakers (β=0.62; SE=0.31; z = 1.99, p =0.04): while the Covered Screen Task featured a higher rejection rate for the Nerdy condition than the No.Persona one (SE=0.43; z = 4.40, p <0.0001), no difference was found in the Truth Value Judgment task (SE=0.45; z = 1.51, p =0.28). Moreover, we found no difference between tasks *within* any of the Persona conditions (Nerdy: SE=0.46;

12 The more complex model with random slopes had singular fit and virtually identical output.

$z = 1.76$, $p = 0.08$; Chill: SE=0.37; $z = 0.06$, $p = 0.95$; No.Persona: SE=0.37; $z = 1.00$; $p = 0.30$).

7 General discussion

In two studies, we found that comprehenders reason about the sociolinguistic profile of the speaker in drawing inferences related to meaning interpretation. Specifically, in the COVERED SCREEN task deployed in Experiment 1, comprehenders assigned a narrower extension to utterances from speakers embodying the social qualities linked to precise speech — i.e., the Nerdy ones; and a broader extension to utterances from speakers featuring the qualities linked to approximate speech — i.e., the Chill ones. This result provides support to Hypothesis 1, according to which comprehenders recruit information about the social characteristics of a speaker to determine the range of values represented by a numeral. In the Truth Value Judgment task deployed in Experiment 2, comprehenders accepted imprecise descriptions uttered by Chill speakers to a higher rate than imprecise descriptions uttered in the absence of social information (per Hypothesis 2A); yet, imprecise utterances by Nerdy speakers were accepted to the same rate as imprecise utterances in the absence of social information, failing to align with either Hypothesis 2A or 2B. We now turn to consider the methodological implications of these findings for semantics and pragmatics, as well as those concerning the connection between pragmatic reasoning and other domains human action.

In general, our results contribute to a more inclusive view of pragmatic reasoning as a socially embedded process, in line with the findings from recent work highlighting the sensitivity of pragmatic reasoning to different social dimensions (see Section 1 for details). Our studies expand on this body of work by providing two novel contributions. From an empirical standpoint, they broaden the range of social dimensions known to shape pragmatic reasoning, providing evidence that comprehenders track not only demographic or ideological information — as previously shown — but also more complex ensembles of identity and personality traits, which in turn coalesce in socially recognizable stereotypes. Most importantly, these clusters of social qualities correspond to those that, conversely, comprehenders infer from speakers' pragmatic choices, as demonstrated in prior work on the social perception of imprecise speech (Beltrama 2018, Beltrama, Solt & Burnett 2022; see Section 2). This highlights a bi-directional relationship between social and pragmatic inferences, similar to what has been found in many domains of speech per-

ception (see Section 1): comprehenders infer specific constellations of social information from a speaker's pragmatic behavior; and recruit the very same traits as a cue to navigate pragmatic reasoning.

Furthermore, our results unveil a parallel between meaning interpretation and domains of cognition that goes beyond the realm of meaning narrowly construed. Work in social psychology and philosophy illuminated how people are able to form an impression of someone's personality and identity from what they do or what they look like (Fiske et al. 1999, Fiske, Cuddy & Glick 2007, Judd et al. 2005, Goodwin, Piazza & Rozin 2014, Fiske 2018, Stolier, Hehman & Freeman 2020); and that these impressions shape different realms of human action, including decision-making (e.g., Jenkins et al. 2018) and epistemic reasoning (e.g., Fricker 2007, Davis 2016; Section 1 for further details). Our findings highlight a crucial similarity between pragmatic reasoning and these domains. In particular, by virtue of being underlied by traits pertaining to High Competence vs. High Warmth traits respectively, and by having been effectively introduced as such in the experiments, Nerdy and Chill stereotypes can be seen as occupying distinct quadrants of the two-dimensional space that is commonly taken to provide the scaffolding of processes of person perception (see Fiske 2018; Section 2 for further details). The upshot is that pragmatic reasoning can be shaped by social constructs and clusters of traits that are *very similar* to those whose influence has been demonstrated in other areas of linguistic behavior and beyond.

At the same time, it is important to point out two respects in which the generalizabilty of our findings is limited. First, the effect of social information on pragmatic reasoning does not extend to all the social variables we kept track of in the studies. Most notably, we found no effect of speaker gender. We see this result as quite striking, especially in light of gender's prominent role in other domains of epistemic reasoning (Fricker 2007, McCready & Winterstein 2017, Mazzarella & Pouscoulous 2020 a.o.), as well as in language perception and production. A possible explanation is that the stereotypes utilized in our studies might have been so salient that they obscured the contrast between other social dimensions. This possibility is suggested by the idea, articulated in the sociolinguistics literature, that stereotypes are holistic constructs, whose value and salience goes beyond the "compositional" sum of individual demographic and personality traits that (Agha 2003, Podesva 2011, D'Onofrio 2020). Specifically, the stereotypes of the speaker as an intellectually invested, high-Competence Nerdy person vs. a sociable, high-Warmth Chill person might have been perceived sim-

ilarly across different genders — consistent with the observation that personae along these lines have been discussed in connection to male (Kiesling 2009) and female characters (Bucholtz 2001, Pratt 2021) alike. Moreover, it's important to note that, as also shown in prior experimental work, effects of gender on pragmatic reasoning are modulated by discourse factors. For example, McCready & Winterstein (2017) show that comprehenders' judgments of convincingness are jointly shaped by the speaker gender and the topic of the conversation; accordingly, the impact of speaker gender on the assessment of imprecise statements might have been similarly subordinate to further discourse factors which were not manipulated in our study. Furthermore, we only found a limited effect of participants' similarity to the speaker across both studies: while in Experiment 1 precise interpretations with Nerdy speakers are highest for participants who don't see themselves as Nerds, no interaction between Persona and Similarity was found, and no corresponding effect on the interpretation of Chill speakers' utterances. In Experiment 2, no interaction or effect involving Similarity *tout court* was found. Yet, effects of social information on language processing and decision making have often been shown to be modulated by respondents' own identity (Park & Judd 1990, Judd, Ryan & Park 1991, Niedzielski 1999, Babel 2010, Walker & Campbell-Kibler 2015, Levon 2016, Wade 2020). Taken together, these considerations suggest a view in which certain specific aspects of pragmatic inference are influenced by certain specific dimensions of social variation; but not one in which this happens at *any* level. They thus raise the question of whether the outcome of imprecision resolution is equally sensitive to other sociolinguistic dimensions besides the stereotype-level distinctions manipulated in our study; and why, if this is the case, these effects didn't emerge in our studies.

Another limitation concerns the empirical scope of our study. As discussed in Section 2, (im)precision was selected as a case study in that it provided a promising testbed to observe the effects of interest. But of course, not all cases of meaning indeterminacy are necessarily intertwined with social stereotypes as deeply as (im)precision — or at least, we do not know that yet. It remains to be seen whether, for example, the resolution of vague predicates (e.g., *tall*) is subject to similar sociolinguistic dynamics as those highlighted for numerals (though see Beltrama, He & Schwarz 2024 for evidence that social considerations matter here as well). In light of this, we see our findings on (im)precision as a proof of concept that social perception *can* impact meaning resolution; and that the same question *should* therefore extended, in whatever way is appropriate, to other domains of pragmatic reasoning.

Broadening the empirical base of this work will also make it possible to better understand how the interaction between social information and meaning interpretation can be captured in formal models of pragmatic inference. While we don't see our results as providing conclusive support to any particular framework, we believe that recent developments of game-theoretic frameworks provide an especially promising route (Burnett 2017, 2019, Henderson & McCready 2019, McCready & Henderson 2020). In this approach, the signaling and uptake of social meanings is captured with the same probabilistic tools deployed to formalize pragmatic inferences, capitalizing on the idea interlocutors can cooperatively (and recursively) reason about what identity traits they want each other to recognize, and calibrate their conversational and inferential behavior accordingly. To our knowledge, work in this tradition hasn't explicitly investigated how sociolinguistic perception impacts pragmatic inferences; but a setup where conversational agents come with their own sets of prior probabilities — which differ depending on agential background, information, assumptions and overall goals — emerges as especially conducive to capture the sensitivity of pragmatic reasoning to patterns of sociolinguistic variation.

Our findings also provide a methodological contribution, as they afford a novel angle to consider the context-sensitivity of judgments commonly used in meaning-based tasks — i.e., inference and Truth-Value judgment tasks. In this regard, it's important to note that the effect of social information — at least in connection to Chill speakers — is general enough to apply across two related, yet distinct kinds of meaning inferences, both of which have been at the core of experimental paradigms aiming to tap into the outcome of meaning interpretation: comprehenders' determination of the intended extension of a particular expression; and their assessment of the acceptability of a description *vis-à-vis* a known, established body of facts. This suggests that, to a certain extent, the impact of social stereotypes on pragmatic reasoning impacts comprehenders' interpretation process above and beyond the specific inference deployed in a task. In light of these considerations, the question arises as to why responses to Nerdy speakers' utterances are only influenced by social information in the COVERED SCREEN task, but not the TVJ one — a pattern of results that wasn't predicted by any of the hypotheses we put forward. However, one possibility we would like to suggest, and ultimately test in future work, is that this asymmetry between stereotypes might be linked to epistemic considerations similar to those that had originally motivated Hypothesis 2B — even though the specific prediction of

that hypothesis is not borne out in the data. To see this, it is first crucial to consider that, in the TVJ task, a rejection response — i.e., labeling the response as 'wrong' — is crucially *prejudicial*: it commits the comprehender to expressing a negative assessment of the speaker' conversational behavior as someone who failed to correctly represent the relevant facts. This contrasts with the Covered Screen paradigm, in which a participant's selection of the COVERED screen simply indicates the belief that the speaker must have been looking at a different phone. Accordingly, Nerdy speakers' social perception as agents invested in descriptive accuracy, while not strong enough to lead comprehenders to *accept* their imprecise descriptions to a higher rate (as per Hypothesis 2B), might have nevertheless made respondents hesitant to see Nerdy speakers as conversationally blameworthy — effectively offsetting the higher rate of rejections relative to the baseline condition predicted by Hypothesis 2A. At the same time, we note that, following this line of reasoning, we should predict a lower rate of rejections in the TVJ task relative to the Covered Screen task: if rejection is prejudicial, and participants generally want to avoid blaming the speaker, their responses should be more charitable. Yet, not only are the overall acceptance rates in Experiment 2 *not* higher; with the Nerdy persona, they even tend to be *lower*. In light of this observation, further empirical evidence will be needed to support this explanation. In particular, online behavioral measures tracking the time-course of interpretation might be especially promising to illuminate participants' reasoning with respect to possible differential preferences for precise vs. imprecise interpretations across the two tasks — even in the face of similar overall rejection rates. For example, Aparicio (2017) reports evidence from a self-paced reading task indicating that imprecise interpretations, cued either linguistically or contextually, incurred higher reading times than precise ones, turning out to be more costly to comprehenders than imprecise ones. Similar time-course evidence, following appropriate revision of the two paradigms, could provide a suitable basis to investigate differences in compehenders' preferences for precise vs. imprecise interpretation across the two tasks that — including those possibly linked to the additional prejudicial component involved in TVJs' rejections. Combined with the evidence from the present investigations, this work will further contribute to understanding how the outcome of TVJs is affected by contextual and epistemic factors that go beyond the descriptive meaning of the sentence being adjudicated — in line with the recent body of work that independently unveiled the impact

of pragmatic considerations on these experimental measures (Sikos, Kim & Grodner 2019, Scontras & Pearl 2021, Waldon & Degen 2020).

8 Conclusion

Our results provide a novel perspective on link between the socio-indexical and descriptive dimensions of meaning: they highlight the role of speakers' social identity in guiding meaning interpretation; and point to different, promising directions of investigation to shed further light on the relationship between sociolinguistic perception and pragmatic reasoning.

Appendix

Link to supplementary materials.

References

Acton, Eric. 2019. Pragmatics and the social life of the English definite article. *Language* 95(1). 37–65. https://doi.org/10.1353/lan.2019.0010.

Acton, Eric & Chris Potts. 2014. That straight talk: Sarah Palin and the sociolinguistcs of demonstratives. *Journal of Sociolinguistics* 18(1). 3–31. https://doi.org/10.1111/josl.12062.

Agha, Asif. 2003. The social life of a cultural value. *Language and Communication* 23(3–4). 231–273. https://doi.org/10.1016/S0271-5309(03)00012-0.

Aparicio, Helena. 2017. *Processing context-sensitive expressions: The case of gradable adjectives and numerals*. Chicago, IL: University of Chicago dissertation. https://doi.org/10.6082/M1G15Z1F.

Babel, Molly. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39(4). 437–456. https://doi.org/10.1017/s0047404510000400.

Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40(1). 177–189. https://doi.org/10.1016/j.wocn.2011.09.001.

Beltrama, Andrea. 2018. Precision and speaker qualities: The social meaning of pragmatic detail. *Linguistics Vanguard* 4(1). https://doi.org/10.1515/lingvan-2018-0003.

Beltrama, Andrea. 2020. Social meaning in semantics and pragmatics. *Language and Linguistics Compass* 14(9). https://doi.org/10.1111/lnc3.12398.

Beltrama, Andrea. 2021. *Just* perfect, *simply* the best: An analysis of emphatic exclusion. *Linguistics and Philosophy* 45(2). 321–364. https://doi.org/10.1007/s10988-021-09326-x.

Beltrama, Andrea, Joyce He & Florian Schwarz. 2024. *It's not just imprecision: Stereotypes guide vagueness resolution in implicit comparisons*. Poster to be presented at Sinn und Bedeutung 29.

Beltrama, Andrea, Stephanie Solt & Heather Burnett. 2022. Context, precision, and social perception: A sociopragmatic study. *Language in Society* 52(5). 805–835. https://doi.org/10.1017/s0047404522000240.

Beltrama, Andrea & Laura Staum Casasanto. 2021. The social meaning of semantic properties. In Lauren Hall-Lew, Emma Moore & Robert J. Podesva (eds.), *Social meaning and linguistic variation: Theorizing the third wave.* Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108578684.004.

Benor, Sarah Bunin. 2004. *Talmid chachams* and *tsedeykeses*: Language, learnedness, and masculinity among Orthodox Jews. *Jewish Social Studies* 11(1). 147–170. https://doi.org/10.1353/jss.2005.0001.

Bergen, Leon. 2016. *Joint inference in pragmatic reasoning*. Cambridge, MA: Massachusetts Institute of Technology dissertation. https://doi.org/1721.1/106430.

Bonnefon, Jean-François, Aidan Feeney & Gaëlle Villejoubert. 2009. When some is actually all: Scalar inferences in face-threatening contexts. *Cognition* 112(2). 249–258. https://doi.org/10.1016/j.cognition.2009.05.005.

Breheny, Richard, Heather J. Ferguson & Napoleon Katsos. 2013. Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition* 126(3). 423–440. https://doi.org/10.1016/j.cognition.2012.11.012.

Bucholtz, Mary. 1999. "Why be normal?": Language and identity practices in a community of nerd girls. *Language in Society* 28(2). 203–223. https://doi.org/10.1017/S0047404599002043.

Bucholtz, Mary. 2001. The whiteness of nerds: Superstandard English and racial markedness. *Journal of Linguistic Anthropology* 11(1). 84–100. https://escholarship.org/uc/item/0663j2rg.

Burnett, Heather. 2017. Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of Sociolinguistics* 21(2). 238–271. https://doi.org/10.1111/josl.12229.

Burnett, Heather. 2019. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy* 42(5). 419–450. https://doi.org/10.1007/s10988-018-9254-y.

Campbell-Kibler, Kathryn. 2011. The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change* 22(3). 423–441. https://doi.org/10.1017/S0954394510000177.

Crain, Stephen & Cecile McKee. 1985. Acquisition of structural restrictions on anaphora. *North East Linguistics Society (NELS)* 16(1). 94–110. https://doi.org/20.500.14394/36490.

Cremers, Alexandre, Lea Fricke & Edgar Onea. 2023. The importance of being earnest: How truth and evidence affect participants' judgments. *Glossa Psycholinguistics* 2(1). 1–17. https://doi.org/10.5070/g6011172.

Cummins, Chris, Uli Sauerland & Stephanie Solt. 2012. Granularity and scalar implicature in numerical expressions. *Linguistics and Philosophy* 35(2). 135–169. https://doi.org/10.1007/s10988-012-9114-0.

D'Onofrio, Annette. 2015. Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics* 19(2). 241–256. https://doi.org/10.1111/josl.12115.

D'Onofrio, Annette. 2018. Personae and phonetic detail in sociolinguistic signs. *Language in Society* 47(4). 513–539. https://doi.org/10.1017/S0047404518000581.

D'Onofrio, Annette. 2019. Complicating categories: Personae mediate racialized expectations of non-native speech. *Journal of Sociolinguistics* 23(4). 346–366. https://doi.org/10.1111/josl.12368.

D'Onofrio, Annette. 2020. Personae in sociolinguistic variation. *WIREs Cognitive Science* 11(6). e1543. https://doi.org/10.1002/wcs.1543.

Davis, Emmalon. 2016. Typecasts, tokens, and spokespersons: A case for credibility excess as testimonial injustice. *Hypatia* 31(3). 485–501. https://doi.org/10.1111/hypa.12251.

Doran, Ryan, Gregory Ward, Meredith Larson, Yaron McNabb & Rachel E. Baker. 2012. A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language* 88(1). 124–154. http://www.jstor.org/stable/41348885.

Drager, Katie K. 2015. *Linguistic variation, identity construction and cognition* (Studies in Laboratory Phonology 2). Berlin: Language Science Press. https://doi.org/10.17169/LANGSCI.B75.22.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453–476. https://doi.org/10.1111/j.1467-9841.2008.00374.x.

Fairchild, Sarah & Anna Papafragou. 2018. Sins of omission are more likely to be forgiven in non-native speakers. *Cognition* 181. 80–92. https://doi.org/10.1016/j.cognition.2018.08.010.

Fiske, Susan T. 2018. Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science* 27(2). 67–73. https://doi.org/10.1177/0963721417738825.

Fiske, Susan T., Amy J.C. Cuddy & Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences* 11(2). 77–83. https://doi.org/10.1016/j.tics.2006.11.005.

Fiske, Susan T., Juan Xu, Amy C. Cuddy & Peter Glick. 1999. (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues* 55(3). 473–489. https://doi.org/10.1111/0022-4537.00128.

Fricker, Miranda. 2007. *Epistemic injustice: Power and the ethics of knowing*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001.

Gazdar, Gerald. 1980. Pragmatics and logical form. *Journal of Pragmatics* 4(1). 1–13. https://doi.org/10.1016/0378-2166(80)90014-4.

Glass, Lelia. 2015. Strong necessity modals: Four socio-pragmatic corpus studies. *University of Pennsylvania Working Papers in Linguistics* 21(2): *Selected papers from New Ways of Analyzing Variation (NWAV) 43*. Sabriya Fisher (ed.). 79–88. https://repository.upenn.edu/handle/20.500.14332/45055.

Goodwin, Geoffrey P., Jared Piazza & Paul Rozin. 2014. Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology* 106(1). 148–168. https://doi.org/10.1037/a0034726.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Speech acts* (Syntax and Semantics 3), 41–58. New York: Academic Press. https://doi.org/10.1163/9789004368811_003.

Social Meaning and Linguistic Variation. 2021. In Lauren Hall-Lew, Emma Moore & Robert J. Podesva (eds.). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108578684.

Hay, Jennifer & Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48(4). 865–892. https://doi.org/10.1515/LING.2010.027.

Henderson, Robert & Elin McCready. 2019. Dogwhistles, trust and ideology. *Amsterdam Colloquium 22*. 152–160. https://archive.illc.uva.nl/AC/AC2019/uploaded_files/inlineitem/Henderson_and_McCready_Dogwhistles_trust_and_ideolo.pdf.

Horn, Laurence. 1984. Towards a new taxonomy of pragmatic inference: Q-based and R-based implicature. In Deborah Schiffrin (ed.), *Meaning, form, and use in context: Linguistic applications*, 11–42. Washington: Georgetown University Press. https://doi.org/10822/555477.

Huang, Yi Ting, Elizabeth Spelke & Jesse Snedeker. 2013. What exactly do number words mean? *Language Learning and Development* 9(2). 105–129. https://doi.org/10.1080/15475441.2012.658731.

Jasbi, Masoud, Brandon Waldon & Judith Degen. 2019. Linking hypothesis and number of response options modulate inferred scalar implicature rate. *Frontiers in Psychology* 10. https://doi.org/10.3389/fpsyg.2019.00189.

Jenkins, Adrianna C., Pierre Karashchuk, Lusha Zhu & Ming Hsu. 2018. Predicting human behavior toward members of different social groups. *Proceedings of the National Academy of Sciences* 115(39). 9696–9701. https://doi.org/10.1073/pnas.1719452115.

Jeong, Sunwoo. 2021. Deriving politeness from an extended Lewisian model: The case of rising declaratives. *Journal of Pragmatics* 177. 183–207. https://doi.org/10.1016/j.pragma.2021.02.017.

Judd, Charles M., Laurie James-Hawkins, Vincent Yzerbyt & Yoshihisa Kashima. 2005. Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology* 89(6). 899–913. https://doi.org/10.1037/0022-3514.89.6.899.

Judd, Charles M., Carey S. Ryan & Bernadette Park. 1991. Accuracy in the judgment of in-group and out-group variability. *Journal of Personality and Social Psychology* 61(3). 366–379. https://doi.org/10.1037/0022-3514.61.3.366.

Kao, Justine T., Jean Y. Wu, Leon Bergen & Noah D. Goodman. 2014. Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences* 111(33). 12002–12007. https://doi.org/10.1073/pnas.1407479111.

Kiesling, Scott F. 2009. Style as stance: Can stance be the primary explanation for patterns of sociolinguistic variation? In Alexandra Jaffe (ed.), *Stance: Sociolinguistic perspectives*, 171–194. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195331646.003.0008.

Kiesling, Scott F. 2018. Masculine stances and the linguistics of affect: On masculine ease. *NORMA* 13(3-4). 191–212. https://doi.org/10.1080/18902138.2018.1431756.

Kinzler, Katherine D. 2021. Language as a social cue. *Annual Review of Psychology* 72. 241–264. https://doi.org/10.1146/annurev-psych-010418-103034.

Klecha, Peter. 2018. On unidirectionality in precisification. *Linguistics and Philosophy* 41(1). 87–124. https://doi.org/10.1007/s10988-017-9216-9.

Krifka, Manfred. 2009. Approximate interpretations of number words: A case for strategic communication. In Erhard Hinrichs & John Nerbonne (eds.), *Theory and evidence in semantics*, 109–132. Stanford, CA: CSLI Publications.

Labov, William. 1966. *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 75(3). 522–551. https://doi.org/10.2307/417059.

Leffel, Timothy, Ming Xiang & Christopher Kennedy. 2016. Imprecision is pragmatic: Evidence from referential processing. *Semantics and Linguistic Theory (SALT)* 26. 836–854. https://doi.org/10.3765/salt.v26io.3937.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/5526.001.0001.

Levon, Erez. 2016. Gender, interaction and intonational variation: The discourse functions of High Rising Terminals in London. *Journal of Sociolinguistics* 20(2). 133–163. https://doi.org/10.1111/josl.12182.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(1). 339–359. https://doi.org/10.1007/BF00258436.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In William J. Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 403–439. Netherlands: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-009-2037-8_16.

Mahler, Taylor. 2022. *Social identity information in projection inferences: A case study in social and semantic-pragmatic meaning*. Columbus, OH: The Ohio State University dissertation.

Mazzarella, Diana & Nausicaa Pouscoulous. 2020. Pragmatics and epistemic vigilance: A developmental perspective. *Mind and Language* 36(3). 355–376. https://doi.org/10.1111/mila.12287.

Mazzarella, Diana, Emmanuel Trouche, Hugo Mercier & Ira Noveck. 2018. Believing what you're told: Politeness and scalar inferences. *Frontiers in Psychology* 9. https://doi.org/10.3389/fpsyg.2018.00908.

McCready, Elin & Robert Henderson. 2020. Social meaning in repeated interactions. In Christine Howes, Stergios Chatzikyriakidis, Adam Ek & Vidya Somashekarappa (eds.), *Probability and Meaning conference (PaM 2020)*, 69–72. Gothenburg: Association for Computational Linguistics. https://aclanthology.org/2020.pam-1.9.

McCready, Elin & Grégoire Winterstein. 2017. Negotiating epistemic authority. *New Frontiers in Artificial Intelligence*: *SAI-isAI 2016 Workshops, LENLS, HAT-MASH, AI-Biz, JURISIN and SKL, Kanagawa, Japan, November 14-16, 2016, Revised Selected Papers*. 74–89. https://doi.org/10.1007/978-3-319-61572-1_6.

Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Social Psychology* 18(1). 62–85. https://doi.org/10.1177/0261927X99018001005.

Noveck, Ira. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78(2). 165–188. https://doi.org/10.1016/S0010-0277(00)00114-1.

Papafragou, Anna & Julien Musolino. 2003. Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition* 86(3). 253–282. https://doi.org/10.1016/s0010-0277(02)00179-8.

Park, Bernadette & Charles M. Judd. 1990. Measures and models of perceived group variability. *Journal of Personality and Social Psychology* 59(2). 173–191. https://doi.org/10.1037/0022-3514.59.2.173.

Pinkal, Manfred. 1995. *Logic and lexicon: The semantics of the indefinite*. Trans. by Geoffrey Simmons (Studies in Linguistics and Philosophy (SLAP) 56). Dordrecht: Springer Science+Business Media. https://doi.org/10.1007/978-94-015-8445-6.

Podesva, Robert J. 2011. Salience and the social meaning of declarative contours: Three case studies of gay professionals. *Journal of English Linguistics* 39(3). 233–264. https://doi.org/10.1177/0075424211405161.

Podesva, Robert J., Jermay Reynolds, Patrick Callier & Jessica Baptiste. 2015. Constraints on the social meaning of released /t/: A production and perception study of U.S. politicians. *Language Variation and Change* 27(1). 59–87. https://doi.org/10.1017/S0954394514000192.

Pratt, Teresa. 2021. Affect in sociolinguistic style. *Language in Society* 52(1). 1–26. https://doi.org/10.1017/s0047404521000774.

Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics, 1998 revision. In Jae Hak Yoon & Andreas Kathol (eds.), *Papers in semantics* (Working Papers in Linguistics 49). The Ohio State University, Department of Linguistics. A revised version was published in Roberts 2012.

Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6). 1–69. https://doi.org/10.3765/sp.5.6.

Sauerland, Uli & Penka Stateva. 2011. Two types of vagueness. In Paul Égré & Nathan Klinedinst (eds.), *Vagueness and language use* (Palgrave Studies in Pragmatics, Language and Cognition (PSPLC)), 121–145. London: Palgrave Macmillan. https://doi.org/10.1057/9780230299313_6.

Schwarz, Florian, Jacopo Romoli & Cory Bill. 2016. Reluctant acceptance of the literal truth: Eye tracking in the Covered Box paradigm. *Sinn und Bedeutung (SuB)* 20. 61–78. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/286.

Schwarz, Florian & Jeremy Zehr. 2021. Tutorial: Introduction to PCIbex – An open-science platform for online experiments: Design, data-collection and code-sharing. *Annual Meeting of the Cognitive Science Society (CogSci)* 43. https://escholarship.org/uc/item/1ng1q4c6 (10 November, 2021).

Scontras, Gregory & Lisa S. Pearl. 2021. When pragmatics matters more for truth-value judgments: An investigation of quantifier scope ambiguity. *Glossa: a journal of general linguistics* 6(1). https://doi.org/10.16995/glossa.5724.

Sikos, Les, Minjae Kim & Daniel J. Grodner. 2019. Social context modulates tolerance for pragmatic violations in binary but not graded judgments. *Frontiers in Psychology* 10. https://doi.org/10.3389/fpsyg.2019.00510.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language and communication* 23(3–4). 193–229. https://doi.org/10.1016/S0271-5309(03)00013-2.

Solt, Stephanie. 2014. An alternative theory of imprecision. *Semantics and Linguistic Theory (SALT)* 24. https://doi.org/10.3765/salt.v24i0.2446.

Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi & Deirdre Wilson. 2010. Epistemic vigilance. *Mind and Language* 25(4). 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x.

Squires, Lauren. 2013. It don't go both ways. Limited bidirectionality in sociolinguistic perception. *Journal of Scoiolinguistics* 17(2). 200–237. https://doi.org/10.1111/josl.12025.

Stolier, Ryan, Eric Hehman & Jonathan B. Freeman. 2020. Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour* 4(4). 361–371. https://doi.org/10.1038/s41562-019-0800-6.

Syrett, Kristen, Christopher Kennedy & Jeffrey Lidz. 2009. Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics* 27(1). 1–35. https://doi.org/10.1093/jos/ffp011.

Van Der Henst, Jean-Baptiste, Laure Carles & Dan Sperber. 2002. Truthfulness and relevance in telling the time. *Mind and Language* 17(5). 457–466. https://doi.org/10.1111/1468-0017.00207.

Wade, Lacey. 2020. *The linguistic and the social intertwined: Linguistic convergence toward southern speech*. Philadelphia, PA: University of Pennsylvania dissertation.

Waldon, Brandon & Judith Degen. 2020. Modeling behavior in truth value judgment task experiments. *Society for Computation in Linguistics* 3. 10–19. https://doi.org/10.7275/SG32-AQ30.

Walker, Abby & Kathryn Campbell-Kibler. 2015. Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology* 6. https://doi.org/10.3389/fpsyg.2015.00546.

Yoon, Erica J., Michael Henry Tessler, Noah D. Goodman & Michael C. Frank. 2020. Polite speech emerges from competing social goals. *Open Mind* 4. 71–87. https://doi.org/10.1162/opmi_a_00035.

Andrea Beltrama
Department of Linguistics
3401-C Walnut St, Suite 300
University of Pennsylvania
Philadelphia, PA 19104-6305
andrea.beltrama@gmail.com

Florian Schwarz
Department of Linguistics
3401-C Walnut St, Suite 300
University of Pennsylvania
Philadelphia, PA 19104-6305
florians@sas.upenn.edu