AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 5, 1245–1263
https://doi.org/10.1037/xge0001345

Real-Time Inference in Communication Across Cultures: Evidence From a Nonindustrialized Society

Rachel Ryskin[1], Miguel Salinas[2], Steven Piantadosi[3], and Edward Gibson[2]
[1] Department of Cognitive & Information Sciences, University of California, Merced
[2] Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology
[3] Department of Psychology, University of California, Berkeley

In everyday communication, speakers and listeners make sophisticated inferences about their conversation partner's intended meaning. They combine their knowledge of the visuospatial context with reasoning about the other person's knowledge state and rely on shared assumptions about how language is used to express communicative intentions. However, these assumptions may differ between languages of nonindustrialized—where conversations often primarily take place within a, so-called, *society of intimates*—and industrialized cultures—*societies of strangers*. Here, we study inference in communication in the Tsimane', an indigenous people of the Bolivian Amazon, who have little contact with industrialization or formal education. Using a referential communication task, we probe how Tsimane' speakers refer to objects in the world around them when there are potential ambiguities (e.g., referring to a cup when there are multiple cups in view) across different visual contexts. Using an eye-tracking task, we probe the real-time inferences that Tsimane' listeners make about the speaker's intentions. We find that Tsimane' speakers use visual (color, size) contrasts to disambiguate referents (e.g., "Hand me the *small* cup"), much like English speakers, and they predictively direct their gaze to objects in a contrast set when they hear a modifier (e.g., "small"). Despite myriad cultural and linguistic dissimilarities between the two populations, the qualitative patterns of behavior and eye-gaze of Tsimane' and English speakers were strikingly similar, suggesting that the communicative expectations underlying many everyday inferences may be shared across cultures.

Keywords: communication, language, cross-cultural, eye-tracking

A primary function of language is to allow us to describe the world around us to others (Tomasello, 1999). When communicating about an object that they perceive, the speaker's goal is to design a linguistic utterance that will allow the listener to identify the same object. The listener's goal is to decode the linguistic signal and link its meaning to a percept in their environment. These types of commonplace exchanges rely on a number of cognitive faculties and social processes, from visual perception to reasoning about the knowledge state of the conversation partner. The linguistic forms that are chosen by speakers, and how they are interpreted by listeners, reflect both shared cognitive mechanisms (e.g., the ability to judge which of two similar objects is smaller) and cultural knowledge (e.g., conventional mappings between words and meanings). Yet, they have primarily been studied in so-called WEIRD (western, educated, industrialized, rich, and democratic) cultures (Henrich et al., 2010). These samples, though most accessible to many researchers, may lead to biased conclusions about the basic nature of human cognitive processes and behaviors (Gurven, 2018). For example, aspects of perception and decision-making that were thought to be universal (e.g., perception of musical dissonance, spatial reasoning, and cost–benefit analysis), seem to be adaptations to WEIRD environments (e.g., Flynn, 2007; Henrich et al., 2001; Jacoby et al., 2019; Pitt et al., 2021). Communication—sitting at the nexus of cognition and culture—may be particularly important to study cross-culturally. In what follows, we review the literature on reference in communication and discuss how behavioral patterns thought to be universal may be subject to cultural adaptation. We then report empirical findings from a study of production and comprehension of referential expressions in the Tsimane', a nonindustrialized group indigenous to the Bolivian Amazon. We observe striking similarities in communicative behaviors, suggesting that basic aspects of how we design and interpret references to objects in our environment may be invariant to culture and rely primarily on shared cognitive mechanisms, such as visuospatial processing and pragmatic inference.

This article was published Online First April 27, 2023.
Rachel Ryskin https://orcid.org/0000-0002-9516-4467
This work was supported by Grant 1760874 from the National Science Foundation, Division of Research on Learning (to Steven Piantadosi). Special thanks to Tomás Huanca and Esther Conde for coordinating our fieldwork, to many interpreters, including Manuel Roca and Robin Nate, for their hard work, and to Jerry Zhu and Xinzhu Fang, for their help with data collection and coding. We are also grateful to Paula Rubio-Fernandez for discussions of this work and to Daphna Heller for providing detailed comments on a previous draft of the manuscript.

All data and code for this paper can be found at https://osf.io/bjs85/. The ideas and some of the data presented here were previously disseminated as a conference presentation at the 2019 Meeting of the Human Sentence Processing Society and as a preprint (https://psyarxiv.com/ga83j/).
Correspondence concerning this article should be addressed to Correspondence concerning this article should be addressed to Rachel Ryskin, Department of Cognitive & Information Sciences, University of California, Merced, Merced, CA 95343, United States. Email: rryskin@ucmerced.edu

Producing and Understanding Reference

Studies of English and languages from a few other industrialized cultures have shown that there is a tight coupling between how speakers use referential expressions (e.g., "this dog," "the small cup," etc.) and how listeners predictively infer what the speaker is referring to. This link motivates the construal of speakers and listeners as rational agents engaged in a cooperative task (Degen et al., 2020; Goodman & Frank, 2016; Heller et al., 2016; Zaslavsky et al., 2018). As proposed by Grice (1975), speakers design utterances that will be optimally informative for the listener (Olson, 1970; Clark & Wilkes-Gibbs, 1986), and listeners make sophisticated pragmatic inferences about the speaker's intended message (Chambers et al., 2002; Sedivy et al., 1999). For instance, when they hear "Hand me the big…," listeners anticipate the referent to belong to a pair with a smaller item (e.g., a big cup and a small cup). This inference is thought to reflect the listener's expectation that speakers aim to say enough to deliver an unambiguous message, but no more than necessary (i.e., the speaker would not have said "big" unless omitting it might cause confusion).[1]

Further, the likelihood of this contrastive inference appears to be closely tied to the statistical distribution of referents and adjectives that listeners experience in the world. While size adjectives are only used to describe an object when a larger or smaller version is also present, color adjectives are often produced even in situations where the referent is unique (Brown-Schmidt & Konopka, 2011; Pechmann, 1989; Rubio-Fernández, 2016; Sedivy, 2005). Some have argued that this redundancy is due to the visual salience of color (Arts et al., 2011; van Gompel et al., 2019). Other accounts of this asymmetry have argued that the redundant use of color adjectives is consistent with there being a more stable mapping between color words and the color concepts they identify, relative to the mapping between size words and concepts (Degen et al., 2020; Kennedy & McNally, 2005; Sedivy, 2003). For example, "red" consistently picks out more or less the same property, whether it is in reference to a "red cup" or a "red building" (but see Cohen & Murphy, 1984). In contrast, the same size adjective can be used to refer to many different conceptual sizes (e.g., "big cup" vs. "big building"). It follows that including a size adjective in an object description, in the absence of a contrast, provides minimal information about the identity of the referent.[2] Thus the speaker's choice to include a size adjective should be highly diagnostic of a contrast, whereas the inclusion of a color adjective should be less diagnostic. Indeed, listeners are less likely to draw a contrastive inference when they hear a color adjective (Sedivy, 2003) than when they hear a size adjective (but see Aparicio et al., 2016).

The Role of Culture

Human languages are highly diverse in their sound inventories and patterns, their morphology and syntax, and even the ways that words carve up meaning space (Evans & Levinson, 2009; Thompson et al., 2020). Their forms are the product of cultural evolution constrained by cognitive pressures (Smith & Kirby, 2008). In particular, the industrialization of a culture can impact referential communication in multiple ways. For instance, education (in the WEIRD tradition) affects how people construe similarities between entities in the world (Tanaka & Taylor, 1991). Industrialized populations tend to group objects into categories based on taxonomic properties while traditional societies often draw category boundaries related to their experience with the functions of objects (Luria, 1976). Cross-cultural differences in conceptual structure may have implications for what constitutes a relevant contrast set (e.g., two similar objects differing only in size or color) and what properties (size vs. color) of an object are more or less stable across contexts. For example, cultures differ in terms of the average individual's ecological expertise (Atran & Medin, 2008): a typical WEIRD college undergraduate may see a red bird and a blue bird as two instances of the same category in a color contrast, whereas an individual more familiar with local bird species may consider these to be two instances of different categories (e.g., cardinal and blue jay). This difference in the perception of the environment could have an impact on the way that referential expressions are designed (e.g., "the red bird" vs. "the cardinal") and the inferences that listeners might make (e.g., listeners should be less likely to make contrastive inferences regarding a particular referent if they do not perceive it to be in a contrast with any other candidate referent in the environment).

Further, industrialization tends to accompany the transformation from small-scale, so-called "societies of intimates," to larger-scale "societies of strangers" (Givón, 2005; Trudgill, 2012). Societies of intimates are characterized by small group sizes (50–100 people), foraging-based economies, restricted territorial distribution, and kin-based grouping, *inter alia*, such that individuals in these societies interact primarily with a small number of other people and share much prior knowledge with each conversation partner. Given the tight and essential interconnectedness of intimates, the goals and costs of speech acts are thought to differ from those outlined by Grice to describe the norms of communication of societies of strangers (Givón & Young, 2002). Speakers in these societies often wish to exchange information while attempting to minimize the possibility of alienating any of the other intimates (e.g., by being identified as the source of negative information) by avoiding explicit information. Abiding by Gricean principles such as providing the maximum amount of information and avoiding redundancy would be in conflict with these goals. Such flouting of informativity norms has yet to be explored in the Tsimane' or (to our knowledge) in the context of simple referring expressions in societies of intimates.

Finally, despite the universal presence of turn-taking in conversation, cultures differ quantitatively in terms of the lag they permit between conversational turns (Stivers et al., 2009). Given that listeners anticipate what the speaker will say next in concert with the

[1] A longstanding debate in this literature concerns whether a semantic account, which does not invoke pragmatic reasoning, is sufficient to account for these phenomena. The current study will not attempt to tease the pragmatic and semantic accounts apart and the question of cross-cultural generalizability should be equally interesting to both camps.

[2] In the absence of a visually present object differing primarily in size, the adjective can be interpreted as drawing a comparison to a standard size for that type of object (e.g., big relative to a typical cup) or to a recently seen object of the same type (e.g., big relative to a cup that was recently discussed). However, the relevance of these alternative contrast sets may be reduced in common experimental settings where many objects are divorced from their real-life size (i.e., pictures of objects of all sizes occupy the same space on the screen) and there is no expectation that the speaker (often in the form of pre-recorded instructions) would reference objects from earlier trials, such that the size adjective would provide close to no information about the referent when no size contrast is present in the immediate context.

preparation of their own speech (Brehm & Meyer, 2021), differences in the timing of linguistic exchanges may be reflected in different proclivities for predictive processing during comprehension. Thus, here we investigate the role of Gricean norms in nonindustrialized cultures and, in particular, their real-time dynamics.

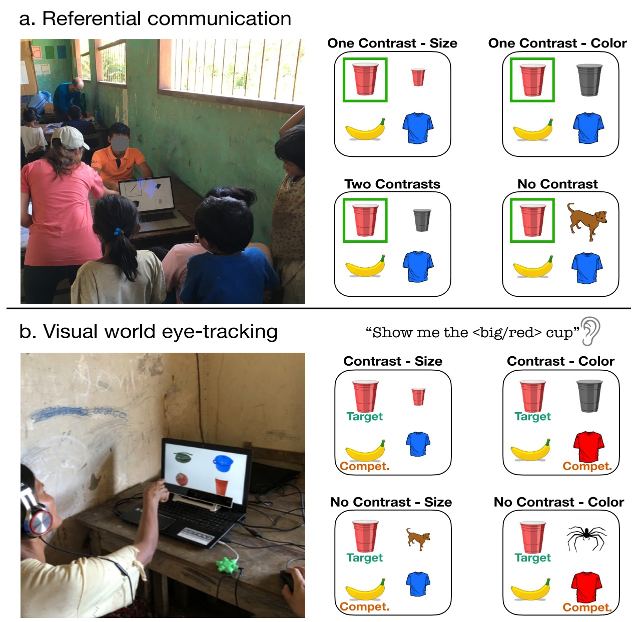
The Present Research

In order to probe the role of culture and industrialization in communication, we study reference production and pragmatic inference in the Tsimane', an indigenous people of Bolivia. The Tsimane' are a nonindustrialized Amazonian group consisting of about 17,000 people from lowland Bolivia who live by farming, hunting, and foraging for subsistence. Most Tsimane' adults have received minimal formal education in reading or writing and Tsimane' children receive much less direct language input than children in WEIRD countries (less than one minute per daylight hour is spent talking to children younger than 4 years of age; Cristia et al., 2017). This is particularly noteworthy for the current purposes because the amount of childhood language exposure and vocabulary knowledge have been related to anticipatory linguistic processing (Borovsky et al., 2012). Tsimane' is one of two languages in a language isolate family (Mosetenan; Sakel, 2004). Unsurprisingly, the language differs in numerous ways from English. Word order appears to be more flexible (e.g., adjectival modifiers can appear before or after nouns) and color words map to the perceptual color space differently than in English (Gibson et al., 2017). For example, the word corresponding most closely to "blue" is used by Tsimane' speakers to refer to a much wider array of colors than those which English speakers traditionally label as blues; the word has higher entropy in Tsimane' and low entropy in English. Additionally, as described in more detail below, Tsimane' speakers were observed to use color adjectives infrequently, in particular, when the referents they modify are natural objects as opposed to artificial (i.e., they are less likely to say "red pepper" than "red shirt"). English speakers use color adjectives quite liberally, even when they are unnecessary for disambiguation.

In the present research, we examined the use and processing of simple, adjective-noun referential expressions (e.g., "a big cup") in Tsimane' and English, varying the visual properties of the context along several dimensions. First, we used a referential communication task (Isaacs & Clark, 1987) to probe the *production* of referential expressions. Participants saw a four-item display and had to communicate to their conversation partner (a translator) which item was in the green square (Figure 1a) while their speech was recorded. To measure the *comprehension* of referential expressions, we used the visual world eye-tracking paradigm (Tanenhaus et al., 1995; Sedivy et al., 1999). Participants listened to referential expressions (e.g., "Show me the big cup") recorded in Tsimane' by a translator (or in English for English speakers) and had to identify which item on their screen the expressions referred to while their eye gaze was recorded (Figure 1b). Second, using these two tasks, we compared the use and interpretation of referring expressions across different visual contexts—critically, the target referent was either a member of a contrast set (e.g., there was a big cup and a small cup in the display) or a singleton (there was only one cup in the display). Third, both in production and comprehension, we compared two types of adjectives—size and color—which are known to elicit different patterns of behavior in English. Production of size adjectives has not been systematically examined in Tsimane' speakers. Fourth,

Figure 1

Experimental Setup and Schematic of Example Trials (a) From Critical Conditions in the Referential Communication Task and (b) From Critical Conditions in the Visual World Eye-Tracking Task



Note. Photos show a typical experimental setup within a one-room schoolhouse, typically the only building in a Tsimane' village. Sources of the images are as follows: red cup (Adapted from Freesvg.org, by OpenClipart 2014, https://freesvg.org/plastic-cup-vector-drawing. In the public domain), gray cup (Adapted from Freesvg.org, by OpenClipart, 2014, https://freesvg.org/plastic-cup-vector-drawing. In the public domain), banana (From Freesvg.org, by OpenClipart, 2016, https://freesvg.org/yellow-banana. In the public domain), red T-shirt (From Freesvg.org, by OpenClipart https://freesvg.org/red-t-shirt. In the public domain), blue T-shirt (From openclipart.org, by Kuba, 2011, https://openclipart.org/detail/118609/azure-tshirt. In the public domain), dog (From cleanpng.com, by Hotep, https://www.cleanpng.com/png-dog-cat-clip-art-brown-dog-pictures-125711/. In the public domain), spider (From Publicdomainpictures.net, by Piotr Siedlecki, https://www.publicdomainpictures.net/en/view-image.php?image=89856&picture=spider, CCO Public Domain), and ear (From openclipart.org, by rematuche, 2017, https://openclipart.org/detail/289462/simple-ear, In the public domain). See the online article for the color version of this figure.

we included both color words that are consistently used by Tsimane' speakers (low entropy: black, red) and color words that are inconsistently used by Tsimane' speakers (high entropy: blue, yellow) (Gibson et al., 2017), as opposed to English speakers who use all of these color words consistently. Finally, we compared two types of referents—natural and artificial—which previous work suggests may lead to differential patterns of color adjective use among Tsimane' speakers (Gibson et al., 2017). We expand on the predicted patterns across all these dimensions below.

In *production*, a rich literature in psycholinguistics attests that English speakers are (a) more likely to produce a modifying adjective (e.g., big) when the referent is in a contrast set compared to when it is not (Pechmann, 1989; Brown-Schmidt & Konopka, 2011; Ryskin et al., 2015, inter alia) and (b) more likely to produce color adjectives redundantly than size adjectives (Sedivy, 2005; Brown-Schmidt & Konopka, 2011; Degen et al., 2020, inter alia). Much less is known about referential communication in Tsimane' but previous work with this group (Gibson et al., 2017) offers some suggestions that they may differ from English speakers in some ways. In a sequential labeling task with natural real-world objects (i.e., they were asked to label, e.g., a green pepper and then a red pepper), Tsimane' speakers rarely used color adjectives for either referent in the sequence (i.e., they would label both peppers using the bare noun), whereas English speakers were more likely to use a color adjective when labeling the second item in a pair (and used more color adjectives overall). It is noteworthy that Tsimane' speakers did use color adjectives contrastively when the referents were artificial, man-made objects (e.g., a shirt) though still much less frequently than English speakers, suggesting that they may view pairs of artificial objects more readily as instances of the same category in a color contrast. Gibson et al. also reported that some color words are used inconsistently (high entropy, e.g., blue and yellow) whereas others are well-known and reliably used by all Tsimane' speakers (low entropy, e.g., black and red). It is unknown to what extent the entropy of the color word might interact with the contrastive use of a color adjective (e.g., redundancy may be absent for high entropy color words but not low entropy color words; on the other hand, high entropy adjectives use may simply be at floor regardless of contrast presence). Thus, in our referential communication task, we probed whether Tsimane' speakers would (a) produce more adjectives when referents are in a contrast set than when they are not, (b) produce more color adjectives than size adjectives redundantly, (c) use color adjectives contrastively more often for artificial referents as compared to natural referents, and (d) use high-entropy color adjectives more contrastively than low-entropy color adjectives or vice versa.

In *comprehension*, eye-tracking studies have shown that English speakers make real-time, pragmatic inferences while interpreting contrastive adjectives (Sedivy et al., 1999; Chambers et al., 2002; Heller et al., 2008; Rubio-Fernandez & Jara-Ettinger, 2020; Ryskin et al., 2015, inter alia). In particular, upon hearing a referring expression containing an adjective (e.g., show me the big…), (a) they are quicker to identify a target referent if it is in a contrast set (which differs along the dimension picked out by the adjective), and (b) this contrastive inference is larger when the adjective is a size adjective than when it is a color adjective. If Gricean assumptions about the speaker's intention to be unambiguous and nonredundant do not generalize to the Tsimane' society of intimates, they may refrain from predictively inferring the nature of the target referent when they hear a modifier. Another important difference between the two languages is that adjectives always appear before the noun in English whereas they can appear before or after the noun in Tsimane'. Contrastive inferences in comprehension could be weaker when word order is freer simply because the listener has less evidence about the statistics of pre-nominal and post-nominal adjectives and their relationship to objects in contrast sets (because there are more options, they experience each a smaller number of times). On the other hand, it might be the case these statistics are aggregated for adjectives regardless of their position relative to the noun, in which case no difference in magnitude would be predicted. Similarly, a contrastive inference may be reduced in Tsimane' when referents are natural because of lack of experience with modification for those objects in daily life,[3] and/or when modifiers are high entropy because the assumption of shared knowledge of the word-meaning mapping may not hold. In addition, to our knowledge, this study is the first to use eye-tracking technology to study cognition in the Tsimane' or any other remote, nonindustrialized group. Thus, in our visual world paradigm task, we tested whether (a) eye-tracking technology is a viable method for studying real-time language processing in fieldwork with nonindustrialized societies, (b) whether Tsimane' participants would engage in anticipatory, contrastive inference rooted in Gricean expectations of informativity, and (c) whether the probability of inference in comprehension would be tied to the probability of an adjective during production (i.e., whether the relative rates of adjective use across conditions in the production study will be reflected in contrast effects on eye movements during comprehension).

Method

All experiment materials, code, and a preregistration for the eye-tracking experiment can be found at https://osf.io/bjs85/.

Participants

In the referential communication task, we collected data from 21 Tsimane' participants.[4] In the eye-tracking task, we collected data from 60 Tsimane' participants and 64 English-speaking participants from the Cambridge, MA area.[5] Each individual participated in only one of the two tasks. Data from an additional 10 Tsimane' participants were collected but never analyzed because of low-quality eye-tracking data (e.g., excessive data loss and poor calibration).[6] All participants were compensated for their participation. All experimental procedures were approved by the Massachusetts Institute of Technology's Committee on the Use of Humans as Experimental Subjects. Informed consent from U.S. participants and assent from Tsimane' participants were obtained, as required by the Committee.

Tsimane' are a forager-horticulturist population inhabiting over 100 villages in the Bolivian lowlands ranging in size between 40 and 550 individuals (Gurven et al., 2007). The Tsimane' are undergoing slow cultural change through contact with Spanish-speaking Bolivians but largely retain a pre-industrial lifestyle (as of data collection in 2018). Villages are typically composed of several extended families. The

[3] Additionally, Tsimane' speakers might be less likely than English speakers to interpret two natural entities (e.g., birds) of different colors as belonging to the same contrast set, as opposed to being different types, but we do not test this question here because the stimuli that differ in color were identical images that were colored differently.

[4] The goal of this task was to establish the presence or absence of contrastive adjective use in the Tsimane' language given previous reports of a conspicuous absence thereof in a less communicative task (Gibson et al., 2017). Contrastive adjective use among English speakers is well-documented. Comparing overall rates of adjective use across the two populations would not be particularly informative for the current question because of the inherent differences between task environments, familiarity with computers/experiments of participants, etc.

[5] In the case of the eye-tracking data, collecting data from an additional sample was more critical for methodological reasons: the eye-tracker being used hadn't, to our knowledge at the time, been used in visual world paradigm studies, and there was very little work using eye-tracking outside the lab environment. Thus, unlike for production data, replicating well-known findings with English speakers was important for interpreting the Tsimane' eye-tracking data.

[6] The relatively high rate of data loss (~14% in Tsimane' participants) is to be expected given the inconsistent lighting and sound conditions typical of fieldwork.

majority of their diet consists of plantains, rice, corn, sweet manioc, and other crops (Kraft et al., 2018). They also regularly fish and hunt. A small proportion of foods is purchased from Bolivian merchants. The villages mostly do not have running water, plumbing, or electricity. Most homes lack walls or doors, and access to healthcare is limited and primarily dependent on the village's proximity to a larger town (Gurven et al., 2020). Some people (typically younger males) travel into town to purchase goods and seek work. This constitutes the primary form of exposure to Spanish.

Tsimane' participants in this study were recruited from six Tsimane' communities near the town of San Borja in the Bolivian Amazon, in collaboration with the Centro Boliviano de Investigación y de Desarrollo Socio Integral (CBIDSI), which provided interpreters, logistical coordination, and expertise in Tsimane' culture. Though no formal language testing was conducted, Tsimane' participants were, generally speaking, not bilingual. Experimenters communicated with participants exclusively through translators who translated from Spanish to Tsimane'. Demographic information collected from the population (including individuals who did not participate in the current experiments) indicated that participants had minimal formal education (mean years of education reported = 3.9, $SD = 4.0$)

Referential Communication Task

Procedure

In the referential communication task, participants sat in front of a computer screen with a four-picture display (see Figure 1a). The conversation partner (a translator) sat on the other side of the table and received the same four images but on pieces of paper that were randomly arranged. On each trial, one of the four pictures on the participant's screen was outlined with a green square, and participants were asked to describe that target image so that their partner could identify which of their four pictures corresponded to it. The participant and the partner could not see each other's pictures but the participants were told that they were the same items. After hearing a description, the partner would show the participant the matching image and, if the participant agreed that it was the right one, they moved on to the next trial (i.e., an experimenter advanced the computer program and distributed the next set of pictures to the partner). If the partner could not tell which picture they were referring to, he asked for clarification. All the participants' productions were audio recorded and transcribed offline and translated into Spanish by a bilingual native Tsimane' speaker. Stimulus presentation and audio recording were controlled using MATLAB and Psychtoolbox (Brainard, 1997). The entire session lasted around 15 min.

Materials and Design

Images were photos collected from a previous eye-tracking experiment stimulus set (Ryskin et al., 2019) and supplemented with a Google image search for local objects and animals. Before data collection began (in Tsimane'), the names of all the objects in the images were obtained through discussion with the translators.

The experimental design manipulated (a) contrast presence (three levels: one-dimension contrast vs. two-dimension contrast vs. no contrast) and (b) type (five levels: size, color-low-entropy-natural, color-low-entropy-artificial, color-high-entropy-natural, and color-high-entropy-artificial). In one-dimension contrast condition trials, the target item (in the green square) was a member of a pair that differed in either size or color (Figure 1a). The other two objects on the screen were distractors. In two-dimension contrast condition trials, the target item was a member of a pair that differed in both size and color. In no contrast condition trials, the target item was a singleton and the size contrast item was replaced by another distractor. The type factor varied the kind of adjective (size vs. two types of color adjectives: low-entropy and high-entropy) and the type of object that was referred to (artificial vs. natural). "Low-entropy" color terms—jäibäs (red) and tsincus (black)—are used consistently and reliably by Tsimane' speakers and "high-entropy" colors—chames/yellow and yushñyus/blue—are used inconsistently (Gibson et al., 2017). Natural objects (e.g., plants and animals) are thought to elicit fewer color adjectives than artificial objects (e.g., dishware and tools). In filler trials, the target items did not differ from the distractor in size or color (two or three of the other pictures were of the same color as the target).

The experiment was made up of 69 trials in total: 51 critical trials (5×3 size contrast conditions + 3×3 color contrast conditions $\times$ 4 types), 15 filler trials, and 3 practice trials. Participants were randomly assigned to one of four pseudo-randomized orders which counterbalanced target items across contrast conditions.

Eye-Tracking

Procedure

The eye-tracking task was modeled on Sedivy et al. (1999). Participants sat in front of a computer (powered from a portable car battery as the school houses in which data collection was conducted were not wired for electricity) and wore headphones (see Figure 1b). On each trial, participants saw four pictures (one in each quadrant of the display) and heard a pre-recorded audio file containing a referring expression (e.g., "show me the big cup").[7] Their task was to point to the referenced (target) object (the big cup). An experimenter used a mouse to click on the picture that the participant pointed to. The locations of the target, competitor (e.g., big banana), size contrast (e.g., small cup), and distractor (e.g., small shirt) were randomly determined for each trial. A Gazepoint 3-HD eye-tracker was positioned under the computer screen and, after a 9-point calibration sequence, continuously recorded eye-gaze coordinates at 120 Hz. Stimulus presentation and event timestamps were controlled and recorded using MATLAB and Psychtoolbox (Brainard, 1997). The entire session lasted around 40 min with a 10 min break in the middle.

Materials and Design

The experimental design consisted of two crossed factors: contrast presence (two levels: contrast vs. no contrast) and type (five levels: size, color-low-entropy-natural, color-low-entropy-artificial, color-high-entropy-natural, and color-high-entropy-artificial). In contrast condition trials, the target item was a member of a pair

[7] In both English and Tsimane', the auditory instructions contained pre-nominal adjectives only. For English speakers, adjective ordering is fixed and adjectives are always pre-nominal. For Tsimane' speakers, word order is freer, however, in order to measure anticipatory contrastive inference, the adjective has to be pre-nominal: if they hear the noun first, there is nothing for listeners to anticipate. We leave it to future work to investigate how word order flexibility may affect contrastive inference

that differed in either size or color (see Figure 1b for an example with size). The competitor was a picture of a singleton object of the same size or color as the target. In no-contrast condition trials, the target item was a singleton and the size contrast item was replaced by another distractor (see Figure 1b). As in the referential communication task, the type factor varied the kind of adjective (size vs. two types of color adjectives: low-entropy [red or black] and high-entropy [blue or yellow]) and the kind of object that was referred to (artificial vs. natural). During filler trials, the target was a singleton and the auditory stimulus did not contain an adjective (e.g., "Ikaviete sapyeyes"/"Show me the rope"). Half of the filler trials contained a pair (e.g., two pencils) among the distractor items.

Images consisted of a subset of the pictures used for the referential communication task. Tsimane' auditory stimuli were recorded at the fieldwork site in Bolivia in a quiet outdoor location on a laptop. One translator provided labels for images that were shown to him. English auditory stimuli were recorded on the same laptop in a quiet indoor setting (by RR). The speakers were instructed to speak slightly more slowly than they would in a typical conversation and to maintain neutral prosody to the extent possible. Each instruction was recorded twice and the more natural/neutral sounding of the two recordings was chosen. Due to the practicalities of fieldwork, minimal post-processing was applied to the audio files. For consistency, the same approach was taken with the English files. The timestamps of word onsets were obtained using the Montreal Forced Aligner (McAuliffe et al., 2017) and extracted via Praat script (Boersma & Weenink, 2022). Additional details of the audio files are available in the supplemental material S1 on OSF (https://osf.io/bjs85/). There was some variation in the prosodic properties of the auditory stimuli between Tsimane' and English.[8] In the adjective window, which was the focus of analyses, the duration was similar across English (mean = 0.84 s, $SD = 0.16$) and Tsimane' (mean = 0.81 s, $SD = 0.13$), but the intensity (English: mean = 62.15 decibels, $SD = 4.01$, Tsimane': mean = 57.49 decibels, $SD = 1.57$) and average pitch (English: mean = 222.71 Hz, $SD = 23.14$, Tsimane': mean = 145.19 Hz, $SD = 19.73$) differed, likely in part because the Tsimane' speaker was male and the English speaker was female. Item-specific analyses indicate that the variation in prosodic properties does not appear to impact the primary effects of interest (see supplemental material S1 at https://osf.io/bjs85/) but caution is warranted in interpreting baseline differences across languages and trial types. We return to this point in the discussion.

The experiment consisted of 120 trials in total: 100 critical trials (10×2 contrast conditions $\times$ 5 types) and 20 filler trials. Participants were randomly assigned to one of two counterbalancing lists. A given target item would appear in both the contrast and no-contrast condition across subjects, paired with the same auditory instruction in both cases. Except for the first three trials which were always filler trials, the order of trials was randomly generated for each participant. The trials were separated into two blocks of 60 trials each to give participants a 10 min break in the middle.

Data Preprocessing

Timestamps of eye-gaze coordinates were aligned to onsets of critical words in the audio stimulus files and categorized according to which of the four quadrants they landed in. Data from incorrect trials (i.e., where the participant selected the wrong target) were excluded, as well as data from participants who had an overall accuracy below 80%.[9] If a participant had $< 50\%$ valid samples or more than 50% of stimulus messages (e.g., a timestamp for stimulus onset) missing within a block due to packet loss between the eye-tracker hardware and the experiment computer, the block was excluded.

Results

All statistical analyses were performed using the MixedModels package (Bates et al., 2020) in Julia (Bezanson et al., 2017). Visualizations were created with the ggplot2 library (Wickham, 2016) in R (R Core Team, 2019). Analysis code and data are available at https://osf.io/bjs85/.

Referential Communication Task

The transcribed and translated utterances were coded for whether they contained a size or color adjective. "Other-initiated repairs" occurred when the speaker produced an utterance that did not uniquely identify the target and the listener had to ask for clarification. These were not counted as utterances containing an adjective even if the speaker then produced an appropriate adjective and communication was successful. In some cases, participants used an adjective that did not describe the size or the color to identify the target (e.g., "ripe" when referring to the red tomato) but still achieved referential success without repair. These cases were also coded as containing zero size or color adjectives. The rates of adjective use across size and color adjective conditions are summarized in Figure 2. The data were analyzed in two mixed-effects logistic regression models—one for size adjective use (including only data from the size trials) and one for color adjective use (including only data from the color trials)—with the presence or absence of an adjective during the trial as the dependent variable. The fixed portion of the models consisted of a dummy-coded, three-level predictor representing the contrast condition (no contrast vs. two contrasts [size and color] vs. contrast [size or color]), with the no contrast condition as the reference level. Due to the small size of the dataset, only random intercepts for participants and items were included. See Table A1 for full model details.

Overall, participants produced more color adjectives than size adjectives in all conditions (see Figure 2a). Participants were more likely to use a size adjective when the target referent was in a size contrast with another object in the display (two contrasts vs. no contrast: $b = 1.54$, $SE = 0.60$, $p = .01$; contrast vs. no contrast: $b = 2.16$, $SE = 0.58$, $p < .001$), and they were (slightly) more likely to use a color adjective when the target referent was in a color contrast with another object in the display (contrast vs. no contrast: $b = 0.76$, $SE = 0.37$, $p = .04$), though this was only a numerical tendency when the contrasting object also differed in size (two contrasts vs. no contrast: $b = 0.65$, $SE = 0.37$, $p = .08$).

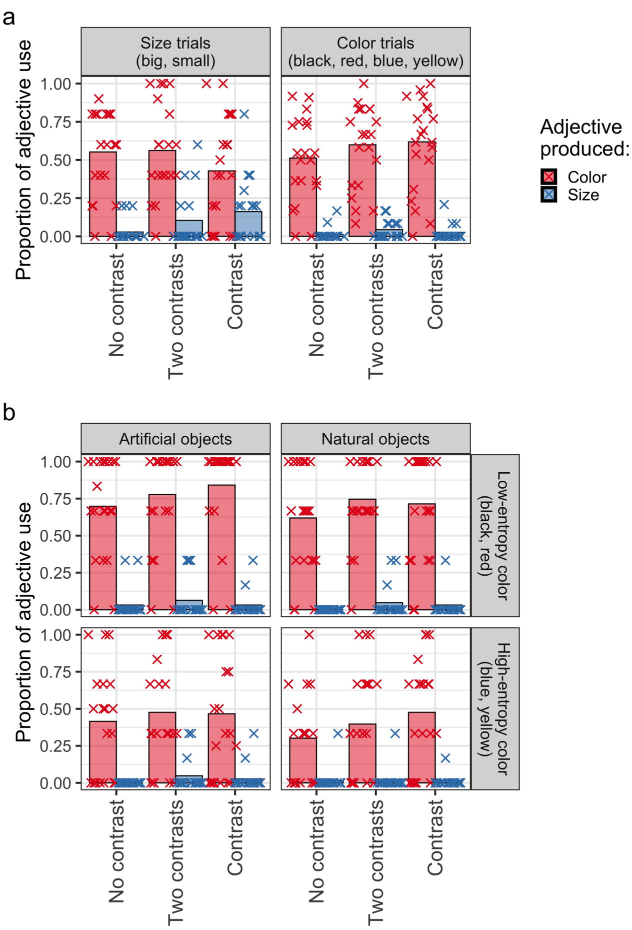
An additional analysis compared contrasts effects across types of color adjective trials, as shown in Figure 2b (dummy coded with low-

[8] To some extent, this is to be expected since the auditory stimuli consist of different words spoken by different speakers. Note that the same audio files were used for contrast and no contrast conditions for a given item so between-item differences are largely orthogonal to the primary comparison of interest. Moreover, all analyses include random intercepts and slopes in order to account for variation between items.

[9] Accuracy for remaining participants was 97% on average for Tsimane' and 99% on average for English participants. Participants excluded due to accuracy are included in the set of 10 excluded (Tsimane') participants.

Figure 2

Summary of Adjective Production Rates Across Conditions in the Referential Communication Task



Note. Crosses indicate individual (Tsimane') participant averages by condition. Bar heights indicate means across all participants. (a) The plot on the left indicates data from trials where size was the relevant dimension by which a contrast could be established (and the corresponding control trials). The plot on the right indicates data from trials where color was the relevant dimension by which a contrast could be established (and the corresponding control trials). Participants were more likely to produce color adjectives than size adjectives in all conditions, and the presence of a size or color contrast increased the likelihood of producing size or color adjectives respectively. (b) Summary of color adjective production rates by referent type (artificial vs. natural) and color adjective type (low-entropy vs. high-entropy). Participants were more likely to produce low-entropy color adjectives than high-entropy color adjectives, and the presence of a color contrast increased the likelihood of producing a color adjective for all referent and color adjective types. See the online article for the color version of this figure.

entropy artificial as the reference level). In the no contrast condition, there was no significant difference between natural and artificial objects with low-entropy colors ($b = -0.52$, $SE = 0.51$, $p = .30$), but participants produced fewer high-entropy color adjectives (natural: $b = -2.55$, $SE = 0.51$, $p < .001$; artificial: $b = -1.84$, $SE = 0.53$, $p < .005$). The trial type did not significantly interact with the effects of contrast presence ($ps > .15$).

Eye-Tracking

The time course of eye-gaze to the target and competitor for size and color adjectives by contrast condition and language is shown in Figure 3a. The "target advantage" (Figure 3b and c) corresponds to the difference between the average proportion of looks to the target and the average proportion of looks to the competitor. As is typically found in visual world paradigm studies, as the auditory instruction unfolded (e.g., "Show me the big cup"), looks to the target consistently increased and looks to the competitor decreased (after a brief initial increase) indicating that participants rapidly identified the referent (e.g., the cup). For both English and Tsimane' participants, this identification appears to have occurred earlier when the target was in a contrast set (e.g., big cup and small cup); looks to the target exceed looks to the competitor sooner in the contrast condition than the no contrast condition. This pattern is particularly noticeable for size adjectives compared to color adjectives.

The eye-gaze data were analyzed in two complementary ways. The first approach used autoregressive logistic mixed effects models (generalized linear mixed models [GLMMs]; Cho et al., 2018). The dependent variable was the presence or absence of a look at the target within each 10 ms bin in an 850 ms time window (at the trial level for each participant). The time window duration corresponds to the average duration between the onset of the adjective and the onset of the noun and was time-locked to the onset of the adjective (for each trial) and offset by 200 ms to account for the oculomotor delay. Looks to the competitor were not analyzed in these models because they are not independent of looks to the target. The second analysis approach used linear mixed effects models to model the target advantage averaged over the adjective window—how much more participants look to the target relative to the competitor. The time window duration was determined by the onsets of the adjective and noun of the auditory stimulus (starting 200 ms after the onset of the adjective and ending 200 ms after the onset of the noun) and differed on a trial-by-trial basis. While the former analysis approach more closely reflects the generative process for the data and is more sensitive, the latter analysis allows for consideration of both target and competitor looks. Both analyses lead to similar conclusions regarding contrastive inferences by Tsimane' and English speakers.

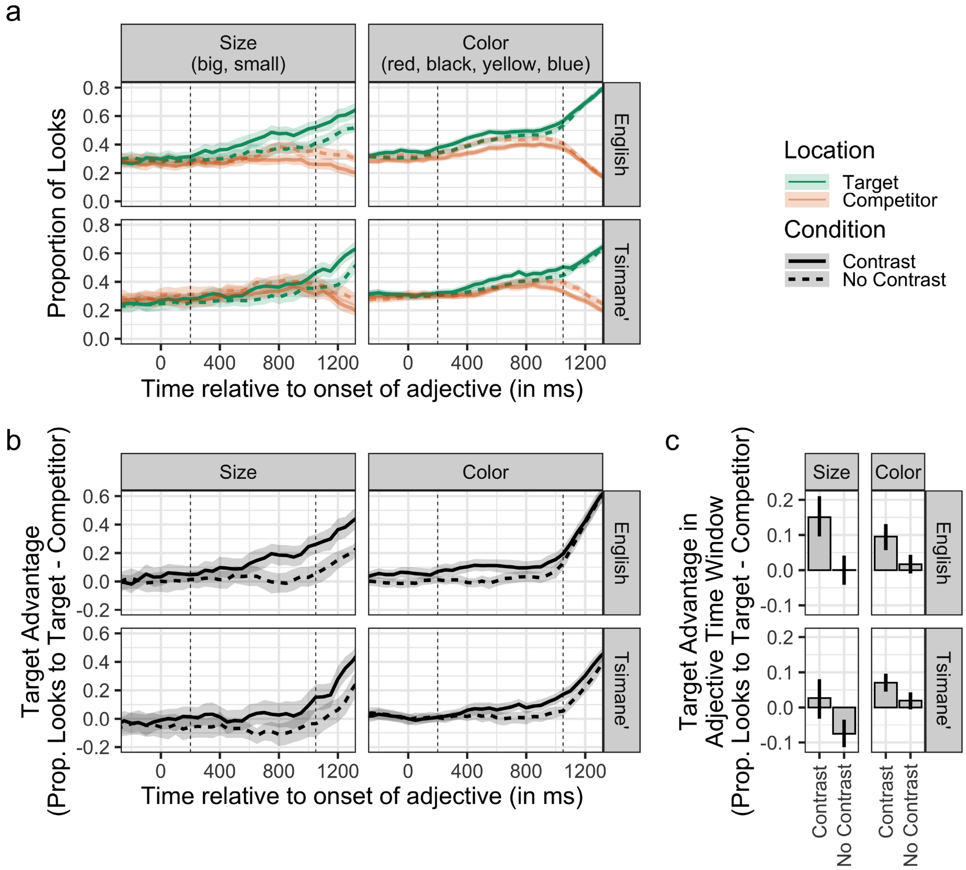
Contrastive Interpretations for Size Versus Color

Autoregressive GLMM of Target Looks. In order to determine the effects of contrast condition and adjective type (size vs. color) on *binary target looks*, for both English and Tsimane' participants (Figure 3a), a GLMM model was fitted with effects-coded fixed effects for contrast condition (no contrast = −1, contrast = 1), adjective type (size = −1, color = 1), language (English = −1, Tsimane = 1), and all two-way and three-way interactions. An AR(1) predictor was also included to account for the autocorrelation (of lag 1) of eye gaze over time. Following (Cho et al., 2018), the random effects structure was chosen by model comparison (Table A2). The selected model included random intercepts for trial order with random slopes for AR(1), random intercepts for subjects with random slopes for contrast, adjective type, and AR(1), and random intercepts for items with random slopes for contrast and AR(1). See Table A3 for full model details.

Participants were more likely to look at the target when it was in a size contrast set than when it was not ($b = 0.08$, $SE = 0.01$, $p < .001$). They made more looks to the target overall on color

Figure 3

Gaze Data by Condition and Language in the Visual World Eye-Tracking Task, With Critical Auditory Instructions of the Form "Show me the ?adjective? ?noun?."



Note. (a) Timecourse of eye gaze to target and competitor locations during size and color adjective trials (all color adjective trials combined). Dashed vertical lines indicate the time window used in the autoregressive GLMM analysis: starting 200 ms after the onset of the adjective and ending 200 ms after the average onset of the noun. (b) The time course of target advantage (eye gaze to competitor subtracted from eye gaze to target) by condition. (c) Target advantage averaged over the adjective window (starting 200 ms after the onset of the adjective and ending 200 ms after the onset of the noun) by condition and population. (Shaded regions and error bars indicate bootstrapped 95% confidence intervals over participant means). Both Tsimane' and English-speaking participants were more likely to look at the target (and less likely to look at the competitor) when the target object was in a contrast set (e.g., a big cup and a small cup) than when it wasn't. This effect was larger for size adjectives than color adjectives. See the online article for the color version of this figure.

adjective trials ($b = 0.10$, $SE = 0.02$, $p < .001$). Tsimane' participants made fewer looks at the target overall ($b = -0.07$, $SE = 0.02$, $p < .001$). Critically, the contrast effect was reduced on color trials relative to size adjective trials (Interaction: $b = -0.03$, $SE = 0.01$, $p = .01$). Neither the contrast effect nor the main effect of adjective type, nor the interaction of contrast and adjective type differed across language groups[10] (Contrast × Language interaction: $b = 0.00$, $SE = 0.01$, $p = .74$; Adjective Type × Language interaction: $b = 0.01$, $SE = 0.02$, $p = 0.75$; Contrast × Adjective Type × Language interaction: $b = 0.00$, $SE = 0.01$, $p = .70$).

LMM of Average Target Advantage. In order to determine the effects of contrast condition and adjective type (size vs. color) on *target advantage* (average proportion of looks to target during the adjective window minus the average proportion of looks to the competitor in the adjective window), for both English and Tsimane' participants (Figure 3c), a model was fitted with effects-coded fixed effects for contrast condition (no contrast = −1, contrast = 1), adjective type (size = −1, color = 1), language (English = −1, Tsimane = 1), and all two-way and three-way interactions. The

[10] Refitting the model with the English group or the Tsimane' group as the reference level indicates that contrast effects (English: $b = 0.07$, $SE = 0.02$, $p < .001$; Tsimane': $b = 0.08$, $SE = 0.02$, $p < .005$) and adjective type effects (English: $b = 0.09$, $SE = 0.03$, $p < .001$; Tsimane': $b = 0.10$, $SE = 0.03$, $p < .001$) are significant for both language groups but the interaction of contrast and adjective type only reaches significance in the English group (English: $b = -0.03$, $SE = 0.02$, $p < .05$; Tsimane': $b = -0.03$, $SE = 0.02$, $p = .14$). However, this difference in the magnitude of interaction effects should not be over-interpreted given that the three-way interaction is not significant.

maximal random effects structure given the design of the study was used. This included random intercepts for trial order, random intercepts for subjects with random slopes for contrast, adjective type, and random intercepts for items with random slopes for contrast. See Table A4 for full model details.

Participants made more looks to the target than the competitor overall (*Intercept* = 0.04, *SE* = 0.01, $p < .005$), indicating that they typically began to identify the target before the onset of the noun. This target advantage was greater when the target was in a contrast ($b = 0.04$, $SE = 0.01$, $p < .001$) but did not differ significantly depending on the adjective type ($b = 0.01$, $SE = 0.01$, $p = .27$). The overall target advantage was reduced for Tsimane' speakers relative to English speakers ($b = -0.03$, $SE = 0.01$, $p < .05$). Crucially, the contrast effect did not differ significantly across languages (Contrast × Language interaction: $b = -0.01$, $SE = 0.01$, $p = .20$) or adjective types (Contrast × Adjective Type: $b = -0.01$, $SE = 0.01$, $p = .09$) and the three-way interaction of contrast, adjective type, and language was not significant ($b = 0.004$, $SE = 0.01$, $p = .63$).[11]

Contrastive Interpretations Across Color Adjectives

The time course of the proportion of looks to the target and competitor for the four kinds of color adjectives by contrast condition and the population is shown in Figure 4a. There are subtle differences in how target identification unfolds within the critical time window across the four types of color adjective trials (e.g., there seems to be an early preference for the target in the low-entropy natural condition, regardless of condition),[12] The contrast condition appears to have an effect across adjective and referent types (perhaps with the exception of low-entropy natural for Tsimane').

Autoregressive GLMM of Target Looks. A model predicting *binary target looks* was fitted with fixed effects for contrast condition (effects-coded: no contrast = −1, contrast = 1), color adjective type (dummy coded with low-entropy-artificial trials as the reference level), language (English = −1, Tsimane = 1), and all two-way and three-way interactions. An AR(1) predictor was also included to account for the autocorrelation of eye gaze over time. Following (Cho et al., 2018), the random effects structure was chosen by model comparison (Table A2). The selected model included random intercepts for trial order with random slopes for AR(1), random intercepts for subjects with random slopes for contrast, adjective type, and AR(1), and random intercepts for items with random slopes for contrast and AR(1). See Table A5 for full model details.

Participants in the low-entropy color artificial trials were more likely to look at the target when it was in a color contrast setting than when it was not ($b = 0.05$, $SE = 0.02$, $p < .05$). They made more looks to the target overall in low-entropy natural trials ($b = 0.12$, $SE = 0.04$, $p < .001$). Tsimane' participants made fewer fixations to target objects relative to English speakers ($b = -0.06$, $SE = 0.04$, $p < .05$). However, the contrast effect did not interact significantly with color adjective type or language (*ps* > .50).

LMM of Average Target Advantage. In order to determine the effects of contrast condition on *target advantage* for the four kinds of color adjectives, for both English and Tsimane' participants (Figure 4c), a model was fitted with effects-coded fixed effects for contrast conditions (no contrast = −1, contrast = 1), color adjective type (dummy coded with low-entropy-artificial trials as the reference level), language (English = −1, Tsimane = 1), and all two-way and three-way interactions. The maximal random effects structure given the design of the study was used. This included random intercepts for trial order, random intercepts for subjects with random slopes for contrast, adjective type, and their interaction, and random intercepts for items with random slopes for contrast. See Table A6 for full model details.

In the low-entropy color artificial trials (reference), the target advantage was greater when the target was in a contrast set ($b = 0.04$, $SE = 0.02$, $p < .01$). The target advantage was also greater in high-entropy natural ($b = 0.07$, $SE = 0.03$, $p < .05$) and low-entropy natural trials ($b = 0.11$, $SE = 0.03$, $p < .001$), but the effect of contrast did not differ significantly across color adjective types or languages (for all two-way and three-way interactions, *ps* > .34).

Discussion

We investigated whether aspects of the way that humans design and interpret communicative signals (referential expressions) are shared between English and Tsimane'.

Production of Referential Expressions

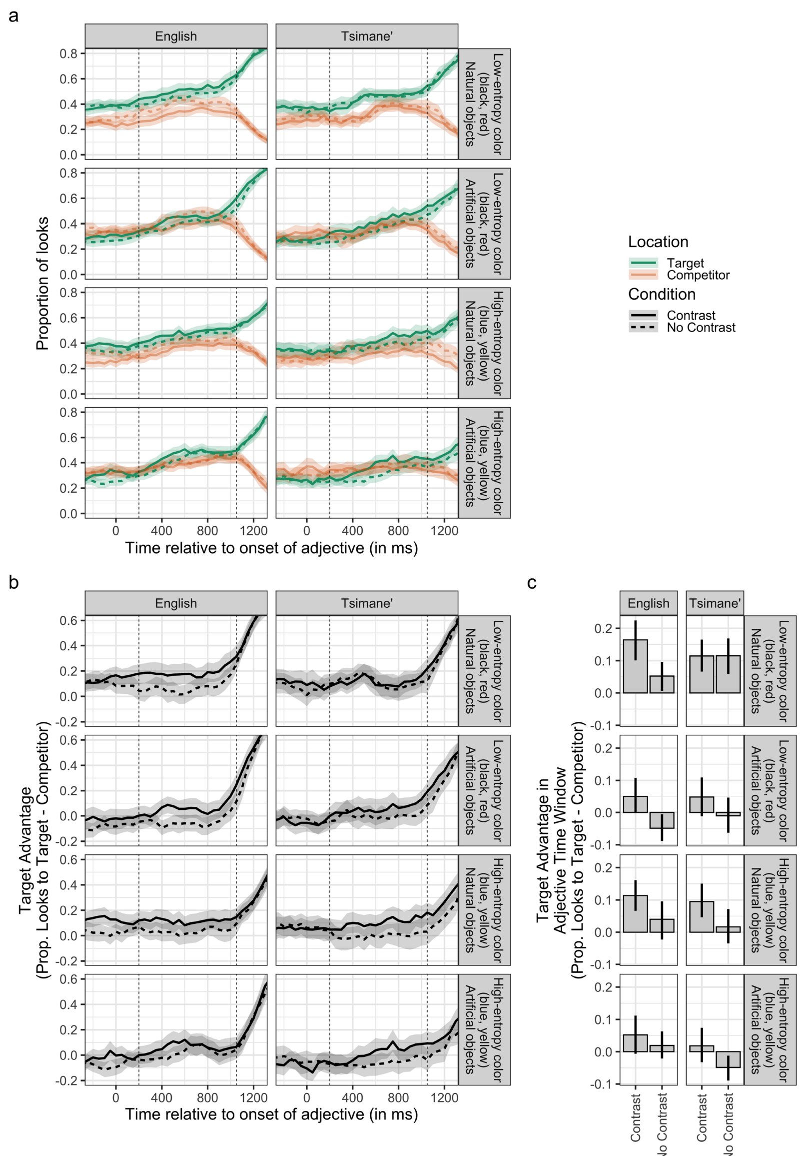
In a referential communication task, Tsimane' speakers produced higher rates of color adjectives than size adjectives. Qualitatively, this pattern is similar to what has been observed with English speakers (e.g., Brown-Schmidt & Konopka, 2011). But the rates of both adjective types were overall lower compared to English speakers (e.g., under 25% for size adjectives in the current data set and around 50% in Brown-Schmidt & Konopka, 2011). This pattern is consistent with previous findings of reduced rates of adjective use in Tsimane' (Gibson et al., 2017). These population differences may derived from a number of factors. The current task is not identical to those used in previous studies with English speakers, though modification rates tend to be fairly invariant to small differences in materials and experimental design (e.g., similar size modification rates in Sedivy, 2005; Ryskin et al., 2015). Color and size adjectives may be less frequently used in the Tsimane' language as a whole—and therefore come to mind less readily for participants in the current experiment —perhaps because it is less common for Tsimane' speakers to find themselves in a situation where two objects are most efficiently differentiated via an adjective of this kind. In everyday communication, there are often many other ways to differentiate two objects, for instance, using gestures (e.g., pointing; Holler & Stevens, 2007), demonstratives, and/or spatial terms (e.g., "that cup next to you"), descriptors related to the speaker's goals (e.g., "the banana that looks tasty") or relying on shared common knowledge (e.g., "the cup you just washed"). The relative share of each type of referential

[11] Refitting the model with a dummy-coded adjective type predictor with size as the reference level indicates that the target advantage during size adjective trials was greater when targets were in a contrast ($b = 0.06$, $SE = 0.01$, $p < .001$) and smaller for Tsimane' speakers than English speakers ($b = -0.05$, $SE = 0.02$, $p < .05$). However, the size of the contrast effect was not significantly different between languages ($b = -0.01$, $SE = 0.01$, $p = .29$)

[12] We interpret this as noise related to some uncontrolled properties of the materials (e.g., particularly salient visual or auditory properties of certain targets). These item effects appear to affect speakers of both languages similarly, suggesting that they may be related to visual features which are shared across participants from both groups. See S1 in the supplemental materials, which are available at https://osf.io/bjs85/ for analysis of auditory features of stimuli.

Figure 4

Gaze Data by Condition and Language in the Visual World Eye-Tracking Task for Different Color Adjective Trial Types, With Auditory Instructions of the Form "Show me the ⟨color adjective⟩ ⟨noun⟩."



Note. (a) The time course of eye gaze to target and competitor locations during color adjective trials (separated by the entropy of color adjectives and the nature of the object they modified). Dashed vertical lines indicate the time window used in the autoregressive GLMM analysis: starting 200 ms after the onset of the adjective and ending 200 ms after the average onset of the noun. (b) The time course of target advantage (eye gaze to competitor subtracted from eye gaze to target) by condition. (c) Target advantage averaged over the adjective window (adjective onset +200 ms to noun onset +200 ms) by condition and population. (Shaded regions and error bars indicate bootstrapped 95% confidence intervals over participant means). Both Tsimane' and English-speaking participants were more likely to look at the target (and less likely to look at the competitor) when the target object was in a contrast set (e.g., a red cup and a black cup) than when it was not. The low-entropy natural object trials appear to be an exception for Tsimane' speakers, but analyses suggest that the contrast effect was not significantly smaller than for low-entropy artificial trials. (Note that color type, low vs. high entropy, is a distinction based on the properties of Tsimane' and is less pertinent to English speakers who use all four color adjectives with high consistency). See the online article for the color version of this figure.

expression in usage may differ across languages. We can speculate that such differences would depend on the distributions of objects that are used by the speakers of the language (e.g., whether they often talk about similar objects differing primarily in size/color), how they group objects into categories (e.g., two birds of different colors may belong to different categories depending on the speaker's knowledge), and the constraints placed on the speakers. For instance, strangers are less likely to share common experiences and therefore cannot rely on that to achieve efficient reference; a language spoken primarily between individuals who share many experiences—intimates—may rely heavily on this shared knowledge and make less use of other types of referential expressions. An alternative explanation may be that Tsimane' speakers may simply be less accustomed to engaging in referential communication tasks. Though the goal of these tasks is to be as ecologically valid as possible for an experimentally controlled production task (Krauss & Weinheimer, 1964, 1966), they are still inspired by activities that are common in WEIRD cultures but may not be familiar to Tsimane' speakers causing them to alter their language usage in some way.[13]

Given the difficulties in interpreting overall group differences between English and Tsimane' speakers, the focus of the current referential study was on whether the production of referring expressions was affected by the presence of a visual contrast. We found that Tsimane' speakers were more likely to include size or color adjectives when the target referents were in a contrast set than when they were singletons, similar to previous findings in English speakers. In this task, Tsimane' speakers designed their utterances according to Gricean principles—including additional information (the adjective) more so when it was necessary for their message to be unambiguous for their conversation partner. There was no evidence that this Gricean pattern differed across adjective types, despite the fact that high-entropy adjectives were less likely to be produced than low-entropy adjectives, replicating prior work (Gibson et al., 2017). In contrast to previous findings, there was no evidence that Tsimane' speakers produced more adjectives when describing artificial objects than natural objects in the current dataset. This may be related to task differences between the referential communication paradigm used here and the object labeling task used by Gibson et al. (2017). For instance, the present task always displayed four objects of different colors and sizes simultaneously, which may have encouraged comparison along the color dimension even when no object of the same category was present. Additionally, the present task was computerized, whereas the previous work used real-world objects; by using photos on a computer display, the present paradigm may have deemphasized the distinction between natural and artificial objects.

Real-Time Inference During Comprehension

In the visual-world paradigm eye-tracking task, Tsimane' and English listeners' gaze reflected the integration of linguistic and visual information with inferences about the intentions of the speaker in real time. Upon hearing a size or color adjective, both English and Tsimane' participants were more likely to make anticipatory looks at the target (and fewer to the competitor) when it was in a contrast set than when it was not, replicating prior findings with speakers of English, as well as a set of other typologically diverse languages (Ryskin et al., 2019; Sedivy et al., 1999; Rubio-Fernandez & Jara-Ettinger, 2020). The magnitude of contrast effects did not appear to differ significantly across cultures. Both groups likely make inferences based on the assumption that a cooperative speaker would not have used an adjective if the object were a singleton. Critically, this inference took place in real-time and drove predictive eye movements within the first second of hearing the onset of the adjective.[14] Despite the fact that in many societies of intimates, like the Tsimane', the communicative goals have been reported to deviate from those laid out by Grice, the inferences that Tsimane' listeners drew during the interpretation of referential expressions were in line with Gricean norms of informativity. Moreover, they were consistent with the patterns of use measured in the production task: Tsimane' speakers were more likely to use an adjective if the target was in a contrast set.

Tsimane' and English speakers drew these inferences when listening to both size and color adjectives. Previous results in the literature were equivocal with respect to the presence of contrastive inferences for color (Sedivy, 2005; Aparicio et al., 2016) in English. Plausibly, these discrepant findings may reflect the fact that the effect of contrast presence on the production of a color adjective, and in turn on the probability of adjective-driven inference during comprehension, is tied to the relationship between the object and the color (e.g., whether it is typical; Degen et al., 2020; Sedivy, 2003; Kreiss & Degen, 2020).[15] It is noteworthy that the analysis of binary looks to target (only) suggests that, for both English and Tsimane' listeners, contrastive inferences elicited by color adjectives were reduced relative to those elicited by size adjectives. A simple explanation could be that there were simply more color trials and both groups habituated to the color adjectives in some way that reduced their tendency to draw contrastive inferences, or that uncontrolled visual or acoustic features of the stimuli differed between size and color trials and made contrasts less salient in the latter. (Item-level analysis in S1 in the online supplemental materials suggests that acoustic/prosodic features are unlikely to have had a substantial impact.) Alternatively, this may point to a shared expectation that color adjectives are, in general, less predictive of a contrast than size adjectives (Degen et al., 2020; Sedivy, 2003). The latter explanation would be consistent with the observation that, in contrast to

[13] Indeed, earlier pilot work used a more complex referential communication task, where speakers were asked to describe four images on cards in the order in which they were laid out so that a listener sitting on the other side of a visual barrier with the same set of images could arrange them to be in the same order, suggested that the Tsimane' speakers found the procedure unnatural. The task used in the current study appeared more approachable to participants but it is possible that some task demands remain.

[14] It is worth noting that Tsimane' speakers were less likely to look at the target and their target advantage was lower overall. This baseline difference might reflect any of a very large number of factors that differ between English and Tsimane'-speaking participants in this study including experience with participating in an experiment, comfort with computers, noise in the environment during data collection, acoustic properties of the stimuli which were recorded by different speakers using a different language (see S1 in the online supplemental materials, which are available at https://osf.io/bjs85/) familiarity with images in the visual display, etc. For this reason, conclusions regarding these baseline differences are fraught and the focus of the current investigation is on the effect of contrast presence and whether it holds across languages/populations and/or adjective types. Of course, baseline differences can create spurious interactions (e.g., when there are floor or ceiling effects) but we have no reason to believe that this is an issue in the current dataset.

[15] This factor was not explicitly controlled in the design of the present materials but stimuli were generally chosen to have plausible colors (e.g., peppers could be red or green but not blue).

previous evidence of lower use of color words in an object labeling task (Gibson et al., 2017), Tsimane' speakers were similar to English speakers in using color adjectives more often than size adjectives (Pechmann, 1989) in the referential communication paradigm. However, the analysis of target advantage does not strongly support the conclusion that color adjectives elicit reduced inference relative to size adjectives. Once both looks to target and competitor are taken into account, the target advantage is only numerically smaller for color adjectives across languages. Thus, we can conclude that both Tsimane' and English listeners do make contrastive inferences when interpreting color adjectives in the current dataset—as expected given that both groups produce more color adjectives when a contrast set is present—but how this inference compares to inferences from size adjectives remains an open question.

Finally, for both English and Tsimane' listeners, the magnitude of the contrastive inference did not differ significantly across the types of color adjective trials (low- vs. high-entropy, natural vs. artificial). Despite potential differences in how Tsimane' might construe contrasts, the frequency with which they encounter modification in everyday life, and the uncertainty regarding the mapping between words and meanings, the patterns of inference were (qualitatively) consistent with the patterns of production (i.e., color entropy and referent nature did not appear to impact contrast effects in production).

Taken together, these findings suggest that, if they are present in Tsimane', "non-Gricean" norms may be primarily applicable to more complex conversational settings where interpersonal relationships are at stake.[16] In the kinds of simple referential exchanges that form the building blocks of conversation, inferences drawn during comprehension appear to be tightly (rationally) tied to the statistics of use given the specific context, across cultures.

Implications and Conclusions

In sum, across a variety of contexts, the patterns of eye-gaze in Tsimane' listeners point to inferences about the speaker's intent that are tied to how adjectives are used to disambiguate referents during production. Moreover, these patterns appear to be similar to those observed in the productions and eye movements of English speakers. In line with prior proposals (Levinson, 2011), the current work suggests that pragmatic aspects of language use are shared across languages and cultures. This finding stands in contrast to the many phonological, syntactic, and semantic differences between English and Tsimane' (Sakel, 2004; Gibson et al., 2017) and the great diversity of these features in the world's languages more broadly (Evans & Levinson, 2009). Notably, word order is freer in Tsimane' than in English, such that prenominal modifiers are less frequent in Tsimane' than in English (relative to all modifiers). This does not appear to impact the magnitude of contrastive inference (at least not to an extent that we can detect), suggesting that the statistics of use, which seem to be tied to probabilistic inferences during comprehension, may be tracked in aggregate, regardless of the position of the modifier. Future work looking at many languages with different levels of flexibility in word order, or potentially using artificial language learning, may shed light on this speculation.

Given the importance of culture in shaping language (Boas, 1938; Evans, 2003; Everett, 2012; Tomasello, 1999; Thompson et al., 2020), it is perhaps not surprising that aspects of language which are formed by the accumulation of conventions (Haspelmath, 1999) differ between "WEIRD" English language users and Tsimane' who live in a nonindustrialized "society of intimates" and have received little formal education. However, the production and interpretation of reference are less conventionalized—they require in-the-moment inferences based on the visuospatial context and social reasoning about the knowledge and communicative goals of the conversation partner—which may explain why the same principles appear to govern its use across cultures.

However, other aspects of communication are likely to be more culturally dependent. For example, politeness expressions differ even among otherwise fairly similar cultures (e.g., the appropriateness of direct requests differs between English/German and Russian/Polish speakers Ogiermann, 2009). These aspects of pragmatics appear quite different in nature and more closely tied to social norms (House & Kádár, 2021), yet whether this distinction is qualitative or simply a matter of degree is an open question (cf., Leech, 1983). In many ways, due to the need to balance ecological validity and experimental control, the tasks used in the current study represent an impoverished form of communication. It may be that a more naturalistic approach (e.g., conversation analysis or targeted language games; Brown-Schmidt & Tanenhaus, 2008), paired with measurement of a wider array of multimodal data streams (e.g., gesture and prosody; Holler & Levinson, 2019) would reveal subtle differences in the ways that Tsimane' and English speakers produce and interpret even simple referring expressions (e.g., prosodic markers of contrast do not appear to be universal; Swerts et al., 2002). The current study provides a foundation for future work exploring a broader range of psycholinguistic phenomena with nonindustrialized groups such as the Tsimane'.

The current data also bear on another important topic in psycholinguistics: substantial prior work has linked the ability to engage in predictive language processing to the development of literacy (Borovsky et al., 2012; Mishra et al., 2012). Here we observe robust evidence of predictive language processing in a population that had received very little formal education. However, one key difference between the previous paradigms and the current investigation is that the present paradigm manipulates the ability to predict upcoming input based on visual and pragmatic information, rather than lexico-semantic or syntactic cues. It seems plausible that a prediction that relies on a rich model of the language and its rules may only manifest after extensive exposure to the statistics of the language (e.g., Levy, 2008). On the other hand, predictions that are based on reasoning about the intentions of another human may emerge early in development (Nadig & Sedivy, 2002), in similar ways across cultures (Callaghan et al., 2005), and without any formal instruction. Further cross-cultural work is needed to test this speculation. However, such a pattern would be consistent with the idea that the absence of prediction effects observed in certain populations (e.g., young children, lower-literacy populations, and second language learners) does not reflect an absence of prediction per se, but rather an implicit representation of the language that is insufficiently tuned to the input and therefore generates incorrect predictions more frequently than that of the control population of literate, young adults (Ryskin et al., 2020).

In addition to the theoretical implications for language, these findings indicate that the visual world paradigm, and eye-tracking in general, can be fruitfully used to study cognitive processes in real-time in

[16] English speakers also flout these norms when the communicative goals are more multilayered, for example, to joke or to deceive.

remote, nonindustrialized groups. Eye-tracking has only rarely been used in fieldwork with non-WEIRD populations (cf., Norcliffe et al., 2015; Rubio-Fernandez & Jara-Ettinger, 2020). As a result, many findings in cognitive science may be more limited in scope than is often considered. For example, claims of cultural effects on perspective-taking during communication (Wu & Keysar, 2007) are based on participants who are far more culturally similar (e.g., American college students and foreign students studying in the United States) than the populations in the present research. Future work may capitalize on the present findings to investigate how pragmatic inferences are affected by differences in visual perspective and/or knowledge state between Tsimane' speakers and listeners (e.g., when a member of a contrast set is visible only to the listener; Heller et al., 2008).

To conclude, in a pair of classic psycholinguistic tasks probing production and comprehension of referential expressions across visual contexts and adjective types, patterns of behavior of Tsimane' and English participants were strikingly similar and in accordance with Gricean maxims. Though it will be important to collect data from additional nonindustrialized cultures with different languages, these results suggest that inferences about the conversation partner's communicative intent are tied to language use patterns and deployed rapidly during online language processing in Tsimane', and the underlying assumptions about principles of cooperativity may be a cross-cultural feature of human communication.

References

H. Aparicio, Xiang, M., & Kennedy, C. (2016, January). Processing gradable adjectives in context: A visual world study. *Proceedings of SALT*, *25*, 413–432 https://doi.org/10.3765/salt.v25i0.3128

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011, January). Overspecification facilitates object identification. *Journal of Pragmatics*, *43*(1), 361–374. https://doi.org/10.1016/j.pragma.2010.07.013

Atran, S., & Medin, D. L. (2008). *The native mind and the cultural construction of nature*. MIT Press.

Bates, D., Alday, P., Kleinschmidt, D., Caldern, J. B. S., Noack, A., Kelman, T., & Baldassari, A. (2020, March). *JuliaStats/MixedModels.jl: V2.3.0*. Zenodo. https://doi.org/10.5281/zenodo.3727845

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Boas, F. (1938). *Handbook of American Indian languages*. U.S. Government Printing Office.

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* [Computer program]. http://www.praat.org/

Borovsky, A., Elman, J. L., & Fernald, A. (2012, August). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Brainard, D. H. (1997, January). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brehm, L., & Meyer, A. S. (2021, March). Planning when to say: Dissociating cue use in utterance initiation using cross-validation. *Journal of Experimental Psychology: General*, *150*(9), 1772–1799. https://doi.org/10.1037/xge0001012

Brown-Schmidt, S., & Konopka, A. E. (2011). Experimental approaches to referential domains and the on-line processing of referring expressions in unscripted conversation. *Information*, *2*(2), 302–326. https://doi.org/10.3390/info2020302

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*(4), 643–684. https://doi.org/10.1080/03640210802066816

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., & Singh, S. (2005, May). Synchrony in the onset of mental-state reasoning: Evidence from five cultures. *Psychological Science*, *16*(5), 378–384. https://doi.org/10.1111/j.0956-7976.2005.01544.x

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002, July). Circumscribing referential domains during real-Time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49. https://doi.org/10.1006/jmla.2001.2832

Cho, S. J., Brown-Schmidt, S., & Lee, W. Y. (2018, February). Autoregressive generalized linear mixed effect models with crossed random effects: An application to intensive binary time series eye-tracking data. *Psychometrika*, *83*(3), 751–771. https://doi.org/10.1007/s11336-018-9604-2

Clark, H. H., & Wilkes-Gibbs, D. (1986, February). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Cohen, B., & Murphy, G. L. (1984). Models of concepts. *Cognitive Science*, *8*(1), 27–58. https://doi.org/10.1207/s15516709cog0801_2

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2017, November). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, *90*(3), 759–773. https://doi.org/10.1111/cdev.12974.

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020, July). When redundancy is useful: A Bayesian approach to "Overinformative" referring expressions. *Psychological Review*, *127*(4), 591–621. https://doi.org/10.1037/rev0000186

Evans, N. (2003). Context, culture, and structuration in the languages of australia. *Annual Review of Anthropology*, *32*(1), 13–40. https://doi.org/10.1146/anthro.2003.32.issue-1

Evans, N., & Levinson, S. C. (2009, October). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Everett, D. L. (2012). *Language: The cultural tool*. Vintage Books.

Flynn, J. R. (2007). *What is intelligence?: Beyond the Flynn effect*. Cambridge University Press.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., & Conway, B. R. (2017, October). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785–10790. https://doi.org/10.1073/pnas.1619666114

Givón, T. (2005). *Context as other minds: The pragmatics of sociality, cognition and communication*. John Benjamins Publishing. Retrieved August 22, 2022, from https://doi.org/10.1075/z.130

Givón, T., & Young, P. (2002). Cooperation and interpersonal manipulation in the society of intimates. In M. Shibatani (Ed.), *Typological studies in language* (Vol. 48, pp. 23–56). John Benjamins Publishing. Retrieved August 18, 2022, from https://benjamins.com/catalog/tsl.48.05giv

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 41–58). Brill.

Gurven, M. (2018). Broadening horizons: Sample diversity and socioecological theory are essential to the future of psychological science. *Proceedings of the National Academy of Sciences*, *115*(45), 11420–11427. https://doi.org/10.1073/pnas.1720433115.

Gurven, M., Kaplan, H., & Supa, A. Z. (2007). Mortality experience of Tsimane Amerindians of Bolivia: Regional variation and temporal trends. *American Journal of Human Biology*, *19*(3), 376–398. https://doi.org/10.1002/(ISSN)1520-6300

Gurven, M., Kraft, T. S., Alami, S., Adrian, J. C., Linares, E. C., Cummings, D., & Trumble, B. (2020, October). Rapidly declining body temperature in a tropical human population. *Science Advances*, *6*(44), Article eabc6599. https://doi.org/10.1126/sciadv.abc6599

Haspelmath, M. (1999, January). Optimality and diachronic adaptation. *Zeitschrift Für Sprachwissenschaft*, *18*(2), 180–205. https://doi.org/10.1515/zfsw.1999.18.2.180

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836. https://doi.org/10.1016/j.cognition.2008.04.008

Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition*, *149*, 104–120. https://doi.org/10.1016/j.cognition.2015.12.008

Henrich, J., Heine, S. J., & Norenzayan, A. (2010, June). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henrich, J., Young, P., Boyd, R., McCabe, K., Albers, W., Ockenfels, A., & Gigerenzer, G. (2001). What Is the role of culture in bounded rationality? In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 343–359). MIT Press.

Holler, J., & Levinson, S. C. (2019, August). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Holler, J., & Stevens, R. (2007, March). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology*, *26*(1), 4–27. https://doi.org/10.1177/0261927X06296428

House, J., & Kádár, D. Z. (2021). *Cross-Cultural pragmatics* (1st ed.). Cambridge University Press. https://www.cambridge.org/core/product/identifier/9781108954587/type/book

Isaacs, E., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, *116*(1), 26–37. https://doi.org/10.1037/0096-3445.116.1.26

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019, October). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, *29*(19), 3229–3243.e12. https://doi.org/10.1016/j.cub.2019.08.020

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, *81*(2), 345–381. https://doi.org/10.1353/lan.2005.0071

Kraft, T. S., Stieglitz, J., Trumble, B. C., Martin, M., Kaplan, H., & Gurven, M. (2018, December). Nutrition transition in 2 lowland Bolivian subsistence populations. *The American Journal of Clinical Nutrition*, *108*(6), 1183–1195. https://doi.org/10.1093/ajcn/nqy250

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*(1–12), 113–114. https://doi.org/10.3758/BF03342817

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*(3), 343–346. https://doi.org/10.1037/h0023705

Kreiss, E., & Degen, J. (2020). Production expectations modulate contrastive inference. *Proceedings of the Cognitive Science Society*, pp. 259–265. https://cognitivesciencesociety.org/cogsci20/papers/0051/0051.pdf

Leech, G. N. (1983). Principles of pragmatics. Longman.

Levinson, S. C. (2011). Universals in pragmatics. In P. C. Hogan (Ed.), *The Cambridge encyclopedia of the language sciences* (pp. 654–657). Cambridge University Press.

Levy, R. (2008, March). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations*. Harvard University Press.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using Kaldi. *Proceedings of the interspeech 2017* (pp. 498–502).

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye- movements are modulated by reading ability—evidence from Indian low and high literates. *Journal of Eye Movement Research*, *5*, Article 3.

Nadig, A., & Sedivy, J. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), Article 8. https://doi.org/10.1111/j.0956-7976.2002.00460.x

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015, October). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience*, *30*(9), 1187–1208. https://doi.org/10.1080/23273798.2015.1006238

Ogiermann, E. (2009, January). Politeness and in-directness across cultures: A comparison of English, German, Polish and Russian requests. *Journal of Politeness Research. Language, Behaviour, Culture*, *5*(2), 189–216. https://doi.org/10.1515/JPLR.2009.011

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, *77*(4), 257–273. https://doi.org/10.1037/h0029436

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*(1), 89–110. https://doi.org/10.1515/ling.1989.27.1.89

Pitt, B., Ferrigno, S., Cantlon, J., Casasanto, D., Gibson, E., & Piantadosi, S. T. (2021, April). Spatial concepts of number, size, and time in an indigenous culture. *Science Advances*, *7*(33), Article eabg4141. https://doi.org/10.31234/osf.io/kq5ma

R Core Team. (2019). *R: A language and environment for statistical computing*.

Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, *7*, Article 8. https://doi.org/10.3389/fpsyg.2016.00153.

Rubio-Fernandez, P., & Jara-Ettinger, J. (2020, June). Incrementality and efficiency shape pragmatics across languages. *Proceedings of the National Academy of Sciences*, *117*(24), 13399–13404. https://doi.org/10.1073/pnas.1922067117

Ryskin, R. A., Benjamin, A. S., Tullis, J., & Brown-Schmidt, S. (2015). Perspective-taking in comprehension, production, and memory: An individual differences approach. *Journal of Experimental Psychology: General*, *144*(5), 898–915. https://doi.org/10.1037/xge0000093

Ryskin, R. A., Kurumada, C., & Brown-Schmidt, S. (2019). Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science*, *43*(8), Article e12769. https://doi.org/10.1111/cogs.2019.43.issue-8

Ryskin, R. A., Levy, R. P., & Fedorenko, E. (2020, January). Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward. *Neuropsychologia*, *136*, Article 107258. https://doi.org/10.1016/j.neuropsychologia.2019.107258

Ryskin, R. A., Salinas, M. A., & Gibson, E. (2022, September). *Contrastive interpretation of adjectives in Tsimane': OSF*. https://osf.io/bjs85/

Ryskin, R. A., Salinas, M. A., Piantadosi, S. T., & Gibson, E. (2021, January). *Real-time pragmatic inference across cultures: Evidence from a non-industrialized society* (Technical Report). PsyArXiv. https://doi.org/10.31234/osf.io/ga83j

Sakel, J. (2004). *A grammar of mosetén*. Walter de Gruyter.

Sedivy, J. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*(1), 3–23. https://doi.org/10.1023/A:1021928914454

Sedivy, J. (2005). Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. In J. C. Trueswell, & M. K. Tanenhaus (Eds.), *Approaches to studying world-*

situated language use: Bridging the language-as-product and language-as-action traditions (pp. 345–364). MIT Press.

Sedivy, J., Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147. https://doi.org/10.1016/S0010-0277(99)00025-6

Smith, K., & Kirby, S. (2008). Cultural evolution: Implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3591–3603. https://doi.org/10.1098/rstb.2008.0145

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., & Levinson, S. C., 2009, June). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics*, *30*(4), 629–654. https://doi.org/10.1006/jpho.2002.0178

Tanaka, J. W., & Taylor, M. (1991, July). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482. https://doi.org/10.1016/0010-0285(91)90016-H

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Thompson, B., Roberts, S. G., & Lupyan, G. (2020, August). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, *4*, 1029–1038, https://doi.org/10.1038/s41562-020-0924-8

Tomasello, M. (1999, October). The human adaptation for culture. *Annual Review of Anthropology*, *28*(1), 509–529. https://doi.org/10.1146/anthro.1999.28.issue-1

Trudgill, P. (2012). On the sociolinguistic typology of linguistic complexity loss. *Language Documentation & Conservation*, *3*, 90–95.

van Gompel, R. P. G., van Deemter, K., Gatt, A., Snoeren, R., & Krahmer, E. J. (2019, April). Conceptualization in reference production: Probabilistic modeling and experimental testing. *Psychological Review*, *126*(3), 345–373. https://doi.org/10.1037/rev0000138

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer.

Wu, S., & Keysar, B. (2007, July). The effect of culture on perspective taking. *Psychological Science*, *18*(7), 600–606. https://doi.org/10.1111/j.1467-9280.2007.01946.x

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018, July). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942. https://doi.org/10.1073/pnas.1800521115

(*Appendices follows*)

Appendix

Table A1
Regression Model Estimates for Analyses of Adjective Production in the Referential Communication Task

Model 1: Size adjectives			Total obs.	315
Fixed-effects parameters	Estimate	*SE*	*z*-value	*p*-value
Intercept (no contrast)	− 4.41	0.58	− 7.64	$< 1 \times 10^{-13}$
Two contrasts	1.54	0.60	2.61	9.2×10^{-3}
Contrast	2.16	0.57	3.77	2×10^{-4}
Variance components	Item ($n = 30$)	Participant ($n = 21$)		
Intercept	0.13	1.11		
Model 2: Color adjectives			Total obs.	756
Fixed-effects parameters	Estimate	*SE*	*z*-value	*p*-value
Intercept (no contrast)	0.01	0.42	0.02	.98
Two contrasts	0.65	0.37	1.75	.08
Contrast	0.76	0.37	2.05	.04
Variance components	Item ($n = 72$)	Participant ($n = 21$)		
Intercept	1.15	2.17		
Model 3: Color adjectives by type			Total obs.	756
Fixed-effects parameters	Estimate	*SE*	*z*-value	*p*-value
Intercept	1.20	0.51	2.38	.02
Two contrasts	0.69	0.54	1.28	.20
Contrast	1.34	0.57	2.35	.02
High-entropy artificial	−1.84	0.53	−3.49	5×10^{-4}
High-entropy natural	−2.55	0.51	−4.98	$< 1 \times 10^{-6}$
Low-entropy natural	−0.52	0.51	−1.04	.30
Two contrasts: high-entropy artificial	−0.27	0.74	−0.37	.71
Contrast: high-entropy artificial	−1.01	0.75	−1.35	.18
Two contrasts: high-entropy natural	−0.09	0.72	−0.13	.90
Contrast: high-entropy natural	−0.23	0.75	−0.31	.76
Two contrasts: low-entropy natural	0.28	0.74	0.37	.71
Contrast: low-entropy natural	−0.67	0.76	−0.87	.38
Variance components	Item ($n = 72$)	Participant ($n = 21$)		
Intercept	1.20	2.50		

(*Appendices continue*)

Table A2

Model Selection for Random Effects in the AR(1) GLMMs

All adjectives		Model 1	Model 2	Model 3	Model 4	**Model 5**	Model 6
Trial order	Int.	x	x	x	x	x	x
	AR(1)		x	x	x	x	x
Subject	Int.	x	x	x	x	x	x
	Contrast			x	x	x	x
	Adjective type					x	x
	Contrast: Adjective type						x
	AR(1)		x	x	x	x	x
Item	Int.	x	x	x	x	x	x
	Contrast				x	x	x
	AR(1)		x	x	x	x	x
LL		− 109, 544.26	− 106, 947.24	− 106, 907.90	− 106, 876.97	**− 106, 865.67**	− 106, 853.62
Npar		12	18	21	24	**28**	33
AIC		219,112.53	213,930.47	213,857.80	213,801.93	**213,787.34**	213,773.24
BIC		219,252.36	214,140.22	214,102.50	214,081.59	**214,113.61**	214,157.77
Color adjectives		Model 1	Model 2	Model 3	**Model 4**	Model 5	
Trial order	Int.	x	x	x	x	x	
	AR(1)		x	x	x	x	
Subject	Int.	x	x	x	x	x	
	Contrast			x	x	x	
	Adjective type					x	
	Contrast: Adjective type						
	AR(1)		x	x	x	x	
Item	Int.	x	x	x	x	x	
	Contrast				x	x	
	AR(1)		x	x	x	x	
LL		− 86, 882.42	− 86, 501.10	− 86, 476.07	**− 86, 451.76**	− 86, 425.63	
Npar		22	26	29	**32**	47	
AIC		173,808.83	173,054.21	173,010.15	**172,967.52**	172,945.26	
BIC		174,060.30	173,351.40	173,341.62	**173,333.29**	173,482.48	

Note. Bold font indicates models with the lowest BIC which are reported in Tables A3 and A5. GLMMs = generalized linear mixed models; LL = log likelihood; Npar = number of parameters; AIC = Akaike's information criteria; BIC = Bayesian information criteria.

Table A3

Regression Model Estimates for Analyses Comparing Target Looks in the Eye-Tracking Task Across Types of Adjectives (Size vs. Color) and Population/Language (Total Observations: 849,528)

Fixed-effects parameters	Estimate	*SE*	*z*-value	*p*-value
Intercept	− 3.81	0.04	− 85.87	$< 1 \times 10^{-99}$
Contrast	0.08	0.01	5.28	$< 1 \times 10^{-6}$
Adjective type	0.10	0.02	5.24	$< 1 \times 10^{-6}$
Language	−0.07	0.02	− 3.45	.00
AR(1)	7.09	0.09	78.00	$< 1 \times 10^{-99}$
Contrast: Adjective type	−0.03	0.01	− 2.53	.01
Contrast: Language	0.00	0.01	0.33	.74
Adjective type: Language	0.01	0.02	0.31	.75
Contrast: Adjective type: Language	0.00	0.01	0.38	.70
Variance components	**Item (*n* = 200)**	**Participant (*n* = 124)**	**(Trial order *n* = 117)**	
Intercept	0.06	0.15	0.02	
Contrast	0.01	0.01	–	
Adjective type	–	0.01	–	
AR(1)	0.17	0.80	0.09	

Note. Items are partially nested within participants because the audio recordings were different depending on the language.

(*Appendices continue*)

Table A4

Regression Model Estimates for Analyses Comparing Target Advantage During the Adjective Time Window in the Eye-Tracking Task Across Types of Adjectives (Size vs. Color) and Population/Language (Total Observations: 10,869)

Fixed-effects parameters	Estimate	*SE*	*z*-value	*p*-value
Intercept	0.04	0.01	3.21	.00
Contrast	0.04	0.01	5.47	$< 1 \times 10^{-7}$
Adjective type	0.01	0.01	1.11	.27
Language	−0.03	0.01	− 2.29	.02
Contrast: Adjective type	−0.01	0.01	− 1.70	.09
Contrast: Language	−0.01	0.01	− 1.28	.20
Adjective type: Language	0.02	0.01	1.85	.07
Contrast: Adjective type: Language	0.00	0.01	0.49	.63
Variance components	**Item (n = 200)**	**Participant (n = 124)**	**(Trial order n = 117)**	
Intercept	0.01	0.00	0.00	
Contrast	0.00	0.00		
Adjective type		0.00		
Contrast: Adjective type		0.00		
Residual variance = 0.31				

Note. Items are partially nested within participants because the audio recordings were different depending on the language.

Table A5

Regression Model Estimates for Analyses Comparing Target Looks in the Eye—Tracking Task Across Subtypes of Color Adjectives and Population/Language (Total Observations: 680,266)

Fixed-effects parameters	Estimate	*SE*	*z*-value	*p*-value
Intercept	− 3.74	0.05	− 73.65	$< 1 \times 10^{-99}$
Contrast	0.05	0.02	2.10	.04
High-entropy artificial	−0.07	0.04	− 1.60	.11
High-entropy natural	0.01	0.04	0.30	.76
Low-entropy natural	0.15	0.04	3.80	.00
Language	−0.06	0.03	− 2.00	.05
AR(1)	7.09	0.09	75.65	$< 1 \times 10^{-99}$
Contrast: High-entropy artificial	0.00	0.03	0.07	.95
Contrast: High-entropy natural	0.00	0.03	0.03	.97
Contrast: Low-entropy natural	−0.02	0.03	−0.65	.51
Contrast: Language	0.00	0.02	−0.13	.89
High-entropy artificial: Language	−0.05	0.04	− 1.20	.23
High-entropy natural: Language	−0.01	0.04	−0.14	.89
Low-entropy natural: Language	0.02	0.04	0.61	.54
Contrast: High-entropy artificial: Language	0.02	0.03	0.82	.41
Contrast: High-entropy natural: Language	0.03	0.03	1.11	.27
Contrast: Low-entropy natural: Language	0.00	0.03	−0.15	.88
Variance components	**Item (n = 160)**	**Participant (n = 124)**	**(Trial order n = 117)**	
Intercept	0.07	0.15	0.03	
AR(1)	0.16	0.82	0.10	
Contrast	0.01	0.01		

(*Appendices continue*)

Table A6

Regression Model Estimates for Analyses Comparing Target Advantage During the Adjective Time Window in the Eye-Tracking Task Across Types of Color Adjectives and Population/Language (Total Observations: 8686)

Fixed-effects parameters	Estimate	*SE*	*z*-value	*p*-value
Intercept	0.00	0.02	0.13	.90
Contrast	0.04	0.02	2.58	.01
High-entropy artificial	0.01	0.03	0.56	.58
High-entropy natural	0.07	0.03	2.42	.02
Low-entropy natural	0.11	0.03	4.20	$< 1 \times 10^{-4}$
Language	0.00	0.02	0.00	1.00
Contrast: High-entropy artificial	−0.02	0.02	−0.87	.38
Contrast: High-entropy natural	0.00	0.02	0.01	.99
Contrast: Low-entropy natural	−0.01	0.02	−0.40	.69
Contrast: Language	−0.01	0.02	−0.71	.48
High-entropy artificial: Language	−0.03	0.03	−0.95	.34
High-entropy natural: Language	−0.01	0.03	−0.43	.67
Low-entropy natural: Language	0.01	0.03	0.24	.81
Contrast: High-entropy artificial: Language	0.01	0.02	0.65	.51
Contrast: High-entropy natural: Language	0.02	0.02	0.88	.38
Contrast: Low-entropy natural: Language	−0.01	0.02	−0.66	.51

Variance components	Item (n = 160)	Participant (n = 124)	(Trial order n = 117)
Intercept	0.01	0.00	0.00
Contrast	0.00	0.01	
High-entropy artificial		0.00	
High-entropy natural		0.01	
Low-entropy natural		0.00	
Contrast: High-entropy artificial		0.00	
Contrast: High-entropy natural		0.00	
Contrast: Low-entropy natural		0.00	
Residual variance = 0.32			

Note. Items are partially nested within participant because the audio recordings were different depending on the language.

Received October 08, 2021
Revision received September 09, 2022
Accepted November 10, 2022 ■