



Original Articles

Contrast perception as a visual heuristic in the formulation of referential expressions

Madeleine Long^{a,1}, Isabelle Moore^{b,1}, Francis Mollica^c, Paula Rubio-Fernandez^{a,d,*}

^a Philosophy Department, University of Oslo, Norway

^b Psychology Department, University of Virginia, United States of America

^c Informatics Department, University of Edinburgh, United Kingdom

^d Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, United States of America

ARTICLE INFO

Keywords:

Referential communication

Contrast perception

Color adjectives

Redundancy

Efficiency

ABSTRACT

We hypothesize that contrast perception works as a visual heuristic, such that when speakers perceive a significant degree of contrast in a visual context, they tend to produce the corresponding adjective to describe a referent. The contrast perception heuristic supports efficient audience design, allowing speakers to produce referential expressions with minimum expenditure of cognitive resources, while facilitating the listener's visual search for the referent. We tested the perceptual contrast hypothesis in three language-production experiments. Experiment 1 revealed that speakers overspecify color adjectives in polychrome displays, whereas in monochrome displays they overspecified other properties that were contrastive. Further support for the contrast perception hypothesis comes from a re-analysis of previous work, which confirmed that color contrast elicits color overspecification when detected in a given display, but not when detected across monochrome trials. Experiment 2 revealed that even atypical colors (which are often overspecified) are only mentioned if there is color contrast. In Experiment 3, participants named a target color faster in monochrome than in polychrome displays, suggesting that the effect of color contrast is not analogous to ease of production. We conclude that the tendency to overspecify color in polychrome displays is not a bottom-up effect driven by the visual salience of color as a property, but possibly a learned communicative strategy. We discuss the implications of our account for pragmatic theories of referential communication and models of audience design, challenging the view that overspecification is a form of egocentric behavior.

1. Introduction

When we refer to the world around us, perception guides the formulation of our message. Imagine that you invite a neighbor over for coffee: you may offer her ‘this cup’ or ‘that cup’, depending on where on the table you placed her decaf. Alternatively, you may preempt any ambiguity by specifying the color or the size of her cup (e.g., ‘The blue cup is the decaf’). These simple examples illustrate how in order to avoid referential ambiguity, speakers must compare the object they want to refer to with its *competitors* (i.e. objects of the same kind in the visual context; e.g., any other cup on the table) and provide a distinctive property that the listener can use to identify the intended referent. When there are no competitors in the visual context, no extra information is needed to pre-empt an ambiguity (e.g., ‘The sugar is on the counter’).

This way of tailoring our referential expressions for our listeners is known as *audience design*. The aim of this study is to investigate the hypothesis that audience design can rely on perceptual heuristics, such as the detection of color contrast. In this view, perceptual heuristics would favor referential expressions that are efficient for both speakers and listeners; that is, descriptions that are easy to produce, but also facilitate the visual search for a referent.

While generally consistent with the above picture, the experimental record of the last 20 years has revealed two intriguing findings in reference production, each employing a different psycholinguistic measure: speakers' looking behavior when formulating referential expressions, and speakers' choice of referential expression. A number of eye-tracking studies have shown that speakers are more likely to be sufficiently informative when they have fixated on a competitor object

* Corresponding author at: Philosophy Department, University of Oslo, Norway.

E-mail address: paula.rubio-fernandez@ifikk.uio.no (P. Rubio-Fernandez).

¹ Equal contribution.

before producing a referential expression (e.g., fixating on a small table before referring to ‘the large table’; Brown-Schmidt & Tanenhaus, 2006; Davies & Kreysa, 2017). However, these studies have also revealed that fixating on a competitor is not necessary in order for a speaker to produce an adequately informative description: Davies and Kreysa (2017) found that when speakers were shown simple displays, 83% of referring expressions were sufficiently informative without any fixations on the competitor object, whereas in more complex displays, 53% of utterances were informative without fixations on the competitor (see also Brown-Schmidt & Tanenhaus, 2006). Importantly, participants who did not fixate on the competitor were not simply ignoring its presence, as evidenced by the fact that these participants used modifiers 62% of the time in trials with competitors compared to only 3% in trials without competitors. Taken together, the results of these eye-tracking studies confirm that while fixating on a competitor may boost informativity, it is not essential, especially in sparser displays.

The second set of intriguing results in the referential communication literature shows that speakers produce adjectives not only when they are necessary to preempt an ambiguity with a competitor, but also when they are unnecessary in the visual context (e.g., Arts, Maes, Noordman, & Jansen, 2011a; Belke, 2006; Maes, Arts, & Noordman, 2004; Nadig & Sedivy, 2002; Sedivy, 2003, 2005). Some researchers have argued that speakers tend to overspecify visually salient properties (such as color) because it is easier than having to identify a referent’s competitors and directly compare them (Pechmann, 1989; Belke & Meyer, 2002; Belke, 2006; Engelhardt, Bailey, & Ferreira, 2006; Koolen, Goudbeek, & Krahmer, 2013; Fukumura & Carminati, 2021). This interpretation is in line with the view that referential overspecification is driven by speaker-internal processes, whereby the speaker fails to adopt the listener’s perspective and produce an optimally informative description (for review and discussion, see Arnold, 2008; Davies & Arnold, 2019). However, this ‘negative view’ of referential overspecification rests on the assumption that overinformative expressions violate the *Maxim of Quantity*, according to which speakers should not provide their listeners with more information than is necessary (Grice, 1975). As a violation of the Gricean maxim, redundancy is understood to be detrimental for the listener (see Davies & Katsos, 2013; Engelhardt et al., 2006; Engelhardt, Demiral, & Ferreira, 2011).

We have recently challenged the negative view of referential overspecification on the grounds that non-restrictive modification can be cooperative in nature (Rubio-Fernandez, 2019, 2021; Rubio-Fernandez, Mollica, & Jara-Ettinger, 2021), hence supporting the view that reference is a *collaborative process* (see Clark, 2006; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992). In this view, the goal of a visually-grounded referential expression is to allow the listener to identify the intended referent fast and easily (Rubio-Fernandez, 2016, 2019; Long, Rohde, & Rubio-Fernandez, 2020; for a review of earlier studies on speaker-listener coordination in referential communication, see Clark & Bangerter, 2004). Whether that interactive goal requires producing the shortest possible referential expression is an empirical question that ultimately depends on the visual context since

redundant adjectives can have *discriminatory value*, despite lacking *informational value*. A color adjective, for example, has informational value if it preempts an ambiguity between several competitors (e.g., various stars in a display of colored shapes), but it may also have discriminatory value if it facilitates the listener’s visual search for the referent (e.g., if a single star is also the only blue shape in the display; see Fig. 1). Supporting the efficiency view, a number of psycholinguistic studies have shown that redundant color adjectives can facilitate the visual search for a referent (Arts, Maes, Noordman, & Jansen, 2011b; Mangold & Pobel, 1988; Paraboni & Van Deemter, 2014; Paraboni, Van Deemter, & Masthoff, 2007; Rubio-Fernandez, 2021; Sonnenschein & Whitehurst, 1982; Tourtouri, Delogu, Sikos, & Crocker, 2019).

The aim of this paper is to develop the efficiency view of referential overspecification by evaluating the role of *speaker-internal processes*. We originally proposed that overspecification may be efficient for the listener by facilitating their visual search for the referent (Rubio-Fernandez, 2016). However, overspecification may also be efficient for the speaker, if it relies on visual cues that trigger the use of modification in appropriate contexts at a low production cost for the speaker (Rubio-Fernandez, 2019). In developing and testing this account, the paper will contribute to two theoretical discussions: the old debate on whether reference production is driven by egocentric processes or by audience design (Arnold, 2008; Dell & Brown, 1991), and the growing discussion on the nature of referential overspecification (Rubio-Fernandez, 2016, 2019, 2021; Tourtouri et al., 2019; van Gompel, van Deemter, Gatt, Snoeren, & Krahmer, 2019; Degen, Hawkins, Graf, Kreiss, & Goodman, 2020; Rubio-Fernandez et al., 2021).

1.1. Visual heuristics and efficiency

We have recently proposed that in situations of co-presence (where speaker and listener share a physical space), speakers do not need to take the listener’s perspective and can rely instead on *visual heuristics* that can inform their choice of referential expression at a low production cost (Rubio-Fernandez, 2019; for earlier proposals of the use of heuristics in referential communication, see also Clark & Marshall, 1981; Dale & Viethen, 2009; Viethen & Dale, 2009; Van Deemter, Gatt, Van Gompel, & Krahmer, 2012; Koolen et al., 2013). Here, we adopt Tversky and Kahneman’s (1974) definition of heuristics as ‘beliefs concerning the likelihood of uncertain events (...) that reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations’ (p. 1124).

Tversky and Kahneman (1974) described three heuristics that are commonly used to assess probabilities under uncertainty, and which lead to systematic biases in reasoning. According to the *adjustment and anchoring heuristic*, people make estimates by starting from an initial value that is adjusted to yield a final answer, with such adjustments often being insufficient. This heuristic has been used to interpret the perspective-taking mistakes frequently observed in referential communication tasks where participants seem to start interpreting language from an egocentric perspective, having to then adjust for this initial bias

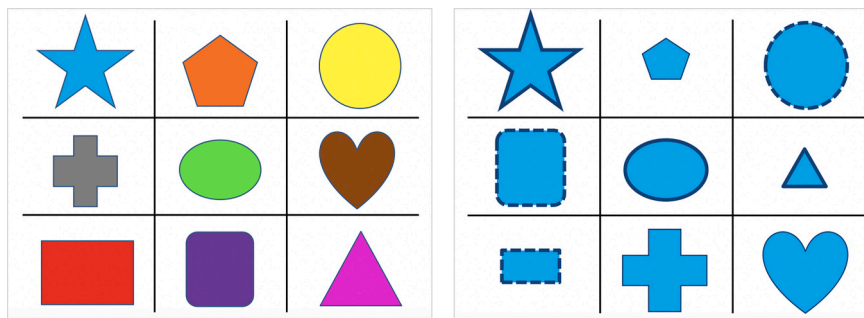


Fig. 1. Sample displays from the polychrome condition (low scene-variation) and monochrome condition (high scene-variation) in Experiment 1.

in order to adopt the speaker's perspective (Epley, Keysar, Van Boven, & Gilovich, 2004; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003).

Here we hypothesize that contrast perception works as a visual heuristic in the formulation of referential expressions. In this view, calculating that color is more efficient in polychrome than in monochrome displays would be the result of a visual heuristic triggered by the perception of color contrast (see Fig. 1). The density of a display also works as a visual heuristic, with denser displays often eliciting higher rates of overspecification than sparser displays (e.g., Clarke, Elsner, & Rohde, 2013; Gatt, Krahmer, Van Deemter, & van Gompel, 2017; Koolen, Krahmer, & Swerts, 2016; Paraboni et al., 2007; Rubio-Fernandez, 2019). We argue that contrast perception is an efficient visual heuristic because identifying distinctive properties of a referent requires contrasting it with its competitors in the visual context. Therefore, relying on properties that are contrastive across the entire visual context is likely to be discriminatory against a referent's competitors, without requiring a specific search for competitors and a direct comparison with the referent.

The development and use of visual heuristics may be related to the processing of contrastive information without direct fixations on competitor objects. Davies and Kreysa (2017) argue that when an object is highly recognizable, it is processed *extrafoveally* (Meyer, Sleiderink, & Levelt, 1998; Morgan & Meyer, 2005). In other words, when contrastive information is easier to identify (e.g., because the display is sparse or the properties are visually salient), the visual system is able to integrate relevant information without having to fixate on competitor objects. Similarly, processing information extrafoveally may facilitate the detection of contrastive properties and trigger the use of modification.

We propose that speakers rely on perceptual contrast as a visual heuristic to produce efficient referential expressions *efficiently*. That is, to produce referential expressions that facilitate the listener's visual search, while requiring limited effort on the speaker's part. Under a contrast perception heuristic, significant perceptual contrast will trigger modification. Thus, if a speaker perceives a sufficiently salient contrast in a visual display, they will produce the corresponding adjective regardless of whether the adjective has informational value (i.e. if competitors have been identified in the display) or only discriminatory value (i.e. if no competitor has been identified). We further hypothesize that the estimation of perceptual contrast is mediated by the relative density of the display, with denser displays eliciting higher rates of modification (Rubio-Fernandez, 2019). The contrast perception heuristic will therefore result in speakers sometimes not producing an adjective (e.g., if they do not perceive a salient contrast, or the display is too sparse to warrant the use of modification), whereas other times they will produce redundant modification (e.g., if they did not identify a competitor but the display is dense with contrasting objects).

While the perceptual contrast heuristic brings together reference production and visual perception, it is different from the widespread view that the redundant use of color adjectives is driven by speaker-internal processes whereby speakers prefer to mention visually salient properties of the referent, rather than searching for potential competitors and producing optimally informative descriptions (e.g., Pechmann, 1989; Belke & Meyer, 2002; Koolen et al., 2013; Fukumura & Carminati, 2021). Unlike these accounts, we conceptualize the perceptual contrast heuristic as a form of audience design that is efficient for both the speaker and the listener.

1.2. Visual heuristics and audience design

Relying on visual heuristics may be understood as a form of 'low-cost pragmatics' in line with Ferreira's (Ferreira, 2019) *feedforward audience design*. The crucial innovation in this framework is that the process of grammatical encoding may relatively automatically (i.e. without the involvement of executive control) encode meaning into linguistic features in such a way as to implement audience design strategies. These

tactic strategies are what Ferreira calls *feedforward audience design*. Importantly, this form of audience design is heavily limited since contextual features, learned through communicative experience, must be available prior to grammatical encoding in order to drive an audience design effect.

Here we propose that the use of contrast perception as a visual heuristic in the formulation of referential expressions (e.g., producing a color adjective to refer to a target in a polychrome display of shapes) is also a form of feedforward audience design. Visual heuristics are relatively effortless to implement but have low accuracy, whereas other referential strategies require more complex reasoning about interlocutors and are more accurate, but also take more time and effort. For example, if speakers and listeners have different perspectives on a visual display (see, e.g., Wardlow-Lane & Ferreira, 2008; Long, Horton, Rohde, & Sorace, 2018), speakers need to engage in recurrent processing audience design, which is more effortful and time consuming (see Buz, Tanenhaus, & Jaeger, 2016; Fedzechkina, Jaeger, & Newport, 2012; Jaeger & Ferreira, 2013; Kurumada & Jaeger, 2015). At a higher level of description, a system of audience design that distinguishes fast and efficient strategies from slower but more accurate strategies is generally consistent with dual process theories (e.g., Evans, 1984; Kahneman, 2003), but may also reflect optimal metareasoning for strategy selection (Lieder & Griffiths, 2017; Milli, Lieder, & Griffiths, 2017).

Dale and Viethen (2009) performed a corpus analysis of 10 referential expressions generated in the same visual contexts by 63 participants. The results of their analyses suggest that referring behavior might be constructed as a combination of lower-level heuristics. Akin to Ferreira's mechanistic framework, Dale and Viethen propose that the specific heuristics that a speaker might use vary depending on their personal past history, and perhaps even on the basis of situation-specific factors that might prompt speakers to be more or less accurate in their referential expressions.

What kind of communicative experiences could lead to the learning of visual heuristics in referential communication? The developmental record might provide some answers to this question. Referential communication studies with toddlers and preschoolers have shown that young children tend to produce underinformative referential expressions that require further clarification (Matthews, Butcher, Lieven, & Tomasello, 2012; Matthews, Lieven, & Tomasello, 2007). These results suggest that the process of learning to uniquely describe a referent often involves a direct comparison of the potential referents by the adult addressee (e.g., 'Which little girl, the one eating an ice-cream or the one stroking a dog?'). As part of this learning process, children come to understand the nature of referential ambiguity and the need for audience design, eventually automatizing the perceptual process whereby they contrast the referent and its competitors in search for discriminatory properties.

Given the contrastive nature of reference, perceptual contrast is a likely candidate for a visual heuristic in the generation of referential expressions. Thus, whereas uniquely describing a referent might require careful comparison with a competitor in some visual contexts (e.g., if they are both visually similar and it is hard to identify a distinguishing property), often enough, a perceptual property that is contrastive across the entire visual context would be an efficient choice (e.g., if all objects are different sizes, patterns or colors). Thus, selecting a contrastive property when producing a referential expression may be efficient not only for the speaker, but also for the listener – which makes it a form of feedforward audience design.

1.3. Perceptual and linguistic factors affecting referential choice

Color adjectives tend to be overspecified more frequently than other types of adjectives, such as size or material (Pechmann, 1989; Rubio-Fernandez, 2019; Tarenskeen, Broersma, & Geurts, 2015). From a perceptual perspective, color contrast is indeed highly salient, playing a critical role in object recognition (Bramão, Reis, Petersson, & Faisca,

2011; Gegenfurtner & Rieger, 2000) and other cognitive processes related to memory, language and attention (Adams & Chambers, 2012; Davidoff, 1991). Under a contrast perception heuristic, the visual salience of color contrast should trigger color overspecification more often than other types of adjectives, as the experimental record confirms (see also Viethen, van Vessum, Goudbeek, & Krahmer, 2017). However, the contrast perception heuristic is triggered not only by color contrast, but also by any other perceptual contrast that is salient enough in a visual context to be mentioned in referential communication. Thus, if the material contrast between various objects is salient enough in a given situation (e.g., glass vs. wood), speakers may use a material adjective redundantly when referring to one of these objects (see Jara-Ettinger & Rubio-Fernandez, 2021).

The visual salience of color has led some researchers to argue that the perception of *scene variation* triggers the use of redundant color adjectives. For example, Koolen et al. (2013) operationalize scene variation as the number of dimensions along which the objects in a scene vary (e.g., size, material or orientation) and argue that in high-variation scenes, speakers are less certain as to which attributes rule out the referent's competitors, and are therefore more likely to mention a salient property such as color. Degen et al. (2020) also endorse Koolen et al.'s scene variation hypothesis (albeit under a different formalization), arguing that speakers' tendency to overmodify with color adjectives increases as the variation in the scene increases.

Scene variation is directly related to perceptual contrast (with each varying dimension in a scene providing a source of perceptual contrast) and it may also work as a 'quick heuristic' (Koolen et al., 2013). However, despite these commonalities, and besides the specific formalization of scene variation (e.g., whether it is understood as the number of dimensions that vary in a scene, or the degree of variation within a dimension), there is a fundamental difference between the scene variation hypothesis and the perceptual contrast heuristic: according to Koolen et al. (2013) and Degen et al. (2020), the perception of scene variation results in the mention of color adjectives, whereas we hypothesize that the perception of contrast results in the mention of the contrasting property – which need not be color.

Following earlier work in Natural Language Generation, Koolen et al. (2013) treat color as a 'preferred attribute' because of its inherent visual salience. The perceptual contrast heuristic is also sensitive to the visual salience of color contrast, predicting that color contrast is more likely to trigger the use of color modification relative to other types of contrast that may not be so perceptually salient as to warrant the mention of the corresponding adjective. However, the perceptual contrast heuristic does not treat color as a preferred attribute that is mentioned by default when other types of contrast are detected. The scene variation hypothesis and the perceptual contrast heuristic therefore make different predictions that will be tested in Experiment 1.

While perception is obviously central to visual heuristics, linguistic factors can also determine referential overspecification (Rubio-Fernandez, 2016). For instance, size adjectives are interpreted in relation to a comparison class, whereas color adjectives encode an absolute property (e.g., a farm may be big compared to a smaller farm, or in relation to the average farm, but it can be red in and of itself; see Kennedy, 2007; Kennedy & McNally, 2005). Thus, color adjectives are used redundantly or non-contrastively more often than size adjectives (Rubio-Fernandez, 2019). However, when a size contrast is sufficiently salient in a visual context (e.g., in a *pop-out display* where the target is the smallest or the largest object in the entire display), size adjectives are used redundantly 60% of the time (Rubio-Fernandez, 2019; for a discussion of the color-size asymmetry in adjective production, see Pechmann, 1989; Degen et al., 2020). We interpret these findings as providing support for the perceptual contrast heuristic.

Another linguistic factor that has been shown to affect the use of redundant adjectives is adjective position. Recent studies have shown that English speakers produce redundant color adjectives in prenominal position more often than Spanish speakers do in postnominal position (e.

g., 'The blue circle' vs 'El círculo azul'; Rubio-Fernandez, 2016, 2019; Wu & Gibson, 2021; see also Kachakeche, Futrell, & Scontras, 2021). This difference supports the view that color adjectives are used redundantly to facilitate the listener's visual search for the referent, since they are a more efficient visual cue in prenominal position. Further support for this view comes from Rubio-Fernandez et al. (2021), who observed that the difference in redundant color rates between English and Spanish speakers disappeared in denser displays, in which redundant color adjectives were efficient even in postnominal position.

Finally, different lexical categories tend to elicit different rates of redundant color adjectives, depending on the strength of the association between the lexical category and the color property. For example, Rubio-Fernandez (2016, 2019) observed that in polychrome displays of four objects, people mentioned the color of geometrical shapes 40% of the time, whereas they did so 95% of the time when the displays contained clothes. Relevant to the present investigation, people produced redundant color adjectives 40% of the time in monochrome displays of clothes (Rubio-Fernandez, 2016), whereas they produced zero rates of color modification in monochrome displays of geometrical shapes (Rubio-Fernandez, 2019).

In conclusion, we hypothesize that perceptual contrast works as a visual heuristic in the production of redundant adjectives. While the salience of color contrast may trigger color overspecification more frequently than other types of contrast elicit the use of other perceptual adjectives, our hypothesis extends to all types of perceptual contrast (as long as it is salient enough in a given context). In addition, linguistic factors such as adjective semantics (absolute vs. gradable), adjective position (prenominal vs postnominal) or lexical category (e.g., geometrical shapes vs clothes) can also affect the production of redundant modification, above and beyond perceptual contrast.

1.4. The present study

Here we tested the hypothesis that perceptual contrast triggers overspecification, working as a visual heuristic that is efficient for both the speaker and the listener. To test this hypothesis, we employed three language-production tasks, and a color-naming experiment. Two of the language-production tasks were purposely designed for this study, while the third one was from a published paper that we re-analyzed here. The aim of the three language-production tasks was to investigate which visual cues trigger the contrast perception heuristic. More specifically, we tested three predictions: (i) color contrast, not scene variation, triggers color overspecification; (ii) color contrast must be detected in the referential domain (rather than across trials) to trigger overspecification, and (iii) the overspecification of atypical colors depends on color contrast.

Experiment 1 compared the role of contrast perception vs scene variation in triggering color overspecification and other redundant modifiers. Then, a re-analysis of the results of Long et al. (2020) examined what type of color contrast perception triggers redundant color modification: color contrast across monochrome displays of different colors, or color contrast within a given display. Experiment 2 investigated whether contrast perception is necessary to trigger overspecification of atypical colors (a well-documented finding in this literature), or unexpected colors are even mentioned in monochrome displays.

Finally, the last experiment in the study aimed to investigate the nature of the perceptual contrast heuristic. A general assumption that is often mentioned as an explanation for color overspecification is that *color is a visually salient property* (e.g., Pechmann, 1989; Belke & Meyer, 2002; Belke, 2006; Engelhardt et al., 2006; Arts et al., 2011b; Koolen et al., 2013; van Gompel et al., 2019; Degen et al., 2020). Despite the widespread reliance on visual salience to explain people's frequent use of color adjectives, psycholinguistic and computational studies on reference production normally work with an intuitive, non-technical notion of visual salience. Koolen et al. (2013), for example, argue that color is a

visually salient property that immediately grabs speakers' attention, such that they produce color adjectives without making sure that color rules out competitors. However, regarding color contrast, such a definition does not explain whether the color of a referent is more salient in polychrome displays than in monochrome displays (see Fig. 1). In other words, does visual salience explain why people produce more redundant color adjectives in polychrome displays than in monochrome displays? Experiment 3 addressed this question in a timed color-naming task, where we removed all informativity and discriminability demands, and asked participants to name the color of a target shape in polychrome and monochrome displays.

Unlike other psycholinguistic studies, our experiments do not directly build on each other. However, all four experiments explore the conditions under which perceptual contrast triggers overspecification, together comprising the first investigation of perceptual contrast as a visual heuristic. To foreshadow our results, we observed that contrast perception triggers overspecification of both color and other perceptual adjectives, and appears to be different from a mere strategy of 'ease of production' (MacDonald, 2013), supporting the view that contrast perception is a learned visual heuristic in the formulation of referential expressions.

2. Experiment 1

Koolen et al. (2013) hypothesized that speakers would be more likely to overspecify the color of a referent when scene variation is high than when it is low, with scene variation operationalized as the number of dimensions along which the objects in a scene differ. More recently, Degen et al. (2020) simulated the experimental conditions in Koolen et al. (2013) and showed how a *Rational Speech Act model* based on Frank and Goodman's (2012) but with continuous semantics predicted an analogous effect of scene variation.

Koolen et al. (2013) and Degen et al. (2020) tested and confirmed their hypothesis in a series of language-production experiments using displays of furniture to elicit referential expressions. In low-variation displays, furniture varied along three dimensions: type, size and orientation; whereas in high-variation displays, furniture varied along four dimensions: type, size, orientation and color. While the results of Koolen et al. (2013) and Degen et al. (2020) were taken to support the scene-variation hypothesis, their experimental design suffers from a confound that leaves open an alternative, more parsimonious interpretation of their results: because the low-variation scenes were monochrome and the high-variation scenes were polychrome, the differential rates of color modification observed in the two conditions may result from the absence and presence of color contrast, rather than from the different levels of scene variation (e.g., whether the objects varied in size or orientation). In our view, the perception of color contrast works as a visual heuristic that triggers the use of color adjectives. Therefore, according to the contrast perception hypothesis, the results of Koolen et al. (2013) and Degen et al. (2020) are an effect of the polychrome and monochrome displays used in the high- and low-variation conditions, rather than an effect of scene variation.

To test these two explanations of what triggers color overspecification, Experiment 1 pitched the scene variation hypothesis against the contrast perception heuristic. We elicited referential expressions in monochrome displays with high scene-variation, which included geometrical figures that varied along four dimensions: shape, size, border weight and border type; and in polychrome displays with low scene-variation, which included geometrical figures that varied

along two dimensions: shape and color (see Fig. 1).² According to the scene variation hypothesis, the monochrome displays should elicit higher color overspecification rates because of their higher scene-variation (e.g., 'the blue star'), whereas according to the contrast perception heuristic, the monochrome displays should not elicit the use of color adjectives, but rather the mention of those properties that are contrastive in the display (i.e. size, border weight or border type; e.g., 'the small star with a thick border'). Thus, while the scene variation hypothesis predicts that color adjectives will be overspecified more often in high-variation scene than in low-variation scenes, the perceptual contrast hypothesis makes a twofold prediction: the perception of color contrast will trigger color overspecification in the polychrome condition, while the contrasting properties detected in the monochrome condition (i.e. size, border type and border weight) will elicit the redundant use of the corresponding adjectives. Importantly, the use of modification would be redundant in both the polychrome and the monochrome conditions (i.e. the target had no competitors in either type of display, so there was no need to describe it).

2.1. Methods

2.1.1. Participants

Thirty-one undergraduate students at University College London took part in the study for monetary compensation. All students were native speakers of English and reported having normal color vision. For a sensitivity analysis, see [Supplementary Materials](#).

Ethics approval for the experiment was obtained from the Ethics Review Panel at UCL. All participants signed a consent form prior to performing the task.

2.1.2. Materials and procedure

Twenty displays of 9 different geometrical shapes were designed such that each shape appeared in one cell of a 3×3 grid (see Fig. 1). A total of 11 possible shapes (arrow, circle, cross, cylinder, heart, oval, pentagon, rectangle, square, star and triangle) and 9 possible colors (blue, brown, green, gray, orange, pink, purple, red and yellow) were randomly combined to create each of the displays. Displays were either polychrome (10 trials), in which all shapes were a different color, or monochrome (10 trials), in which all shapes were the same color. Monochrome displays were designed with high scene-variation such that figures varied by shape (9 different shapes per display), size (big vs small), border weight (thin vs thick) and border type (continuous vs discontinuous); whereas polychrome displays were designed with low scene-variation such that figures only varied by shape (9 different shapes per display) and color (9 different colors per display). The shapes presented in each display were always different, so a bare definite description (e.g., 'the star') would provide sufficient information to identify the target. The colors of the 9 shapes in the polychrome displays were also different.

It is important to bear in mind that the contrast manipulation in the Monochrome condition was not expected to be as visually salient as the color contrast used in the Polychrome condition. As we have earlier acknowledged, color is a visually salient property that is likely to be overspecified more frequently than other properties in contrastive environments. Thus, while the perceptual contrast heuristic predicts that size, border type and border weight should be overspecified in the Monochrome condition, that tendency is likely to be weaker than the tendency to overspecify color in the Polychrome condition. Future studies should try to match the visual salience of different contrastive

² We adopt Koolen et al.'s (2013) notion of scene variation as the number of dimensions along which the objects in a scene may differ, with shape, size, border weight, border type and color counting as five separate dimensions, each having different values. For other formalizations of scene variation, see Davies and Katsos (2013), Gatt et al. (2017) and Degen et al. (2020).

properties in monochrome and polychrome displays.

Participants were randomly assigned to one of two trial-block orders: polychrome-monochrome or monochrome-polychrome. We used a block design to avoid carry-over effects across trials. Prior to commencing the task, participants were instructed to sit beside and behind the experimenter and were given a print-out of numbered empty grids, one for each trial. On each grid, an X marked the position of the target on the experimenter's computer screen. The position of the target changed with each trial. Participants were told that all the shapes were different in each display (which rendered the use of modification redundant) and they had to ask the experimenter to click on the target shape. Participants were asked to avoid using coordinates (e.g., 'top left').

Participants' responses were audio recorded for transcription and coding purposes. Only responses including both a modifier and a noun (e.g., 'the small square') were considered overinformative. Responses were also coded for modifier type (Color vs Other, with 'Other' comprising descriptions of size, border weight and border type, as manipulated in high scene-variation trials). Responses could be coded as including Color and Other overspecification, if both types of modifiers were produced in the same trial (e.g., 'The small blue star'). Data was collected by one of the authors.

2.2. Results

First, we evaluated the scene variation hypothesis (i.e. high scene-variation displays should elicit more color overspecification than low scene-variation displays) using mixed effect logistic regression predicting trial-level color modification (modified = 1, unmodified = 0) with a fixed effect for scene variation level (High vs Low), by-participant random intercepts and slopes for scene variation level, and by-item random intercepts (i.e. the maximal random effect structure for the data as items were not repeated across scene variation conditions). Scene variation was sum coded. We did not include presentation block in the model as it did not improve model fit. All models in the paper were fitted using the brms software package (Bürkner, 2017) for Bayesian regression models in R (R Team, 2017) and Stan (Stan Development Team, 2018). We diagnose significance as an effect with credible intervals excluding 0. All data and analysis code can be found on OSF (<https://osf.io/z86vr/>).

Since the scene variation hypothesis only predicts the use of color adjectives (Degen et al., 2020; Koolen et al., 2013), we limited this analysis to the rates of color modification observed in the two experimental conditions. Contrary to this hypothesis, we observed more color overspecification in low scene-variation (polychrome) displays than in high scene-variation (monochrome) displays (see Fig. 2 and Table 1).

To evaluate the contrast perception hypothesis, we conducted a mixed effect logistic regression predicting over-modification (1 = modified, 0 = unmodified), with fixed effects for modifier type (Color vs Other), display type (Monochrome vs Polychrome), presentation block (First vs Second), and the interaction between modifier and display type, random intercepts and slopes for participants and random intercepts and slopes (as possible) for items (i.e. the maximal random effect structure). Display and modifier type were sum coded. As predicted by the contrast perception hypothesis, there was a significant interaction such that there was more color overspecification in polychrome displays than in monochrome displays and there was more overspecification of 'other' types of modifiers (i.e. size, border weight and border type) in monochrome displays than in polychrome displays (see Fig. 3 and Table 2).

The results of Experiment 1 offer support to the contrast perception hypothesis, with participants overspecifying color more often in polychrome displays than in monochrome displays. More importantly, participants often overspecified the size and border type of the shapes in the monochrome condition, but not their color, further supporting the view that it is the perception of contrast which triggers overspecification of the relevant adjective – rather than scene variation triggering color

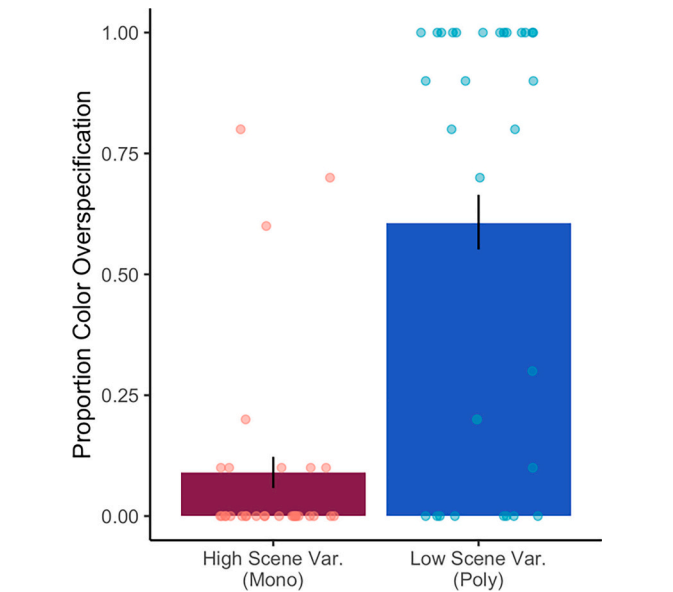


Fig. 2. Mean proportion of color overspecification in Experiment 1 as aggregated by display type to test the scene variation hypothesis. Line ranges reflect 95% bootstrapped confidence intervals and points reflect participant means.

Table 1
Coefficient estimates for the mixed effect model in Experiment 1 testing the scene variation hypothesis.

	Estimate	Est. Error	1-95% CI	u-95% CI
Intercept	−2.02	1.12	−4.40	−0.08
Scene-variation	−4.13	0.93	−6.23	−2.56

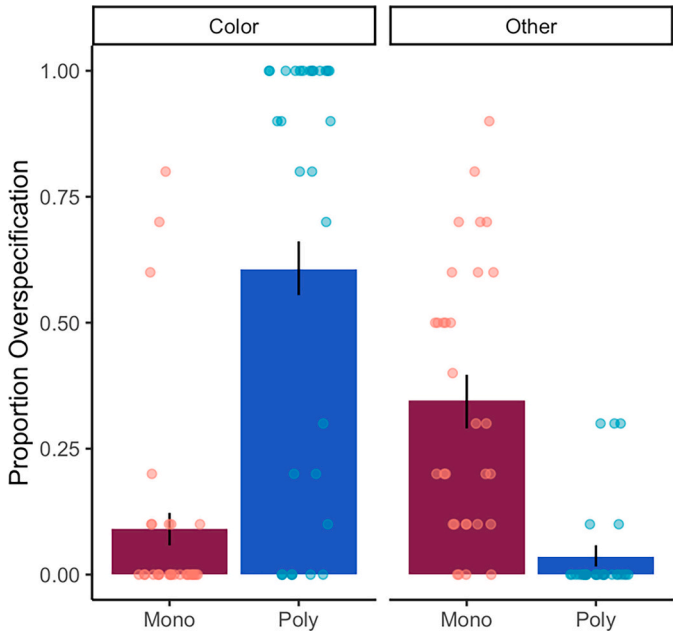


Fig. 3. Mean proportion of color overspecification vs 'other' overspecification (i.e. size, border weight and border type) in Experiment 1 as aggregated by display and modifier type to test the contrast perception hypothesis. Line ranges reflect 95% bootstrapped confidence intervals and points reflect participant means.

Table 2

Coefficient estimates for the mixed effect model in Experiment 1 testing the contrast perception hypothesis.

	Estimate	Est. Error	1-95% CI	u-95% CI
Intercept	-3.14	0.74	-4.80	-1.85
Display Type	-0.69	0.51	-1.71	0.35
Modifier Type	0.80	0.65	-0.47	2.08
Display*Modifier	-3.35	0.60	-4.67	-2.31
Block	0.01	0.76	-1.53	1.50

overspecification by default.

3. Re-analysis of the results of Long, Rohde, and Rubio-Fernandez, 2020

According to the contrast perception heuristic, the perception of color contrast may trigger color overspecification. However, we hypothesize that for this heuristic to be efficient, color contrast should be perceived in the visual context where speaker and listener will make and resolve reference, respectively. For example, in the Monochrome condition of Experiment 1, color changed across trials such that in one display, all the shapes were blue, but in the next display, all the shapes were red, and then green. This kind of color contrast is highly perceptible, but it would not be an efficient heuristic if it triggered color overspecification because both reference production and reference resolution would effectively take place in the absence of color contrast. Thus, if the contrast perception heuristic must apply in a given visual display, it follows from this hypothesis that interspersing multicolor fillers in a block of monochrome trials should increase the use of redundant color adjectives relative to a 'pure' monochrome trial block. In this view, filler trials should trigger the use of color adjectives, which may prime participants to continue using color on subsequent monochrome trials (see Fig. 4). This kind of manipulation was used in a recent study by Long et al. (2020), and we reanalyzed some of their data to further test the contrast perception hypothesis.

3.1. Methods

3.1.1. Participants

Data reported in Long et al. (2020) were re-analyzed for the purpose of this study. In the original study, language production data from 200 native English speakers aged 19–82 with normal color vision and hearing were collected. For the present study, data from the 60 youngest adults in the original sample (ages 19–31) were retained for further analyses.

Ethics approval for the experiment was obtained from the Ethics Review Panel at University of Edinburgh (where the original study was conducted). All participants signed a consent form prior to performing the task.

3.1.2. Rationale for participant inclusion criteria

The aim of the study by Long et al. (2020) was to investigate the

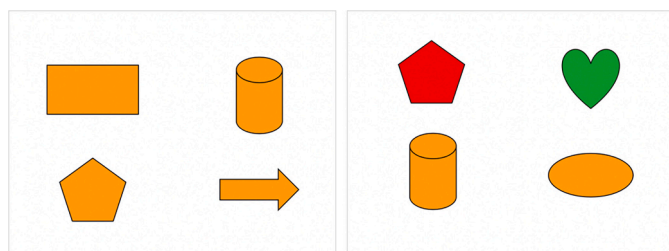


Fig. 4. Sample monochrome (critical) and multicolor (filler) displays from Long et al. (2020).

potential effects of age and cognitive abilities on communicative strategies related to color overspecification. The results of the study revealed a clear difference between younger and older adults' use of color overspecification: In the version of the task without fillers, younger adults overspecified less than older adults, suggesting that they were more inclined towards a strategy of brevity. However, in the version of the task with fillers (which were a hybrid of polychrome and monochrome displays, with 4 shapes of 3 colors), younger adults began overspecifying color at a similar rate as older adults. The influence of age on color overspecification is clear in Fig. 5, which shows color overspecification in monochrome trials across age and version. Based on these results, and the well-documented tendency for adults to become more verbose with age (Gold, Arbuckle, & Andres, 1994), we focused our analysis on the subset of monochrome trials from the youngest 30 participants (aged 19–31) in each version of the task (monochrome with fillers vs monochrome without fillers), for a grand total of 60 participants. In this way, we tried to prevent the influence of age from concealing the effect of version on color overspecification. Moreover, the age of the participants in this re-analysis was also comparable to that of the participants in Experiment 1.

3.1.3. Materials and procedure

Similar to Experiment 1, participants were presented with 20 critical trials: 10 monochrome displays and 10 polychrome displays, in one of two trial-block orders, and asked to indicate the target to the experimenter. However, in this experiment, scene variation was not manipulated (all shapes were the same size and were borderless) and there were only 4 different shapes in each display (see Fig. 4). A total of 10 possible shapes (arrow, circle, cylinder, heart, oval, pentagon, rectangle, square star and triangle) and 9 possible colors (blue, brown, gray, green, orange, pink, purple, red and yellow) were randomly combined to create each of the displays. Crucial to the aim of the experiment, participants were randomly assigned to one of two task versions: one including critical trials only and another one including critical trials plus 40 fillers interspersed. The critical trials were the same in both versions of the task, including the random order in which they were presented. Fillers were a hybrid of polychrome and monochrome displays such that there were 4 shapes of 3 colors (i.e. one of the colors was repeated). The target was a unique color in half of the filler trials, and in the other half, the

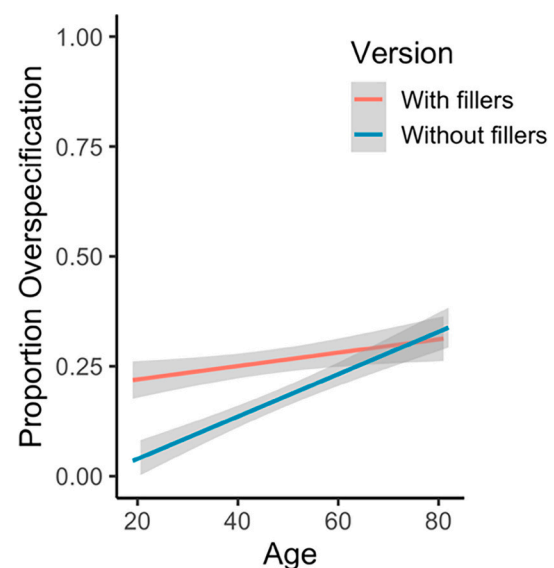


Fig. 5. Regression lines showing color overspecification in monochrome trials by age in each version of the task by Long et al. (2020). Data from the 60 youngest participants were re-analyzed for the present study. The shaded band surrounding each of the regression lines represents a 95% confidence region for the regression fit.

target was the same color as another shape in the display. Of interest was whether the presence of the multi-colored fillers increased color overspecification in monochrome trials as predicted by the color perception hypothesis. Data was collected by one of the authors.

3.2. Results

Using logistic mixed effects regression, we modelled the binary outcome variable of presence/ absence of color overspecification (color = 1, bare noun = 0) with Version (With Fillers vs Without Fillers) as the fixed effect. The model was fit with the maximal random effect structure for participants and items.

As predicted, results revealed an effect of Version on color overspecification, with an increase in color overspecification of about 25% in the version of the task with multicolored fillers (see Table 3; Fig. 6). The 30 participants who took part in the task with fillers overspecified the color of the target in 28% of filler trials. Crucially for our investigation, 11 of those 30 participants overspecified color in at least one monochrome trial, with the first instance of color modification in a critical trial always occurring after the use of color in a filler trial. This pattern of results suggests that it was the perception of color contrast in mutlicolor displays which triggered the use of color modification, with this referential strategy then carrying over to some of the monochrome displays.³

The results of this re-analysis offered support to the contrast perception hypothesis, as participants produced more redundant color adjectives in monochrome trials when multicolored fillers were interspersed in the trial block than when the same block included monochrome trials only. Crucially, color overspecification was first triggered by the perception of color contrast in multicolor displays, with the tendency to mention color then carrying over to monochrome trials in the same block. This confirms that the perception of color contrast across monochrome trials does not trigger color overspecification, suggesting that the perceptual contrast heuristic is an efficient referential strategy that elicits color modification when it would be most efficient; that is, when color contrast is detected in the same visual context where the speaker is producing a referential expression and the listener has to identify the intended referent. We interpret the subsequent mention of color in monochrome trials as a ‘color priming effect’, whereby participants continue using redundant color adjectives even in the absence of color contrast (for participants’ pervasive use of color adjectives across trials, see also Tarenskeen et al., 2015).

4. Experiment 2

The second experiment in the study tried to replicate a well-documented finding in the referential communication literature: speakers are more likely to overspecify atypical colors (e.g., ‘yellow pig’) than typical colors (e.g., ‘yellow banana’) or even variable colors (e.g., ‘yellow notebook’; Sedivy, 2003; Westerbeek, Koolen, & Maes, 2015; Rubio-Fernandez, 2016; Degen et al., 2020). Rubio-Fernandez (2016) interpreted this effect as a form of cooperative behavior, since not

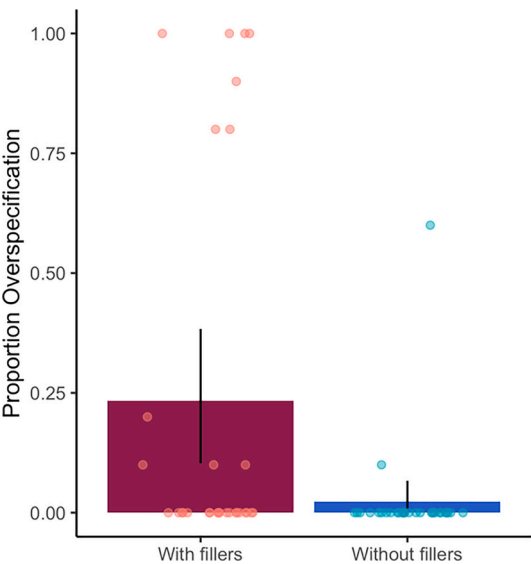


Fig. 6. Mean proportion of color overspecification in monochrome trials from each version of the task (with and without multicolor fillers) in Long et al. (2020). Line ranges reflect 95% bootstrapped confidence intervals and points reflect participant means.

mentioning atypical colors would lead the listener astray (e.g., they would look for a yellow fruit when hearing ‘banana’; see Huettig & Altmann, 2011). Degen et al. (2020), however, remain neutral on this point, leaving open the possibility that “the benefit for listeners and the salience for speakers might simply be a happy coincidence and speakers might not, in fact, be designing their utterances for their addressees” (p. 617).

We investigated this question in Experiment 2, where we used the original polychrome displays in Rubio-Fernandez (2016), including objects of atypical, typical and variable colors, and a monochrome version including the same kinds of objects but in a single color in each display (see Fig. 7). In order to replicate the original results, speakers should mention atypical colors more frequently than typical and variable colors in polychrome displays. More critical to our research question, if speakers mention atypical colors because they are unexpected and therefore salient, then they should do so at comparable rates in polychrome and monochrome displays since a yellow pig, for example, is equally odd in either type of display. However, if speakers are being cooperative and mention atypical colors to spare the listener unnecessary effort in their visual search, then they should do so only in polychrome displays since the color of the target (be that typical or atypical) is uninformative in monochrome displays. It therefore follows from the perceptual contrast hypothesis that speakers should not use redundant color adjectives in the monochrome condition, even when the target’s color is atypical.

4.1. Methods

4.1.1. Participants

Thirty-eight participants were recruited for the study using Amazon’s Mechanical Turk and directed to complete the task on Qualtrics. Workers were restricted to those located within the United States (according to their IP address) and with a 95% HIT approval rate after they had completed more than 500 HITs. Half the participants completed the polychrome version of the task and half completed the monochrome version. Participants were also asked to provide their native language with the understanding that this would not affect their eligibility to participate in the study, in order to minimize deception. Those who were not native English speakers were excluded from analyses (6 in total). The final pool consisted of 16 participants in the polychrome condition and

Table 3
Coefficient estimates for the re-analysis of the results of Long et al. (2020) testing what type of color contrast perception triggers redundant color modification.

	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept	0.23	0.06	0.12	0.34
Without Fillers	−0.21	0.08	−0.35	−0.06

³ As expected, an analysis including the 200 participants from the original study (including younger and older adults) did not reveal an effect of Version on color overspecification, since the influence of Age clearly concealed the effect of contrast (see Fig. 5).

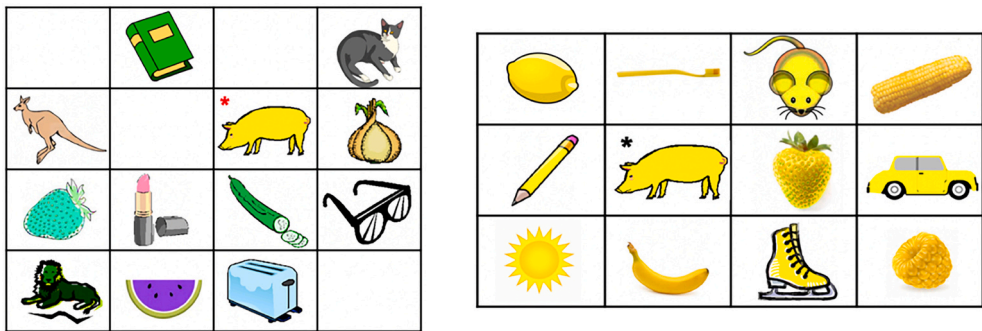


Fig. 7. Sample polychrome and monochrome displays from Experiment 2.

16 in the monochrome condition. Due to a programming error, one item for one participant in the polychrome condition was not displayed (resulting in 0.3% data loss). For a sensitivity analysis, see [Supplementary Materials](#).

Ethics approval for the experiment was obtained from the Institutional Review Panel at MIT. All participants completed a consent form prior to performing the task.

4.1.2. Materials and procedure

For the Polychrome condition, the materials from [Rubio-Fernandez \(2016; Experiment 2\)](#) were adapted to be used online. Thirty displays of 12 animals, fruits, and artifacts were designed such that each image appeared in one cell of a 4 × 4 grid (see [Fig. 7](#)). In each display, 4 animals and fruits were presented in typical colors (e.g., a yellow banana), another 4 were presented in an atypical color (e.g., a yellow pig) and 4 artifacts were presented in variable colors (e.g., a blue notebook). In each display, a red asterisk marked the target object. The materials from the original Polychrome condition were adapted for the Monochrome condition such that all the objects in a display were the same color as the target. Thirty displays were created in the following colors: blue, green, orange, red and yellow. The grids in the Monochrome condition were 3 × 4 (i.e. they did not contain any empty cells) to create a stronger monochrome effect.

To increase cooperative behavior, participants were told that they were paired with a virtual partner and had to instruct their partner to click on the target object marked by the red asterisk. Participants were discouraged to use coordinates (e.g., ‘top left’) when referring to the target, as the virtual partner’s display would be a scrambled version of theirs. Participants were shown one display per trial and had to complete the instruction ‘Click on the...’ written under the display.

Following [Rubio-Fernandez \(2016\)](#), participants were told that a pilot study had revealed that participants sometimes miscommunicated because the person giving the instruction did not notice that there were two objects of the same type (e.g., a small box and a big box) and failed to produce an appropriate description. In the actual experiment, the displays never contained two objects of the same type, or objects of different sizes. This ‘cautionary instruction’ was only intended to make participants behave cooperatively, after [Rubio-Fernandez \(2016\)](#) observed that it had a significant effect on participants’ performance.

The objects in the displays were always different from one another, so a one-word continuation (e.g., ‘Click on the... elephant’) would be sufficient to communicate with the presumed virtual partner. Of interest was whether in the Polychrome condition the rate of color overspecification was higher for atypically-colored objects (e.g., yellow pig) when compared to typically-colored objects (e.g., yellow banana) or variably-colored objects (e.g., blue notebook), as it had been observed in the original study by [Rubio-Fernandez \(2016\)](#). It was also of interest whether these effects were modulated by the number of colors in the display, such that color typicality did not affect the rates of color overspecification in the Monochrome condition. Only responses that included both a color and a noun (e.g., “yellow pig”) were considered

instances of overspecification.

4.2. Results

[Fig. 8](#) shows the proportion of color overspecification by display and color type. To estimate the influence of Display (Monochrome vs Polychrome) and Color Type (Atypical, Typical, Variable) on overspecification, we conducted a mixed effect logistic regression model with fixed effects for display, color type and their interaction, and random intercepts and slopes for participant and item (i.e. the maximal random effect structure). Since we predicted an increase in overspecification in polychrome/atypical-color trials, we dummy coded the predictors with polychrome/atypical color as the reference level. Replicating [Rubio-Fernandez \(2016\)](#), we observed an increase in overspecification in polychrome/atypical-color trials compared to polychrome/typical-color and polychrome/variable-color trials (see [Table 4](#)). As predicted by the contrast perception hypothesis, we found an overall decrease in overspecification in monochrome displays relative to polychrome displays. More directly relevant to our research question, we observed a decrease in overspecification in monochrome/atypical-color trials compared to polychrome/atypical-color trials, while we did not observe an effect of color typicality across the monochrome displays. The results of Experiment 2 therefore replicate the

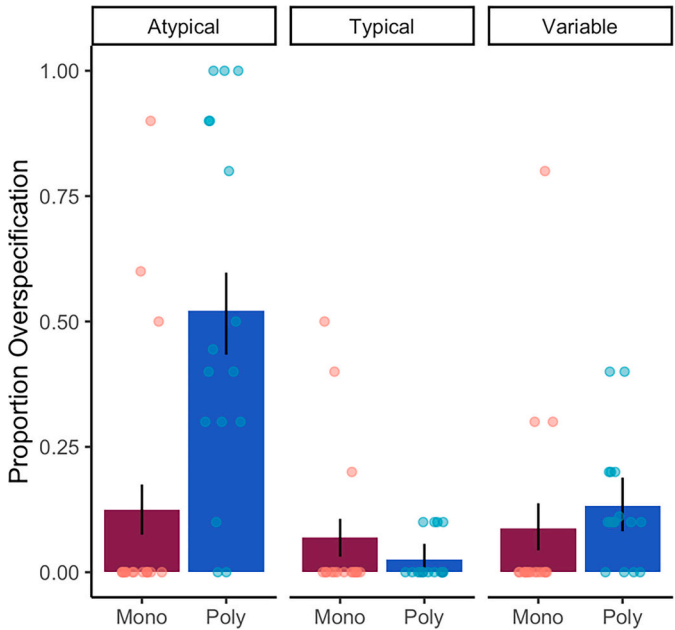


Fig. 8. Mean proportion of color overspecification by display type and target color in Experiment 2. Line ranges reflect 95% bootstrapped confidence intervals and points reflect participant means.

Table 4
Coefficient estimates for the mixed effect model in Experiment 2.

	Estimate	Est. Error	l-95% CI	u-95% CI
Intercept	0.14	0.74	−1.35	1.59
Target Typical	−8.69	3.14	−16.74	−4.62
Target Variable	−3.87	1.16	−6.42	−1.89
Monochrome	−19.86	8.68	−42.00	−7.65
Mono*T Typical	2.99	5.95	−10.52	13.47
Mono*T Variable	1.41	3.33	−6.42	6.85

results of Rubio-Fernandez (2016) in the Polychrome condition, and support the predictions of the contrast perception hypothesis in the Monochrome condition.

When participants in Experiment 2 had to refer to targets of atypical colors, they overspecified their color in polychrome displays, but not when referring to the same targets in monochrome displays. These results support the view that overspecifying atypical colors is a cooperative and efficient communicative strategy (Rubio-Fernandez, 2016, 2019), rather than a form of egocentric behavior whereby speakers mention atypical colors simply because they are unexpected, and hence visually salient. More generally, the results of Experiment 2 support the view that color contrast perception works as a visual heuristic, triggering the use of color modification in polychrome displays.

5. Experiment 3

In the reference production literature, color is generally treated as an inherently salient property (e.g., Pechmann, 1989; Belke & Meyer, 2002; Belke, 2006; Engelhardt et al., 2006; Arts et al., 2011b; Koolen et al., 2013; van Gompel et al., 2019; Degen et al., 2020). However, the working definition of visual salience as a perceptual quality that captures people's attention does not allow us to determine whether color contrast makes color more or less salient, relative to a monochrome display. Thus, the unique color of a target in a polychrome display may pop-out against all other shapes, making it visually salient. On the other hand, the uniform color of a monochrome display may make the target color more salient by virtue of not having any color competitors.

The last experiment in the study tried to investigate these two possibilities in a color-naming task, in which participants had to name the color of a target shape as fast as possible and their voice onset time (VOT) was measured. Crucially, color naming does not involve any pragmatic consideration, such as informativity (e.g., are there any target competitors in the display?) or discriminability (e.g., are there any other blue shapes in the display?). Thus, color naming should allow us to explore the extent to which color contrast facilitates the mention of color (perhaps because color is perceived as more salient in polychrome displays than in monochrome displays), or it hinders its production relative to monochrome displays.

If color naming is easier in polychrome displays than in monochrome displays, it could be argued that there is no such thing as a contrast perception heuristic (that is used for feedforward audience design), and that speakers are simply driven by *ease of production* (MacDonald, 2013). Alternatively, if color naming is easier in monochrome displays than in polychrome ones, then that finding would be more in line with the view that relying on color contrast to produce efficient color modification is a pragmatic strategy that is learned through communicative experience and readily available in reference production (see Ferreria, 2019).

5.1. Methods

5.1.1. Participants

Fifty-four native English speakers with normal color vision and hearing were recruited from the MIT Experiment Pool to participate in the study for monetary compensation. Participants included students and local residents in the Cambridge area. One participant was excluded

from analysis as data was only recorded for half the trials due to microphone error. For a sensitivity analysis, see [Supplementary Materials](#).

Ethics approval for the experiment was obtained from the Institutional Review Panel at MIT. All participants signed a consent form prior to performing the task.

5.1.2. Materials and procedure

Similar to Experiment 1, 32 displays of 9 different geometrical shapes were designed such that each shape appeared in one cell of a 3 × 3 grid (see Fig. 9). Nine possible shapes (circle, cross, heart, pentagon, oval, rectangle, square, star and triangle) and 8 possible colors (blue, brown, green, orange, pink, purple, red and yellow) were combined to create each of the displays. Before the shapes appeared on the screen, a black box appeared around one cell in the grid to direct participants' attention towards the location where the target shape would appear 1 s later. Displays were either polychrome (16 trials), in which all shapes were a different color, or monochrome (16 trials), in which all shapes were the same color, with target colors counterbalanced across trials. Crucially, the same colors were used in the polychrome and monochrome conditions, so a significant difference between the two conditions could not be due to the specific colors used in each condition. There were 2 polychrome and 2 monochrome warm-up trials at the beginning of the experiment, which were excluded from analyses. Trials were presented in the same random order to all participants. The presentation computer produced a tone when the figures appeared on the grid, which was then used to calculate VOT.

Participants were instructed to name the color of the target shape as quickly as possible. In order to measure VOT, a microphone was placed 15 cm from the participant's mouth, as per Hattori, Sumita, and Taniguchi (2014). Using Psychophysics Toolbox Version 3 and MATLAB, sound drivers were set to low latency, with a 44,100 Hz frequency and 2 sound channels for stereo capture of voice onset. The resulting WAV files were then analyzed using Audacity. With the waveform amplitude in Audacity set to 1.0, we measured the first time the amplitude of the sound wave was above 0 dB. We then verified that the segment was indeed the onset of speech rather than an artifact by creating a looped clip of the point in question with 0.2 s before and after onset and listening to the segment with 0.75× playback speed. Once the point of speech onset was established, it was recorded in milliseconds. The tone marking the presentation of the shapes was measured using the same methodology. Then, the tone onset time was subtracted from the onset of speech in order to determine VOT. Due to a microphone error, data were not recorded for 12 trials, resulting in 0.71% data loss. Data was collected by one of the authors.

5.2. Results

Fig. 10 shows the VOT for monochrome and polychrome displays. To estimate the influence of Display (Monochrome vs Polychrome) on VOT, we conducted a mixed effect linear regression model with a fixed effect for Display, a random intercept and slope for participants and a random intercept for items (i.e. the maximal random effect structure). We dummy coded Monochrome as the reference level. We found that polychrome displays have a significant delay in VOT compared to monochrome displays ($\beta = 164$, 95%CI = [114–215]).

The results of Experiment 3 suggest that naming the color of a referent is easier in monochrome displays than in polychrome displays, resulting in shorter voice onset times. Color naming may be harder in polychrome displays because of the lexical competition from the other colors in the display. These results confirm that when people overspecify color in polychrome displays and refrain from doing so in monochrome displays, it is not because they are guided by ease of production. A reasonable hypothesis as to why speakers use redundant color adjectives in polychrome displays (where color is harder to name), but not in monochrome displays (where color is easier to name), is that color is a

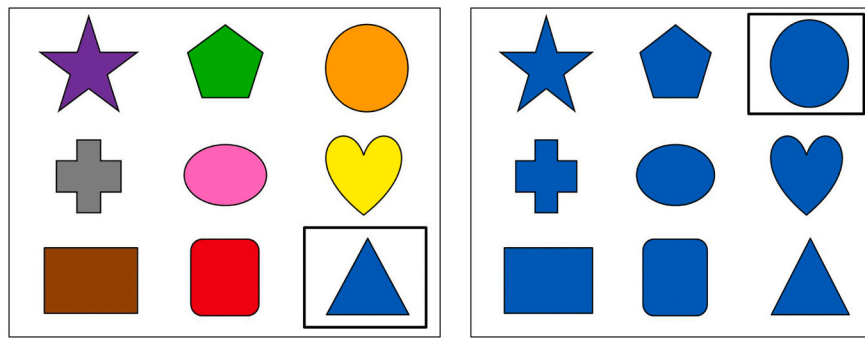


Fig. 9. Sample polychrome and monochrome displays from Experiment 3. The frame around the target appeared 1 s before the shapes so that participants would be ready to name the color of the target as fast as possible.

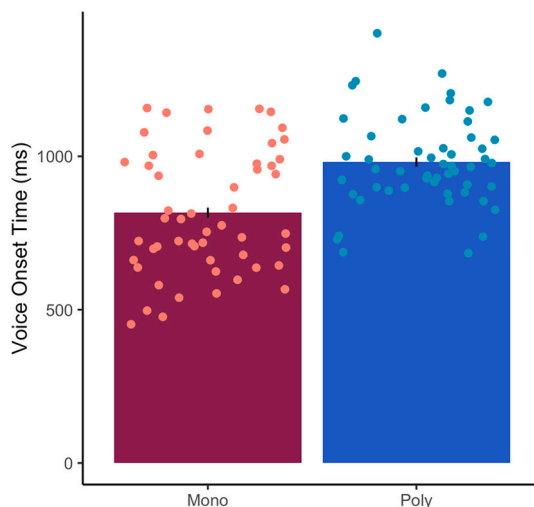


Fig. 10. Mean voice onset time by display type in Experiment 3. Line ranges reflect 95% bootstrapped confidence intervals and points reflect participant means.

more efficient visual cue for the listener in polychrome displays (Rubio-Fernandez, 2016, 2019). Thus, these findings offer indirect support for the view that the contrast perception heuristic is a learned contextual cue (see Ferreria, 2019), and not a speaker bias to minimize production effort (Zipf, 1949).

6. General discussion

In a series of language-production experiments, we tested the hypothesis that contrast perception works as a visual heuristic in the formulation of referential expressions. In this view, speakers can rely on visual heuristics such as the perception of color contrast or the relative density of a display to formulate efficient referential expressions, rather than scanning the visual context searching for a referent's competitors. Whereas referential overspecification has sometimes been interpreted as a form of egocentric behavior whereby speakers fail to tailor their message to the needs of their interlocutors (Engelhardt et al., 2006, 2011), a number of psycholinguistic studies have shown that redundant referential expressions can facilitate the listener's visual search for a referent (Arts et al., 2011b; Mangold & Pobel, 1988; Paraboni et al., 2007; Paraboni & Van Deemter, 2014; Rubio-Fernandez, 2021; Sonnenschein & Whitehurst, 1982; Tourtouri et al., 2019). Thus, producing a redundant referential expression by relying on a perceptual contrast heuristic would be an efficient communicative strategy that may benefit both the speaker and the listener, rather than either of the two (see Arnold, 2008).

The results of Experiment 1 offer support to the contrast perception hypothesis, with participants producing redundant color adjectives more often when referring to targets in polychrome displays than monochrome displays. In the latter, participants overspecified properties of the target shapes that were contrastive in the display (namely, size, border type and border weight), further supporting the contrast perception hypothesis. In addition, a re-analysis of the results of Long et al. (2020) confirmed that color contrast works as a visual heuristic when it is detected in the same visual context where the speaker is formulating a referential expression. Thus, while the color of the displays changed from one trial to the next in the monochrome condition, the perception of color contrast across trials did not trigger color overspecification. Instead, it was the perception of color contrast in the multicolor filler trials which triggered the production of redundant color adjectives almost 30% of the time, priming the use of color in subsequent monochrome trials. These results confirm that the contrast perception heuristic is an efficient audience design strategy that is triggered in visual contexts where color has more discriminatory power.

Experiment 2 replicated the well-documented finding that speakers tend to overspecify atypical colors more often than typical or variable colors (Degen et al., 2020; Rubio-Fernandez, 2016; Sedivy, 2003; Westerbeek et al., 2015). However, previous studies had not established whether this effect is a speaker-internal process driven by surprise (e.g., the unexpected identification of a pink banana), or a form of audience design whereby the speaker avoids having the listener look for an object by its typical color (e.g., a yellow fruit when the target banana is pink). In this study, atypical colors were mentioned more than typical and variable colors in polychrome displays, replicating previous findings. However, atypical colors were not mentioned in monochrome displays, similar to what was observed with typical and variable colors. Our results therefore support the view that overspecifying atypical colors is a cooperative referential strategy that prevents the listener from mistakenly looking for a prototypical object (Rubio-Fernandez, 2016). If speakers mention atypical colors more often than typical colors simply because they are unexpected, they should do so in both polychrome and monochrome displays, since atypical colors are not any less atypical in monochrome displays – just less *efficient* as a visual cue for the listener. Participants' mention of atypical colors in polychrome displays but not in monochrome displays therefore challenges the view that atypical color overspecification is a form of egocentric behavior driven only by surprise, and suggests instead that it is a form of audience design.

Finally, the results of Experiment 3 cast further doubt on the egocentric view of color overspecification since participants were faster to name the color of a target if it appeared in a monochrome display rather than a polychrome one. If participants in referential communication tasks were insensitive to their interlocutor's needs, they would produce redundant color adjectives when color is easier to name. However, the results of Experiments 1 and 2 revealed the opposite pattern of results, with participants preferring to overspecify color in polychrome displays, when it is more efficient for the listener's visual

search. Therefore, the results of the three experiments plus a re-analysis reported in this study support the view that color overspecification is a cooperative and efficient referential strategy (Rubio-Fernandez, 2016, 2019, 2021; Long et al., 2020; Rubio-Fernandez et al., 2021). The results of Experiment 3 can also be interpreted as indirectly supporting the view that speakers' reliance on contrast perception to produce modified descriptions is a learned contextual cue (Ferreria, 2019), rather than a speaker bias to minimize production effort (Zipf, 1949).

The contrast perception hypothesis may be relevant to another well-documented finding in the referential communication literature: the effect of *perceptual grouping* on color overspecification. Koolen et al. (Koolen, 2019; Koolen et al., 2016; Koolen & Fliervoet, 2017; Koolen, Houben, Huntjens, & Krahmer, 2014) observed that speakers are more likely to overspecify the color of a referent when it is closer to its competitors than when they are further apart in the visual scene. Koolen and colleagues interpret these results as an effect of competitor distance on the perceived relevance of the various competitors in a scene, with closer competitors being perceived by speakers as more relevant than more distant competitors. The results of Experiments 1, 2 and 3 suggest that perceptual grouping might be related to contrast perception, which is likely to be stronger if the objects are closer in space. Future studies should investigate this possibility to further understand what triggers the production of redundant color adjectives across different visual contexts.

In addition, given the artificial nature of the visual displays used in this study, future work should also try to replicate the present findings with more naturalistic stimuli. Research on perceptual grouping has indeed employed images from more naturalistic scenes (e.g., Koolen, 2019; Koolen et al., 2016), moving away from the artificial displays that are often used in psycholinguistic studies on reference production. Extending the investigation of perceptual heuristics from laboratory experiments to more naturalistic situations is important because realistic scenarios are rarely monochrome, leaving open the question of how speakers detect different types of visual contrast and how they encode them in their referential expressions. Future psycholinguistic studies should therefore rise to this methodological challenge to better understand the connection between perceptual contrast and overspecification.

7. Concluding remarks on audience design

The main pragmatics debate in reference production has revolved around two theoretical positions (Arnold, 2008; Davies & Arnold, 2019): some researchers see reference production as determined by speaker-internal processes (e.g., Pechmann, 1989), while others construe it as a collaborative process that involves audience design (e.g., Clark & Wilkes-Gibbs, 1986). While there is no principled reason why reference production could not involve both speaker-internal processes and audience design, these two views have often been treated as mutually exclusive, with speaker-internal processes acting as a default.

In a series of recent studies comparing the production of referential expressions by English and Spanish speakers, Rubio-Fernandez (2016, 2019; see also Goudbeek & Krahmer, 2012; Wu & Gibson, 2021; Kachakeche et al., 2021) observed that English speakers produced redundant color adjectives more often than Spanish speakers. The authors interpreted this difference as evidence of efficient cooperation between speakers and listeners, with eye-tracking evidence of incremental processing confirming that prenominal color adjectives are a more efficient visual cue than postnominal color adjectives (Rubio-Fernandez et al., 2021; see also Rubio-Fernandez & Jara-Ettinger, 2020). In a recent discussion of the results of these studies, Fukumura and Carminati (2021) challenged this conclusion on the grounds that speaker-internal processes could explain this cross-linguistic difference (Pechmann, 1989): English speakers produce more redundant color adjectives because they are less certain about the presence of competitors at the time when they would have to encode an adjective, as opposed to Spanish speakers, who have more time to decide whether

color is necessary to uniquely identify the referent.

Participants in Rubio-Fernandez (2016, 2019, Rubio-Fernandez et al., 2020) were presented with simple displays of four different geometrical shapes in four different colors, intended to facilitate scene recognition at a glance (Oliva, 2005).⁴ In addition, Rubio-Fernandez (2019) timed the presentation of the displays to 1.5 s so that Spanish speakers did not have more time to decide whether they should mention the color of the target. Most importantly, participants were explicitly told in the instructions that *all shapes were different in each display* (ruling out the possibility of competitors). Despite all these provisions, Fukumura and Carminati (2021) rejected the interpretation that audience design affects the use of redundant color adjectives in different syntactic positions because speaker-internal processes could explain these cross-linguistic differences (Pechmann, 1989).

It is undeniable that the incrementality of language production gives speakers of languages with postnominal modification more time to decide whether to encode an adjective or not. However, the fact that speaker-internal processes are at play in reference production should not be used as evidence against audience design (for a parallel argument, see MacDonald, 2013). The results of another two experiments in Rubio-Fernandez (2019; Experiments 1a and 1b) speak to this issue: when participants had to ask the experimenter to click on a target in a dense display of nine different shapes, they used redundant color adjectives 75% of the time, whereas they did so less than 50% of the time in displays of four shapes. However, when participants were presented with the same displays but the target was marked for both interlocutors, the production of redundant color adjectives dropped to zero in displays of four shapes and 4% in displays of nine shapes. That is, participants mentioned color when the Experimenter had to identify the target following their description, but not when the target was part of their common ground.

If color overspecification was only driven by the speaker's uncertainty about the presence of competitors, these experiments should have revealed comparable results since the displays were identical across the two tasks. However, the fact that participants adopted different referential strategies depending on the epistemic state of their interlocutor suggests that audience design is compatible with speaker-internal processes: referring to a target in a dense display of objects increases speakers' uncertainty about the presence of competitors relative to sparser displays, but the use of modification can nonetheless be modulated by audience design (e.g., by a cooperative intention to facilitate the listener's harder visual search in a dense display; see Clarke et al., 2013).

To conclude, the view that perceptual contrast works as a visual heuristic in the formulation of referential expressions is in line with Ferreria (2019) feedforward audience design, according to which speakers can make use of contextual cues prior to the onset of utterance production and rely on previously learned strategies that facilitate communication (see also Dale & Viethen, 2009; Jaeger & Ferreira, 2013; Kurumada & Jaeger, 2015; Van Deemter et al., 2012). The view that not all forms of audience design are cognitively costly undermines the widespread assumption that speakers are being egocentric when they rely on low-level cues to formulate referential expressions. Here we have argued that speakers' reliance on visual heuristics may be efficient for both speakers and listeners, who can optimize their use of cognitive resources when engaging in referential communication.

Acknowledgements

This research was supported by a FRIPRO Grant from The Research

⁴ Behavioral studies on visual cognition have shown that observers can recognize the gist of a scene in less than 100 ms (Oliva, 2005; Potter, 1976). These findings cast doubt on the argument that English speakers presented with displays of 4 simple geometrical shapes of different colors could not tell that all the shapes were different by the time they produced a description of the target.

Council of Norway (*Vision shapes language*; Ref. 275505) awarded to PRF. The authors gratefully acknowledge this funding. Special thanks to Herb Clark for a helpful discussion of this and other works on referential communication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2021.104879>.

References

- Adams, R. C., & Chambers, C. D. (2012). Mapping the timecourse of goal-directed attention to location and colour in human vision. *Acta Psychologica*, 139, 515–523.
- Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and cognitive processes*, 23, 495–527.
- Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011b). Overspecification facilitates object identification. *Journal of Pragmatics*, 43, 361–374.
- Arts, A., Maes, A., Noordman, L. G., & Jansen, C. (2011a). Overspecification in written instruction. *Linguistics*, 49, 555–574.
- Belke, E. (2006). Visual determinants of preferred adjective order. *Visual Cognition*, 14, 261–294.
- Belke, E., & Meyer, A. S. (2002). Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during “same”-“different” decisions. *European Journal of Cognitive Psychology*, 14, 237–266.
- Bramão, I., Reis, A., Petersson, K. M., & Faisca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica*, 138, 244–253.
- Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, 54, 592–609.
- Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28.
- Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers’ subsequent pronunciations. *Journal of Memory and Language*, 89, 68–86.
- Clark, H. H. (2006). Pragmatics of language performance. In L. R. Horn, & G. Ward (Eds.), *The handbook of pragmatics* (pp. 365–382). Oxford: Blackwell.
- Clark, H. H., & Bangerter, A. (2004). Changing ideas about reference. In I. Noveck, & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 25–49). London: Palgrave Macmillan.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge: Cambridge University Press.
- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13, 259–294.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39.
- Clarke, A. D., Elsnar, M., & Rohde, H. (2013). Where’s Wally: The influence of visual salience on referring expression generation. *Frontiers in Psychology*, 4.
- Dale, R., & Viethen, J. (2009). Referring expression generation through attribute-based heuristics. In *Proceedings of the 12th European workshop on natural language generation (ENLIG)* (pp. 58–65) (Athens, Greece).
- Davidoff, J. (1991). *Cognition through color*. Cambridge, MA: MIT Press.
- Davies, C., & Arnold, J. E. (2019). Reference and informativeness: How context shapes referential choice. In C. Cummins, & N. Katsos (Eds.), *Handbook of experimental semantics and pragmatics* (pp. 474–493). Oxford, UK: Oxford University Press.
- Davies, C., & Katsos, N. (2013). Are speakers and listeners ‘only moderately Gricean’? An empirical response to Engelhardt et al. (2006). *Journal of Pragmatics*, 49, 78–106.
- Davies, C., & Kreysa, H. (2017). Looking at a contrast object before speaking boosts referential informativeness, but is not essential. *Acta Psychologica*, 178, 87–99.
- Degen, J., Hawkins, R. X., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is rational: A Bayesian approach to “overinformative” referring expressions. *Psychological Review*, 127(4), 591–621.
- Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the “model of the listener”. In D. Napoli, & J. Kegl (Eds.), *Bridges between psychology and linguistics*. New York: Academic Press.
- Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean maxim of quantity? *Journal of Memory and Language*, 54, 554–573.
- Engelhardt, P. E., Demiral, Ş. B., & Ferreira, F. (2011). Over-specified referring expressions impair comprehension: An ERP study. *Brain and Cognition*, 77, 304–314.
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87, 327.
- Evans, J. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, 75, 451–468.
- Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. In, 109. *Proceedings of the National Academy of Sciences of the United States of America* (pp. 17897–17902).
- Ferreria, V. S. (2019). A mechanistic framework for explaining audience design in language production. *Annual Review of Psychology*, 70, 29–51.
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998.
- Fukumura, K., & Carminati, M. N. (2021). Overspecification and incremental referential processing: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Gatt, A., Krahmer, E., Van Deemter, K., & van Gompel, R. P. (2017). Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science*, 41, 1457–1492.
- Gegenfurtner, K., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, 10, 805–808.
- Gold, D. P., Arbuckle, T. Y., & Andres, D. (1994). Verbosity in older adults. In M. L. Hummert, J. M. Wiemann, & J. F. Nussbaum (Eds.), *Interpersonal communication in older adulthood: Interdisciplinary theory and research* (pp. 107–129). Sage Publications, Inc.. <https://doi.org/10.4135/9781483326832.n6>
- Goudbeek, M., & Krahmer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering, and referential overspecification. *Topics in Cognitive Science*, 4(2), 269–289.
- Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Speech acts* (pp. 41–58). New York: Academic Press.
- Hattori, M., Sumita, Y. I., & Taniguchi, H. (2014). Measurement of voice onset time in maxillectomy patients. *The Scientific World Journal*, 2014, 925707.
- Huetig, F., & Altmann, G. T. (2011). Looking at anything that is green when hearing “frog”: How object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology*, 64, 122–145.
- Jaeger, T. F., & Ferreira, V. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences*, 36, 359.
- Jara-Ettinger, J., & Rubio-Fernandez, P. (2021). *The social basis of referential communication: Speakers construct reference based on listeners’ expected visual search*. Preprint available at <https://psyarxiv.com/fzuvh/>. under review.
- Kachakeche, Z., Futrell, R., & Scontras, G. (2021). Word order affects the frequency of adjective use across languages. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43, 43.
- Kahneman, D. (2003). A perspective on judgement and choice. *American Psychologist*, 58, 697–720.
- Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30, 1–45.
- Kennedy, C., & McNally, L. (2005). Scale structure and the semantic typology of gradable predicates. *Language*, 81, 345–381.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11, 32–38.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89, 25–41.
- Koolen, R. (2019). On visually-grounded reference production: Testing the effects of perceptual grouping and 2D/3D presentation mode. *Frontiers in Psychology*, 10.
- Koolen, R., & Fliervoet, Y. (2017). Eye movements during reference production: Testing the effects of perceptual grouping on referential overspecification. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th annual meeting of the cognitive science society* (pp. 2439–2444) (London).
- Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The effect of scene variation on the redundant use of color. *Cognitive Science*, 37, 395–411.
- Koolen, R., Houben, E., Huntjens, J., & Krahmer, E. (2014). How perceived distractor distance influences reference production: Effects of perceptual grouping in 2D and 3D scenes. In *Proceedings of the 36th annual meeting of the cognitive science society* (pp. 2507–2512) (Québec City).
- Koolen, R., Krahmer, E., & Swerts, M. (2016). How distractor objects trigger referential overspecification: Testing the effects of visual clutter and distance. *Cognitive Science*, 40, 1607–1647.
- Kurumada, C., & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, 83, 152–178.
- Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, 124, 762–794.
- Long, M., Horton, W. S., Rohde, H., & Sorace, A. (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition*, 170, 25–30.
- Long, M., Rohde, H., & Rubio-Fernandez, P. (2020). The pressure to communicate efficiently continues to shape language use later in life. *Scientific Reports*, 10(1), 1–13.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 226.
- Maes, A., Arts, A., & Noordman, L. (2004). Reference management in instructive discourse. *Discourse Processes*, 37, 117–144.
- Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology*, 7, 181–191.
- Matthews, D., Butcher, J., Lieven, E., & Tomasello, M. (2012). Two-and four-year-olds learn to adapt referring expressions to context: Effects of distractors and feedback on referential communication. *Topics in Cognitive Science*, 4, 184–210.
- Matthews, D., Lieven, E., & Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify referents for others: A training study. *Child Development*, 78, 1744–1759.
- Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, 66, B25–B33.
- Milli, S., Lieder, F., & Griffiths, T. L. (2017). When does bounded-optimal metareasoning favor few cognitive systems? In, 31. *Proceedings from AAAI-17: The 31st Association for the Advancement of artificial intelligence conference on artificial intelligence* (pp. 4422–4428). Palo Alto, CA: Association for the Advancement of Artificial Intelligence Press.

- Morgan, J. L., & Meyer, A. S. (2005). Processing of extrafoveal objects during multiple-object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 428–442.
- Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13, 329–336.
- Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.
- Paraboni, I., & Van Deemter, K. (2014). Reference and the facilitation of search in spatial domains. *Language, Cognition and Neuroscience*, 29, 1002–1017.
- Paraboni, I., Van Deemter, K., & Masthoff, J. (2007). Generating referring expressions: Making referents easy to identify. *Computational Linguistics*, 33, 229–254.
- Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, 27, 89–110.
- Potter, M. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509–522.
- Rubio-Fernandez, P. (2016). How redundant are redundant colour adjectives? An efficiency-based analysis of colour overspecification. *Frontiers in Psychology*, 7.
- Rubio-Fernandez, P. (2019). Overinformative speakers are cooperative: Revisiting the Gricean Maxim of Quantity. *Cognitive Science*.
- Rubio-Fernandez, P. (2021). Color discriminability makes over-specification efficient: Theoretical analysis and empirical evidence. *Humanities and Social Sciences Communications*, 8(1), 1–15.
- Rubio-Fernandez, P., & Jara-Ettinger, J. (2020). Incrementality and efficiency shape pragmatics across languages. *Proceedings of the National Academy of Sciences*, 117, 13399–13404.
- Rubio-Fernandez, P., Mollica, F., & Jara-Ettinger, J. (2021). Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. *Journal of Experimental Psychology: General*.
- Sedivy, J. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, 32, 3–23.
- Sedivy, J. C. (2005). Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. In J. Trueswell, & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions* (pp. 345–364). Cambridge, MA: MIT Press.
- Sonnenschein, S., & Whitehurst, G. J. (1982). The effects of redundant communications on the behavior of listeners: Does a picture need a thousand words? *Journal of Psycholinguistic Research*, 11, 115–125.
- Stan Development Team. (2018). *RStan: The R interface to Stan* (R package version 2.17.2.). Retrieved from <https://mc-stan.org>.
- Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: Saliency, absoluteness, and consistency [Special Issue Models of Reference]. *Frontiers in Psychology*, 6.
- Team, R. (2017). *R: A language and environment for statistical computing [software tool]*. Vienna, Austria: R Foundation for Statistical Computing.
- Tourtouri, E. N., Delogu, F., Sikos, L., & Crocker, M. W. (2019). Rational over-specification in visually-situated comprehension and production. *Journal of Cultural Cognitive Science*, 3(2), 175–202.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Van Deemter, K., Gatt, A., Van Gompel, R., & Krahmer, E. (2012). Toward a computational psycholinguistics of reference production. *Topics in Cognitive Science*, 4, 166–183.
- van Gompel, R. P. G., van Deemter, K., Gatt, A., Snoeren, R., & Krahmer, E. J. (2019). Conceptualization in reference production: Probabilistic modeling and experimental testing. *Psychological Review*, 126(3), 345–373.
- Viethen, J., & Dale, R. (2009). Referring expression generation: What can we learn from human data?. In *Proceedings of the 2009 workshop on the production of referring expressions: Bridging the gap between computational and empirical approaches to reference*. Amsterdam, The Netherlands.
- Viethen, J., van Vessel, T., Goudbeek, M., & Krahmer, E. (2017). Color in reference production: The role of color similarity and color codability. *Cognitive Science*, 41, 1493–1514.
- Wardlow-Lane, W., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1466–1481.
- Westerbeek, H., Koolen, R., & Maes, A. (2015). Stored object knowledge and the production of referring expressions: The case of color typicality. *Frontiers in Psychology*, 6.
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31, 183–194.
- Wu, S., & Gibson, E. (2021). *Word Order Predicts Cross-linguistic Differences in the Production of Redundant Color and Number Modifiers*. Cognitive Science.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press.