

RUNNING HEARD: OVERSPECIFICATION AND INCREMENTAL PROCESSING

Overspecification and incremental referential processing: An eye-tracking study

Kumiko Fukumura

Maria Nella Carminati

Division of Psychology, Faculty of Natural Sciences

University of Stirling

© 2021, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy without authors' permission. The final article will be available, upon publication, via its DOI: [10.1037/xlm0001015](https://doi.org/10.1037/xlm0001015)

Address correspondence to:

Kumiko Fukumura

Psychology, Faculty of Natural Sciences

University of Stirling

Stirling FK9 4LA

UK

kumiko.fukumura@stir.ac.uk

Abstract

Using eye-tracking, we examined if over-specification hinders or facilitates referent selection, and the extent to which this depends on the properties of the attribute mentioned in the referring expressions and the underpinning processing mode. Following spoken instructions, participants selected the referent in a visual display while their eye movements were monitored. The referring expressions were presented either simultaneously with the displays, so the attributes could be incrementally processed *in sequence*, or before the display presentation, so the attributes could be processed *in parallel from the outset of search*. Experiment 1 showed that when the attributes were processed incrementally, how quickly an earlier-mentioned attribute discriminated determined whether a late-mentioned, over-specified attribute contributed to discrimination: When color was mentioned first and is fully discriminating, the referent was selected fast regardless of the second-mentioned pattern, whereas when pattern was mentioned first and is fully discriminating, the second-mentioned color facilitated discrimination. Experiment 2 found that under incremental processing, color mention after a fully discriminating pattern increased fixations but delayed referent selection relative to a pattern-only description; under parallel processing, however, color mention immediately eliminated alternatives and sped up referent selection. Experiment 3 showed that pattern mention after a fully discriminating color delayed referent selection and tended to reduce fixations relative to a color-only description in both processing modes. Hence, additional attributes can speed up referent selection but only when they can discriminate much faster than alternative attributes mentioned in a more concise description, *and* critically, when they can be used *early* for referent search.

Keywords: over-specification; language comprehension; eye-tracking; serial processing; parallel processing

It is often assumed that redundancy is detrimental to communication. Grice (1975) argued that providing more information than is required is not only “a waste of time” but also violates a *Maxim of Quantity* that speakers should be optimally informative (e.g., “Make your contribution as informative as required”; “Do not make your contribution more informative than is required”). Inspired by Grice’s maxims, some theoretical models have proposed that in referential communication, speakers would only mention attributes that are strictly necessary to identify an intended referent, thereby avoiding not only *under-specification* but also *over-specification* (Dale, 1989; Frank & Goodman, 2012). Under these models, speakers should be more likely to say *green bow* rather than *spotty bow* when *green*, but not *spotted*, discriminates the referent from alternatives. This should be reversed when pattern, not color, identifies the referent uniquely. If speakers select less informative modifiers or do not use modifiers at all, that will result in under-specification. If they select both (e.g., *green spotty bow*), that will lead to over-specification. A prediction that could be derived from these models is therefore that listeners should find discriminating descriptions with the fewest possible attributes easier to process than over-specified descriptions.

The question is, though, whether over-specification indeed hampers communication. In referential communication, over-specification is ubiquitous; speakers often mention attributes that are not required for unique identification (e.g., *large red car* when *large car* is sufficient for identification) (e.g., Deutsch & Pechmann, 1982; Pechmann, 1989) and this happens even the attributes do not discriminate (e.g., Belke & Meyer, 2002; Gatt, Krahmer, Van Deemter, & Van Gompel, 2016; Koolen, Krahmer, & Swerts, 2015). Moreover, over-specification does not seem to occur randomly. Some attributes are more likely to be over-specified than others; for instance, color tends to be over-specified more often than other attributes such as pattern (e.g., Fukumura, 2018; Tarenskeen, Broersma, & Geurts, 2015) and size (e.g., Belke & Meyer, 2002; Engelhardt & F.Ferreira, 2014; Gatt et al., 2016; Pechmann, 1989), though the rates of over-specification vary, depending on the typicality of color for given objects (Degen, Hawkins, Graf, Kress, & Goodman, 2020; Rubio-Fernández, 2016; Sedivy,

2003; Westerbeek, Koolen, & Maes, 2015). Likewise, an attribute that fully discriminates (i.e., ruling out all referential alternatives) is more likely to be over-specified (Van Gompel, Gatt, Van Deemter, Snoeren & Krahmer, 2019) and placed in an earlier position (Fukumura, 2018) than an attribute that only partially discriminates (i.e., ruling out some alternatives but not all).

Hence, some researchers have argued that over-specification occurs because it helps rather than hinders successful information transfer. In this view, some attributes are more likely to be over-specified than others because those attributes are more likely to help comprehension. For instance, color over-specification occurs frequently because salient attributes like color may facilitate referent identification (e.g., Arts, Maes, Noordman, & Jansen, 2011a; Degen et al., 2020; Rubio-Fernandez, 2016, 2019; Sonnenschein & Whitehurst, 1982). Also, speakers not only over-specify but also place in an early position more salient attributes such as color, particularly when they are fully discriminating, perhaps because the earlier placing of those attributes is likely to speed up referent identification (Fukumura, 2018). Moreover, findings suggest that speakers tend to over-specify more often in tasks that emphasize comprehension accuracy (Arts et al., 2011b). However, other researchers proposed that over-specification results from production-internal constraints. For instance, speakers over-specify because they cannot determine which attribute discriminates before starting to speak (Pechman, 1989), and/or because they tend to choose more preferred or salient attributes (Dale & Reiter, 1995; Schriefers & Pechmann, 1988).

Crucially, these debates rely on evidence in production. To date, relatively few studies have examined the impact of over-specification on comprehension, and the existing evidence in comprehension is rather mixed. Some findings suggest that over-specification is detrimental to referential processing (Davies & Katsos, 2013; Engelhardt, Bailey, & F.Ferreira, 2006; Engelhardt, Demiral, & F.Ferreira, 2011; Sedivy, Tanenhaus, Chambers, & Carlson, 1999), whereas other findings suggest that over-specification can facilitate referent identification (Arts et al., 2011a;

Tourtour, Delogu, Sikos, & Crocker, 2019). Thus, the goal of the current study was to determine whether and under what circumstances over-specification hinders or facilitates comprehension.

Most previous studies have examined the over-specification of a single prenominal modifier. The assumption has been that listeners expect speakers to produce modifiers optimally informatively, by avoiding over-specification as well as under-specification, and use such expectations to identify the referent. Consistent with this, in an eye-tracking study by Sedivy et al. (1999), participants fixated the referent faster when a color adjective (e.g., *yellow*) contrasted two category exemplars (a yellow comb co-present with a pink comb and a yellow bowl) than otherwise (a yellow bowl co-present with a pink comb and a yellow comb). The adjective alone did not discriminate fully (i.e., there were always two yellow objects in the context in the above example), so the findings were taken to indicate that the listeners' expectations for optimal informativeness helped them resolve the referential ambiguity that resulted from the adjective. However, in Sedivy et al., there were always preceding instructions where another color adjective was used contrastively to identify one of two category exemplars (*Touch the pink comb* in the above example); hence it is possible that the contrastive use of the adjective in these preceding instructions might have primed the contrastive interpretation in the subsequent request (*Now touch the yellow bowl/comb*).

Indeed, other findings suggest that listeners are not always sensitive to the over-specification of a modifier. It has been reported that adjectives such as color do not necessarily trigger contrastive interpretations (Sedivy, 2003; Fukumura & Van Gompel, 2017). In fact, unlike *relative* or *dimensional* attributes such as size, color is said to be an *absolute* or *partially dimensional* attribute (Bierwisch, 1987; Martin, 1969): Whether a car is large or not depends on its comparison with another car, whereas the color of a car does not necessarily depend on comparison with another car, though of course, color intensity can vary (i.e., some cars can be redder than other cars and speakers can differentiate the differences, Monroe, Hawkins, Goodman, & Potts, 2017). Fukumura and Van Gompel (2017) showed in reading that the over-specification of a size adjective (*the small towel* in

the context of only one towel), but not the over-specification of a color adjective (*the white towel* in the context of only one towel), immediately delayed reading times for the noun phrase. This led them to argue that over-specified size adjectives hinder comprehension, not because they violate readers' general expectations that speakers use modifiers optimally informatively, but because readers fail to identify the referential alternative implied by the meaning of the size adjective.

Moreover, if it is the noun that determines whether the modifier is over-specified *and* if the modifier precedes the noun in the description, the effect of modifier over-specification emerges at the noun rather than at the modifier itself. Hence, we do not know if the modifier is over-specified until we encounter the noun, and critically, any cost associated with modifier over-specification could be overridden by the potential processing advantage resulting from the higher discriminability of the noun. For example, in response to color-modified nouns as in *Find the blue ball*, Tourtouri et al. (2019) reported more fixations to the referent following the noun *ball* when the display contained only one ball (thus *blue* was over-specified) rather than two balls (*blue* was minimally specified), which was interpreted as indicating an advantage of over-specification. But this effect was significant only by items and, as suggested above, it might have been driven by the discriminability of the noun: Whereas the color modifier itself did not discriminate fully in any of the conditions (i.e., there was more than one blue object in the display), the noun discriminated the referent fully in the over-specified context, but not in the minimally specified context.

Finally, and importantly, in these previous studies, over-specified descriptions were not compared with more concise descriptions with fewer attributes; hence the findings may not inform us about whether over-specification helps or hinders comprehension. When descriptions contain more attributes, they should take longer to process because of the length difference. Interestingly though, findings by Arts et al. (2011a) suggest that over-specified descriptions with more attributes could be advantageous relative to more concise descriptions with fewer attributes: The inclusion of additional attributes, as in *large round white button* or *round button at the top left* relative to *round*

button, led to faster referent identification. Critically, Arts et al. presented the referring expressions prior to the presentation of the visual displays. Response times were measured from the onset of the display, so it is possible that over-specified descriptions led to faster response times because the measure did not include the time taken to process the referring expressions. The question is whether over-specification would also help referent identification under different set-ups. The experiments we report below investigated just that.

Current study

We thus investigated whether and under what circumstances over-specification might hinder or facilitate referential processing. The study had two overarching aims. The first aim was to test predictions derived from findings in production. Specifically, Fukumura (2018) showed that speakers over-specified color more often, and they tended to place color in an earlier position than pattern particularly when color was fully discriminating (i.e., ruling out all alternatives) rather than when it was only partially discriminating (i.e., ruling out some but not all alternatives). As discussed in Fukumura (2018), these tendencies can be explained in terms of production-internal processing constraints: Salience and discriminability jointly determine the availability of an attribute in the speaker's message. That is, speakers preferentially produce more available attributes earlier, so that they can start speaking sooner without needing to keep readily available information in memory (cf. Bock, 1982; Belke, 2006; Pechmann, 1989). In the current study, we ask whether the over-specification preferences observed in production in Fukumura (2018) might also reflect comprehension preferences because speakers favor referring expressions that can facilitate not just incremental production but also referent search in comprehension. If so, over-specified attributes should more likely help referent search if they are more salient and fully discriminating and importantly, if they are available *early* for referent search in comprehension.

Critically, the availability of attributes for referent search may depend on *how listeners process them*. The second aim of the current study was thus to investigate the listener-internal processing constraints that may modulate the impact of over-specification. Our starting point is the hypothesis that how listeners process over-specified attributes depends on *what has been said earlier* rather than *what is yet to be said*, at least under some circumstances. Specifically, one fundamental characteristic of language comprehension is that language comprehension proceeds incrementally (e.g., Altmann & Kamide, 1999; Marslen-Wilson & Tyler, 1980; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Research has shown that in the simultaneous presence of the visual display, the attributes mentioned in the referring expressions are processed immediately in sequence (e.g., Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Rubio-Fernández, Mollica, & Jara-Ettinger, 2020; Rubio-Fernández & Jara-Ettinger, 2020). For instance, in response to instructions with multiple adjectives such as *Touch the starred yellow square*, listeners in Eberhard et al. (1995) immediately launched their search for the referent based on the first-mentioned attribute alone: Fixations to the target were faster when the first adjective (*starred*) identified the referent uniquely, ruling out all referential alternatives that had no stars, than when the second adjective (*yellow*) identified the referent uniquely, ruling out alternatives that had different colours.

Note that incremental referential processing is possible only when listeners have immediate access to the referential context. However, speakers do not always refer to things that are in the listener's view at that point in time; hence, listeners sometimes need to search for the referent after having heard the referring expression. Spivey et al. (2001) presented participants with spoken queries (*Is there a red vertical?*) either in the presence of the visual display or prior to its presentation. In either case, a conjunction of two adjectives (*red, vertical*) uniquely identified the target since each referential alternative shared one attribute with the target; that is, some objects were red but not vertical, whereas other objects were not red but vertical. As the number of

referential alternatives increased, the response times to the spoken questions slowed down. This set size effect was smaller when the spoken descriptions were given in the presence of the visual display than prior to the visual display, and this happened both when the first modifier was color (*red vertical*) and when it was orientation (*vertical red*). The idea was that, when the visual display is concurrently available, linguistic processing and the visual search of the referent are rapidly integrated, such that listeners selectively attend to the subset of the referents with the feature denoted by the first modifier, thereby reducing the cost of the increased set size. This was not possible when the referring expressions were given prior to the visual display, because all attributes were available *in parallel* from the outset of the search.

Parallel processing may have made the search process somewhat inefficient in Spivey et al. (2001), because the two attributes, color and orientation, were both relatively salient, and this created competition between them. But it could be advantageous in other situations. In production experiments, language users do not recall sentences verbatim; they often “regenerate” sentences by placing the most salient or available information early (e.g., McDonald, Bock, & Kelly, 1993; Potter & Lombardi, 1990; Tanaka, Branigan, McLean, & Pickering, 2011). If listeners also exploit such “flexibility” during parallel processing, over-specified later-mentioned attributes could be used early when they are highly salient. That is, when the referring expressions have been heard prior to the search for the referent, all the attributes may be used in parallel; hence later-mentioned attributes could facilitate referent selection particularly if they are most salient and thus mapped onto the context faster. By contrast, when the referring expressions are incrementally processed in sequence, how listeners process later-mentioned attributes may be dependent on the properties of earlier-mentioned attributes. For instance, if the first attribute already rules out referential alternatives fully before listeners encounter a second attribute, any further information about the referent provided by this second attribute may not be of any use for referent search.

The current study thus focused on how over-specification affects referent search and the extent to which this depends on the properties of the attributes as well as the two processing modes we have just discussed. All our experiments employed the visual world paradigm, where participants were asked to select the referent using a computer mouse in response to spoken instructions (*Click on the ...*) while we recorded their eye movements to the visual displays and the timing of referent selection (mouse-clicking). Whilst the eye movement data provide a continuous measure of referential processing from its early onset (e.g., Tanenhaus et al., 1995), mouse-clicking latencies indicate the timing of the final interpretation. Experiment 1 examined how the salience and discriminability of attributes influence incremental referential processes and the way over-specification is processed. Experiments 2 and 3 examined whether having more attributes in a description would be advantageous or detrimental and if this is dependent on the processing mode.

Experiment 1

As discussed earlier, in production, highly salient attributes such as color tend to be over-specified more often and produced earlier than less salient attributes such as pattern particularly when they discriminate fully (i.e., ruling out all alternatives) rather than only partially (i.e., ruling out some but not all alternatives) (Fukumura, 2018). Experiment 1 examined whether and how incremental referential processes in comprehension might be jointly facilitated by the salience and discriminability of attributes. Specifically, although previous studies have reported that fully discriminating first adjectives result in fast referent identification (e.g., Eberhard et al., 1995), they have left open how the discriminability of an over-specified second adjective might facilitate referential processing and if this is dependent on the salience of the attributes.

One possibility is that under incremental processing, if the first adjective fully discriminates, an over-specified second adjective does not facilitate referential processing. That is, the referent is identified faster when the first adjective discriminates fully rather than partially, regardless of whether the second adjective discriminates, but the reverse does not hold: The discriminability of

the second adjective matters only when the first adjective does not discriminate fully, i.e., when the second adjective is *required* for full discrimination. Crucially, though, incremental referential processing may be driven not only by the discriminability of the attributes but also by their degree of salience. If the attribute of the first adjective is fully discriminating *and* highly salient, the referential alternatives may be ruled out *immediately* before the encounter of the second adjective; hence, whether this second adjective discriminates or not has no impact on discrimination. By contrast, if the first-mentioned attribute is *not* salient, even when it discriminates fully, listeners may not be able to rule out referential alternatives immediately before encountering the second adjective. In this case, the attribute of an over-specified second adjective may also contribute to discrimination. That is, under incremental processing, *how quickly* the first attribute discriminates determines whether an over-specified second attribute facilitates referent search.

Experiment 1 tested these hypotheses by presenting two-adjective descriptions in the simultaneous presence of the visual display, following previous findings that showed that the attributes are processed incrementally under such condition (e.g., Eberhard et al., 1995; Tanenhaus et al., 1995). We manipulated the salience of the adjectives, by contrasting color-pattern orders (*green spotty bow*) with pattern-color orders (*spotty green bow*), assuming that color is more salient than pattern (Fukumura, 2018; Tarenskeen et al., 2015). The discriminability of each adjective was varied as follows. In the *both fully discriminating* condition, both the first and the second adjectives were fully discriminating (**Fig.1A**). By contrast, in the *A1 partially discriminating* condition, the first adjective was only partially discriminating, whereas the second adjective was fully discriminating (**Fig.1B**). This was reversed in the *A2 partially discriminating* condition, where the first adjective was fully discriminating, and the second adjective was only partially discriminating (**Fig.1C**). We examined the impact of first adjective discriminability by comparing the *both fully discriminating* condition (Fig.1A) with the *A1 partially discriminating* condition (Fig.1B), given that these conditions only differ with respect to the discriminability of the first adjective; in both

conditions the second adjective was always fully discriminating. Likewise, the impact of second adjective discriminability was examined by comparing the *both fully discriminating* condition with the *A2 partially discriminating* condition (Fig.1C), given that these conditions differ only with respect to the second adjective discriminability; in both conditions the first adjective was always fully discriminating.

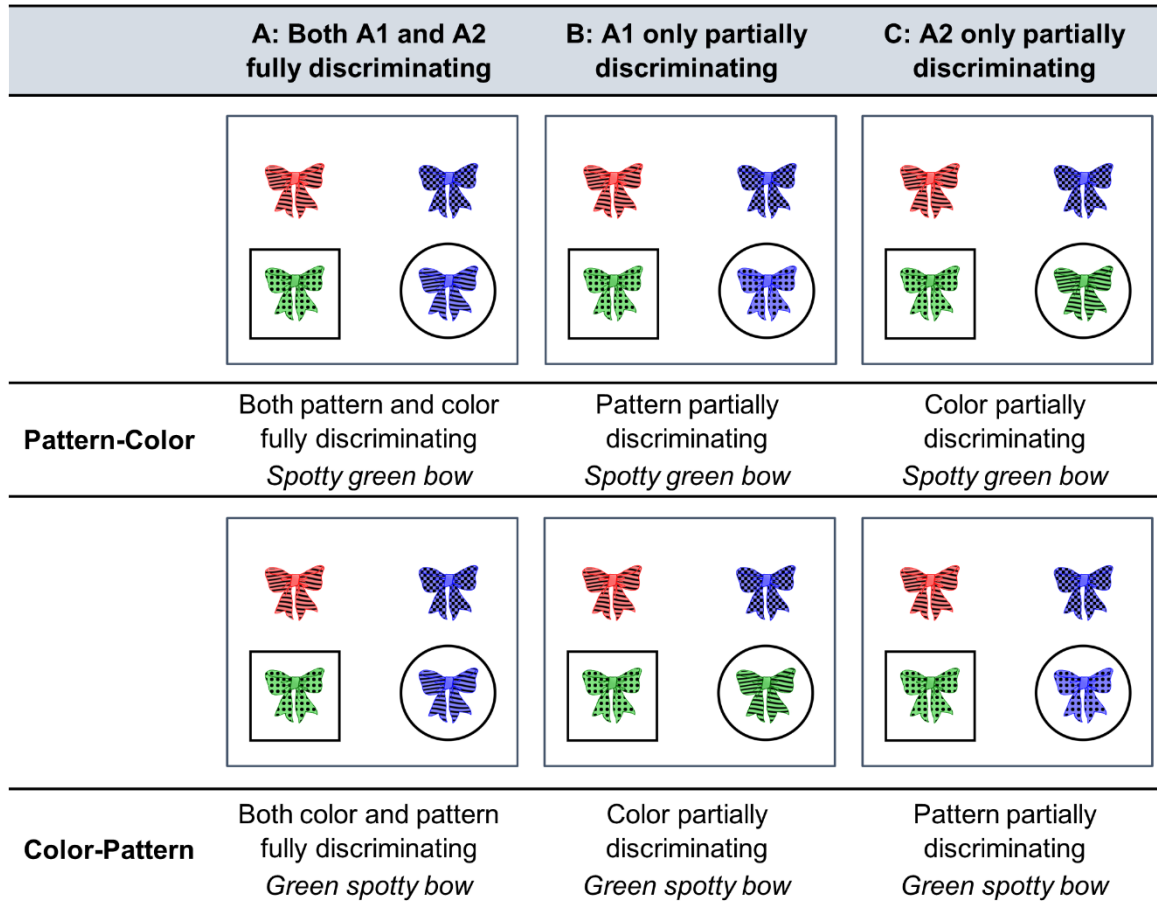


Fig.1. Example visual displays by condition in Experiment 1. A1 = First Adjective, A2 = Second adjective. Box = Target, Circle = Competitor.

If the adjectives are processed incrementally, the first adjective should direct the listener’s attention to the target more and faster when it is fully discriminating than otherwise, triggering more looks to the target and leading to faster referent selection performed via mouse-clicking in the *both fully discriminating* condition than in the *A1 partially discriminating* condition. If the incremental referential processes are facilitated by the salience of the attributes, the discriminability of the first

adjective should have larger impacts when it denotes color (in color-pattern orders) rather than pattern (in pattern-color orders), and the discriminability of the second adjective should also interact with adjective order. In color-pattern orders, if the first-mentioned color adjective fully discriminates, the referent should be identified equally fast, irrespective of whether the second-mentioned pattern adjective fully discriminates or not. By contrast, in pattern-color orders, even when the first-mentioned pattern adjective fully discriminates, the referential alternatives may not be ruled out immediately before the encounter of the second-mentioned color adjective. Hence, a fully discriminating second-mentioned color adjective might facilitate referential processing, increasing looks to the target and speeding up mouse-clicking in the *both fully discriminating* condition relative to the A2 *partially discriminating* condition.

Method

Participants

Forty-eight native speakers of British English aged between 17 and 30 were recruited from the University of Strathclyde student community in exchange for course credits. All participants reported to have no problem with processing color or pattern.

Materials

Forty-eight experimental items were created, each comprising three visual displays and two spoken instructions. Fig.1 shows example visual displays, each having four images of the same object in different colors and patterns.¹ The bows in the box are the targets and the bows in the circle are the competitors, while the remaining bows serve as distractors. We varied the competitor's similarity to the target object such that target and competitor differed in both color and pattern, or they shared either pattern or color. The distractor objects always differed from the target in both

¹ The object type is held constant across conditions, because the informativeness of a modifier is determined by the discriminability of another modifier unlike in the previous studies where the discriminability of the noun (object type) was varied.

color and pattern. The images were taken from Rossion and Pourtois (2004), to which we had applied one of three different colors (red, blue, green) and patterns (striped, spotted, chequered).

The instructions always started with the fragment *Click on the*, followed by two adjectives, denoting the target's color and pattern, and the final noun. We varied the order of the adjectives (pattern-color vs. color-pattern order) and cross-spliced each adjective across the different word order conditions. Each referring expression was combined with the three visual displays, resulting in the following three conditions for each adjective order; (1) *Both fully discriminating* condition, where both the first (A1) and second (A2) adjective identified the target uniquely (Fig.1A); (2) *A1 partially discriminating* condition, where the first adjective was only partially discriminating, whereas the second adjective was fully discriminating (Fig.1B) and; (3) *A2 partially discriminating* condition, where the first adjective was fully discriminating, whereas the second adjective was only partially discriminating (Fig.1C).

The instructions were recorded in stereo at 44.1 kHz by a native speaker of British English. There were three color adjectives (*red, blue, green*) and three pattern adjectives (*spotty, stripy, chequered*). **Table 1** reports the mean onsets of the first and second adjectives in the two instructions. Though some adjectives were longer than others², each adjective occurred in all conditions an equal number of times. Along with the 48 experimental items, 84 filler trials were created to discourage participants from adopting strategic responses. In these filler trials, color and pattern were non-discriminating features; the target was identified by its category ($n = 30$), size and category ($n = 27$), or size and direction ($n = 27$). We varied the number of category exemplars in the scene; in some filler trials all display objects belonged to different semantic categories, whilst in others they were of the same category. Some filler instructions mentioned two attributes, whereas others mentioned only one. In the filler trials each attribute was used contrastively so there was no over-specification. We also created 10 practice trials.

² *red* = 800 ms, *blue* = 800 ms, *green* = 700 ms, *spotty* = 1000 ms, *stripy* = 1100 ms, *chequered* = 850 ms

Table 1

Mean durations (ms) of the first and second adjectives

	<i>Onset</i>	<i>Offset</i>	<i>Duration</i>
“Click on the”	0	1300	1300 (0)
<i>Pattern-Color order</i>			
Pattern	1300	2283	983 (103)
Color	2283	3011	728 (70)
<i>Color-Pattern order</i>			
Color	1300	2067	767 (47)
Pattern	2067	3045	978 (105)

Note. Standard deviation in parentheses.***Design***

We used a 2 (adjective order: pattern-color vs. color-pattern) \times 3 (discriminability: both fully discriminating vs. A1 partially discriminating vs. A2 partially discriminating) repeated measures design, which resulted in the creation of six lists. Eight participants were randomly assigned to each list. Together with the 84 filler items, the 48 experimental items were randomly distributed across the six lists in a Latin square design, with each list containing only one version of each experimental item. Experimental trials were separated by at least one filler trial and the same color or pattern did not appear more than three times consecutively.

Apparatus and procedure

We used an SR Research EyeLink 1000 eye tracker. The fixations were sampled at 1000 Hz and the tracker had an angular resolution of .02 degrees. Participants were seated at a distance of 57cm from a 19-inch CRT display (1600 \times 1200 pixel resolution) running at a 100 Hz refresh rate. A chin rest was used to minimize participants' head movements. Before the initial trial, at least one calibration was carried out and further calibrations were performed throughout the experiment at regular predefined intervals and when required. Trials began with a drift correction, in which participants had to fixate on a dot, which appeared in the centre of the screen. Then a cross appeared for 2000 ms, and it was followed by the simultaneous presentation of the visual display with the four objects (Fig.1) and the audio recording. Participants used a computer mouse to select one of the

objects in the display. The display remained on the screen for 2000 ms after the mouse click. The positions of the target, the competitor and the distractors were counterbalanced across trials. At the beginning of each trial, the position of the mouse cursor was re-set to the centre of the screen. Following informed consent, participants were told that they would see a display with objects in it and at the same time they would hear instructions to click on one of those objects using the computer mouse. They should listen to the instructions carefully and select the object mentioned in the instruction. To familiarise participants with the procedure, the experiment started with 10 practice trials. Each session lasted approximately 40 mins and participants were offered regular breaks before calibration. The ethical approach of this project was approved by the ethics committees at the University of Strathclyde and the University of Stirling.

Results and Discussion

Mouse-clicking latency

We first analyzed mouse-clicking latencies (the timings of mouse-clicking on the target from the onset of the first adjective). **Fig.2** reports the means. In all experiments reported in this paper, reaction times with z-scores exceeding 3.29 were removed before analyses, as they may represent extreme values (Field, Miles, & Field, 2012); this resulted in the removal of less than 1% of data. For statistical analyses, we analyzed log-transformed reaction times in linear mixed effects modelling (e.g., Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (version 3.6.0., R Core Team, 2018). All analyses, including analyses on eye movement data, adopted the maximum random effect structure by including both by-participants and by-items random intercepts and slopes for all the fixed effects (Barr et al., 2013), whilst enforcing zero-correlations between them to avoid over-parameterization or false convergence (Bates, Kliegl, Vasishth, & Baayen, 2015; Bates et al., 2015;

Singmann, & Kellen, 2020). Fixed effects were mean-centred and standardized to reduce collinearity between variables and to facilitate convergence in R (Gelman & Hill, 2007).

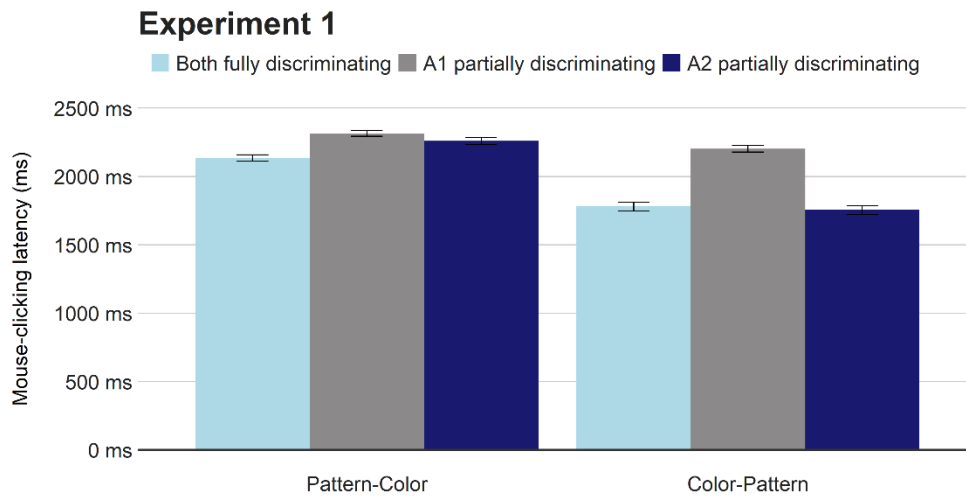


Fig.2. Mean mouse-clicking latencies for the target object (relative to the first adjective onset) in Experiment 1. Error bars represent standard errors.

We first examined if the effect of the first or second adjective discriminability was modulated by adjective order, by running two 2 (adjective order: pattern-color vs. color-pattern) \times 2 (discriminability: both fully discriminating vs. A1/A2 partially discriminating) analyses. Planned comparisons then examined the effect of adjective discriminability for each word order. **Table 2** summarizes the results.

First adjective discriminability. First, we examined the impact of the discriminability of the first adjective on mouse-clicking latencies and if this interacted with adjective order. We compared the *A1 partially discriminating* condition with the *both fully discriminating* condition. The analysis revealed a significant main effect of A1 discriminability, with faster mouse-clicking when the first adjective was fully discriminating (1957 ms) than when it was only partially discriminating (2260 ms). A main effect of adjective order showed shorter mouse-clicking latencies following color-pattern orders (1989 ms) than pattern-color orders (2227 ms). A significant A1 discriminability \times

adjective order interaction indicated that the discriminability of the first adjective led to faster mouse clicking when the first adjective was color (color-pattern orders) than when it was pattern (pattern-color orders), though in the planned comparisons the effect of A1 discriminability was significant in both orders.

Table 2
Analyses of mouse-clicking latencies in Experiment 1

		β	<i>SE</i>	<i>t</i>	<i>p</i>
First adjective (A1) discriminability					
<i>Main analysis</i>	Adjective order	0.07	0.01	9.57	<.001
	A1 discriminability	0.09	0.01	11.58	<.001
	A1 discriminability \times Adjective order	-0.04	0.01	-7.37	<.001
<i>Pattern-Color</i>	A1 discriminability	0.04	0.01	5.98	<.001
<i>Color-Pattern</i>	A1 discriminability	0.13	0.01	11.82	<.001
Second adjective (A2) discriminability					
<i>Main analysis</i>	Adjective order	0.13	0.01	12.96	<.001
	A2 discriminability	0.01	0.01	2.00	.052
	A2 discriminability \times Adjective order	0.02	0.01	2.90	.004
<i>Pattern-Color</i>	A2 discriminability	0.03	0.01	4.67	<.001
<i>Color-Pattern</i>	A2 discriminability	< -0.01	0.01	-0.46	.649

Second adjective discriminability. Second, we examined the impact of the second adjective (A2) discriminability by comparing the A2 *partially discriminating* condition with the *both fully discriminating* condition, and if this effect was modulated by adjective order. A marginally significant main effect of A2 discriminability indicated slightly faster mouse-clicking when the second adjective was fully discriminating (1957 ms) than when it was only partially discriminating (2010 ms). A significant main effect of adjective order confirmed faster mouse-clicking in color-pattern orders (1767 ms) than in pattern-color orders (2198 ms). Most critically, a significant A2 discriminability \times adjective order interaction and the planned comparisons indicated that the second adjective discriminability sped up mouse-clicking in pattern-color orders, but not in color-pattern orders.

Eye movements

Before the analyses, fixations shorter than 40 ms were merged with nearby fixations within $.5^\circ$ visual angle. Remaining fixations shorter than 80 ms were excluded from the analyses, assuming that information is not processed during fixations of such short duration (Rayner & Pollatsek, 1989). We then computed the proportion of time participants spent fixating the target and the competitor for every 100 ms time period following the onset of the first adjective in each trial. These were done using the Eyelink Data Viewer™ software from SR Research. **Fig.3** and **Fig.4** plot the mean fixation proportions for the pattern-color and color-pattern orders, respectively, illustrating the time course of eye movements from the onset of the first adjective.

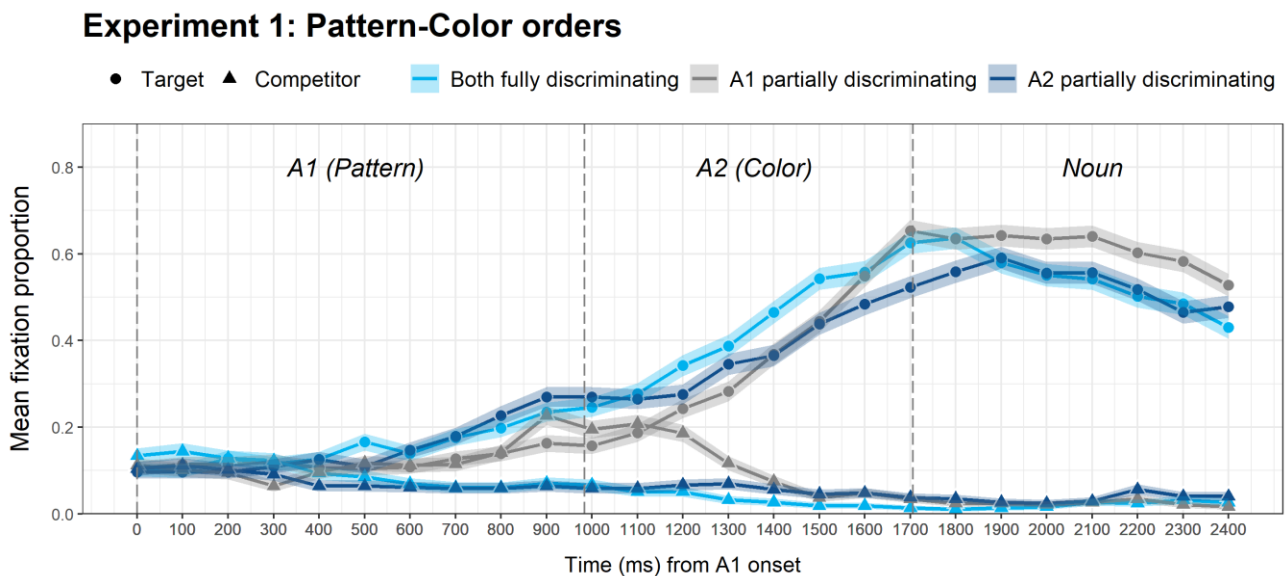


Fig.3. Mean fixation proportions following pattern-color orders in Experiment 1. Vertical dashed lines represent the average onsets of each region. A1 = First adjective, A2 = Second adjective.

Experiment 1: Color-Pattern orders

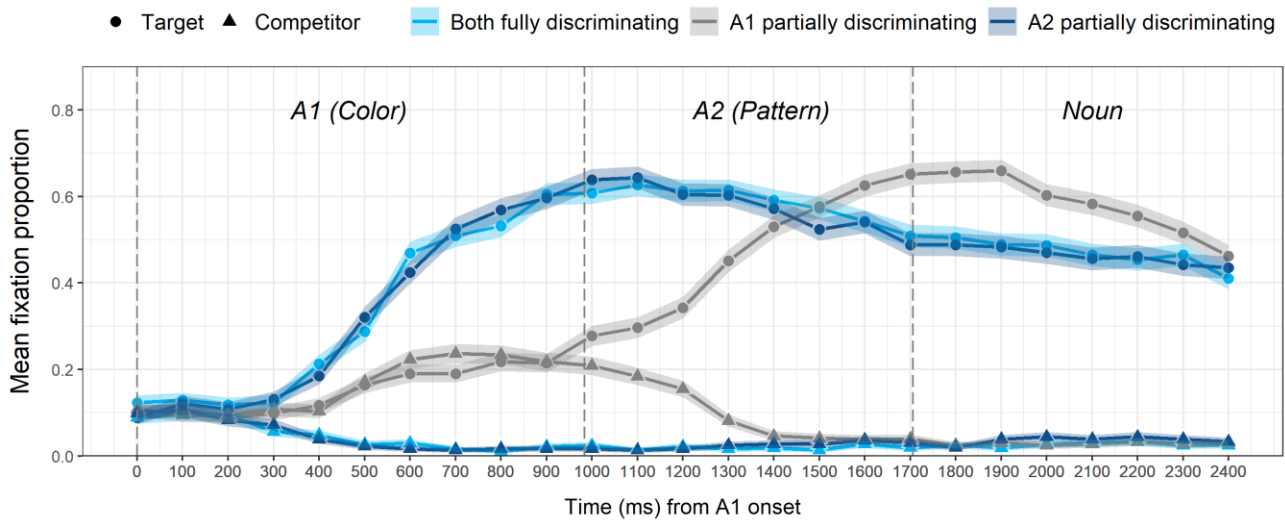


Fig.4. Mean fixation proportions following color-pattern orders in Experiment 1. Vertical dashed lines represent the average onsets of each region. A1 = First adjective, A2 = Second adjective

Since looks to the target and the competitor are not independent (more looks to one object imply fewer looks to the other, and vice-versa), for the statistical analyses on the eye movement data of this experiment and all subsequent experiments, we computed the log gaze probability ratios of looks to the target relative to the competitor $[\ln(P(\text{target})/P(\text{competitor}))]^3$ (Arai, Van Gompel, & Scheepers, 2007) in each 100-ms time period. In this measure, a score of zero indicates that both objects are fixated equally frequently, and a positive score reflects a preference for looking at the target over the competitor, and a negative ratio indicates the opposite. **Fig.5** reports the means of the log gaze probability ratios that were submitted to the mixed effects models. For these models we defined two time windows; the *A1 time window* (left panels in Fig.5), starting from 200 ms after the start of the first adjective (assuming that it takes approx. 200 ms to launch an eye movement, Martin, Shao, & Boff, 1993) and ending at the onset of the second adjective, and the *A2 time window* (right panels in Fig.5), starting from 200 ms after the start of the second adjective and ending at mouse-clicking. We adopted a cut-off point of 1200 ms after the second adjective onset in cases where the

³ We used .05 as a flattening constant before log transformations (to avoid division-by-zero errors). The use of different constants (e.g., .5 or .001) made no difference to the conclusions.

individual participants' mouse-clicking exceeded this point to delimit a period containing most data points.

As in the analyses on the mouse-clicking latencies, we ran two sets of mixed effects regression models. The first set assessed the impact of the discriminability of the first (A1) and second (A2) adjective in the A1 and A2 time windows respectively, investigating whether the effect was modulated by adjective order. The second set of models assessed the impact of A1 and A2 discriminability within each word order. In addition to adjective order and A1/A2 discriminability, the analyses included time (bins) as fixed effects to take into account changes in the rate of fixations over time. As before, fixed effects were all mean-centered and standardized, and random effects included by-participants and by-items random intercepts and slopes for all the fixed effects and interactions, whilst correlations were suppressed. **Table 3** summarizes the fixed effects. Because we are primarily interested in the effects of adjective order and adjective discriminability, we focus our discussion on these two variables.

Experiment 1

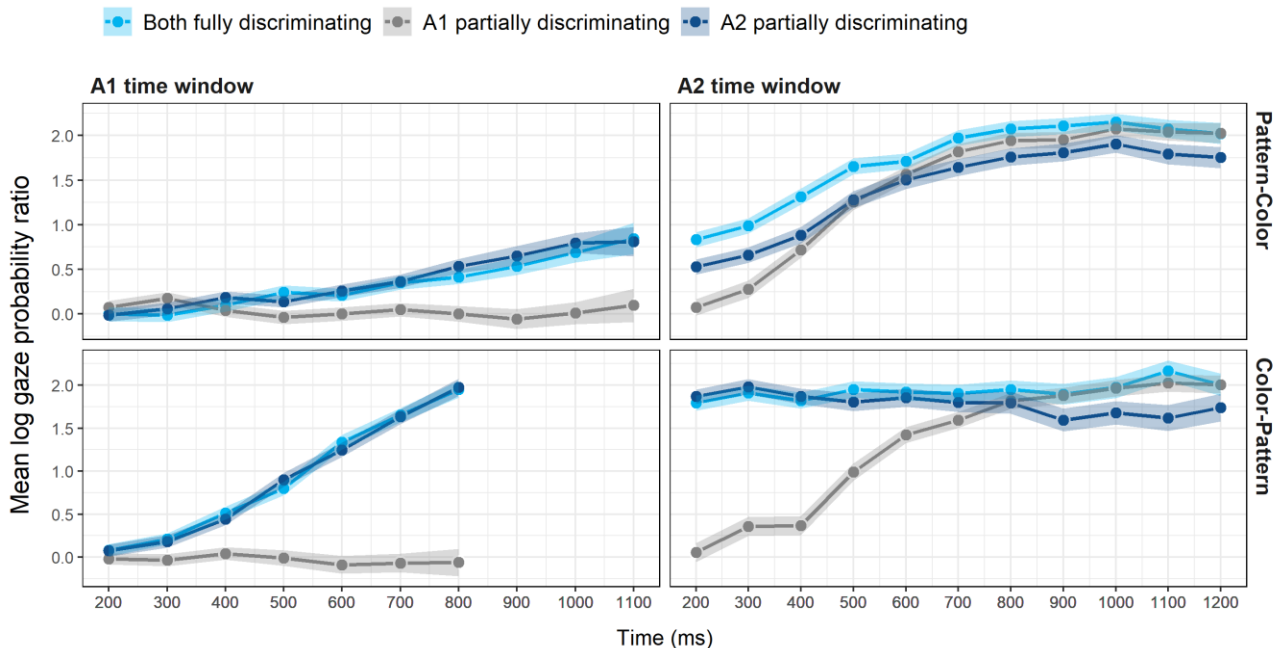


Fig.5. Mean log gaze probability ratios of fixations to the target relative to the competitor in Experiment 1. A1 = First adjective, A2 = Second adjective.

First adjective (A1) discriminability. The analyses revealed a main effect of A1 discriminability; there were more fixations to the target when the first adjective was fully discriminating than when it was only partially discriminating. A main effect of adjective order indicated more fixations to the target in color-pattern orders than in pattern-color orders. A significant A1 discriminability \times adjective order interaction confirmed a larger effect of A1 discriminability in color-pattern orders than in pattern-color orders, though the effect was significant in both orders. Additional analyses revealed that, when the first adjective was fully discriminating (both discriminating condition), there were more fixations to the target in color-pattern orders than in pattern-color orders ($\beta = -0.43$, $SE = 0.04$, $t = -9.61$, $p < .001$), indicating that fully discriminating color attributes boosted target looks more than fully discriminating pattern attributes. By contrast, when the first adjective was only partially discriminating, adjective order had no effect on fixations ($\beta = 0.03$, $SE = 0.05$, $t = 0.63$, $p = .532$), so it was not that mentioning color first was generally advantageous. In fact, when the first adjective was only partially discriminating, there was no preference to look at the target or the competitor either in color-pattern orders ($\beta = -0.02$, $SE = 0.08$, $t = -0.31$, $p = .757$) or in pattern-color orders ($\beta = 0.03$, $SE = 0.06$, $t = 0.43$, $p = .667$).

Second adjective (A2) discriminability. In the A2 time window, a main effect of adjective order indicated more looks to the target in color-pattern orders than in pattern-color orders. The main effect of A2 discriminability was also significant; a fully discriminating second adjective led to more looks to the target than an only partially discriminating second adjective. The effect of adjective order marginally interacted with A2 discriminability. In pattern-color orders, there were more looks to the target when the second attribute (color) was fully discriminating than otherwise. In color-pattern orders, there was no effect of A2 discriminability, reflecting the fact that in both conditions, the first color adjective was fully discriminating and fixation probabilities on the target had already reached a peak by the onset of the second adjective and continued to remain high.

Table 3

Analyses of eye movements in Experiment 1

		β	<i>SE</i>	<i>t</i>	<i>p</i>
First adjective (A1) discriminability (A1 time window)					
<i>Main analysis</i>	Adjective order	-0.20	0.03	-5.69	< .001
	A1 discriminability	-0.30	0.04	-8.56	< .001
	Time	0.22	0.03	6.46	< .001
	Adjective order \times A1 discriminability	0.23	0.03	7.32	< .001
	Adjective order \times Time	-0.16	0.03	-5.12	< .001
	A1 discriminability \times Time	-0.25	0.03	-8.50	< .001
	Adjective order \times A1 discriminability \times Time	0.15	0.03	4.45	< .001
<i>Pattern-Color</i>	A1 (pattern) discriminability	-0.12	0.04	-3.12	.003
	Time	0.09	0.04	2.07	.043
	A1 (pattern) discriminability \times Time	-0.13	0.04	-3.21	.002
<i>Color-Pattern</i>	A1 (color) discriminability	-0.46	0.05	-9.08	< .001
	Time	0.32	0.05	7.00	< .001
	A1 (color) discriminability \times Time	-0.33	0.05	-7.22	< .001
Second adjective (A2) discriminability (A2 time window)					
<i>Main analysis</i>	Adjective order	-0.21	0.04	-5.27	< .001
	A2 discriminability	-0.12	0.03	-3.62	.001
	Time	0.28	0.03	11.06	< .001
	Adjective order \times A2 discriminability	-0.06	0.03	-1.85	.069
	Adjective order \times Time	0.23	0.03	8.52	< .001
	Discriminability \times Time	-0.02	0.02	-1.04	.305
	Adjective order \times A2 discriminability \times Time	0.03	0.02	1.24	.219
<i>Pattern-Color</i>	A2 (color) discriminability	-0.17	0.04	-3.93	< .001
	Time	0.48	0.04	13.43	< .001
	A2 (color) discriminability \times Time	0.00	0.03	0.11	.917
<i>Color-Pattern</i>	A2 (pattern) discriminability	-0.05	0.05	-0.93	.356
	Time	0.02	0.04	0.62	.536
	A2 (pattern) discriminability \times Time	-0.06	0.04	-1.66	.104

Summary of Experiment 1

A fully discriminating first adjective led to more looks to the target (before the onset of the second adjective) and faster mouse-clicking than an only partially discriminating first adjective, for both pattern-color and color-pattern orders. These effects were stronger when the first adjective denoted color (in color-pattern orders) than when it denoted pattern (in pattern-color orders), indicating that color discriminates faster than pattern. Moreover, when color was mentioned first and

fully discriminating, neither target looks nor mouse-clicking latencies were affected by the discriminability of the second-mentioned pattern attribute. By contrast, when pattern was mentioned first and fully discriminating, the second-mentioned color attribute led to more target looks and faster mouse-clicking when it was fully discriminating than when it was only partially discriminating.

These results indicate that in the simultaneous presence of the visual display, referential processing proceeds incrementally, such that *how quickly* an earlier-mentioned attribute discriminates determines whether an over-specified later-mentioned attribute facilitates referential processes. Specifically, when the first attribute is fully discriminating and highly salient, the referent is identified equally fast regardless of how discriminating an over-specified second attribute is. By contrast, when the first attribute is discriminating but *not* salient, the referent cannot be identified very fast, so a fully discriminating over-specified second attribute facilitates referent search. The next question is whether the addition of a second attribute is advantageous or detrimental relative to a more concise one-attribute-only description and how this might depend on attributes and, critically, on the way they are mapped onto the context. Experiments 2 and 3 addressed these issues.

Experiment 2

Some researchers have argued that color is part of a referent's basic conceptual representation (Schriefers & Pechmann, 1988) or that it is a generally preferred attribute (Dale & Reiter, 1995). Indeed, Experiment 1 showed that color discriminates the referent faster than pattern. One possibility is therefore that the additional mention of color would be generally advantageous compared to a fully discriminating pattern-only description. However, as shown in Experiment 1, when referring expressions are heard in the simultaneous presence of the visual display, the search process starts as soon as the first adjective is encountered. Hence, although the additional mention of color after a fully discriminating pattern adjective may boost target fixations, as we saw in Experiment 1, it may not necessarily speed up mouse-clicking relative to a shorter, pattern-only

description. By contrast, when the listener has already heard both adjectives prior to the visual display, both attributes could be used in parallel from the outset of the search. Under such situations, the additional inclusion of color could be facilitatory, not only increasing target fixations but also speeding up mouse-clicking, since color discriminates much faster than pattern, as demonstrated in Experiment 1.

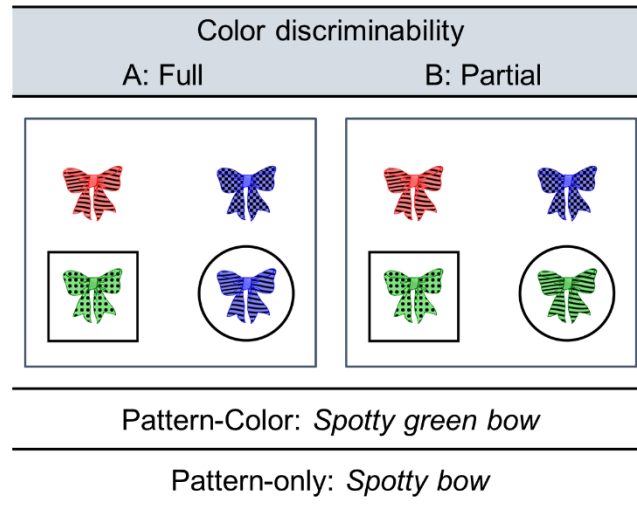


Fig.6. Example displays in Experiment 2. Box = Target, Circle = Competitor.

To test these predictions, Experiment 2 examined whether the over-specification of color in pattern-color descriptions (*spotty green bow*) could be advantageous relative to pattern-only descriptions (*spotty bow*). Thus, pattern was always mentioned first and fully discriminating. Color was either mentioned (pattern-color descriptions) or not mentioned (pattern-only descriptions), and in the display, color was either fully discriminating (**Fig.6A**) or only partially discriminating (**Fig.6B**). Critically, we presented the referring expressions either simultaneously with the visual display (*concurrent display*, as in Experiment 1, Experiment 2A) or 1000 ms before the presentation of the display (*auditory first display*) (Experiment 2B).

Methods

Participants

Forty-eight participants took part in Experiments 2A and 2B each. They were native speakers of British English, aged between 17 and 30, and recruited from the University of Stirling student community in exchange for course credits.

Materials

Each experiment involved 48 experimental items, with each item comprising two versions of spoken instructions and visual displays. Fig.6 shows sample displays, where the bow in the box is the target: Participants saw either display A or B, and heard either “*Click on the spotty green bow*” (pattern-color) or “*Click on the spotty bow*” (pattern only). In Fig.6A, color was fully discriminating (there is only one green bow), while in Fig.6B color was partially discriminating (there is another green bow, the competitor, marked in a circle). In the spoken instructions, pattern was always mentioned first and was fully discriminating in both *pattern-color* (two adjective) and *pattern-only* (one adjective) conditions. The onset of the noun in the one adjective condition and the onset of the second adjective in the two adjective condition were synchronised.

Apparatus and procedure

In Experiment 2A, the spoken instructions and visual displays were presented simultaneously, as in Experiment 1. In Experiment 2B, participants were shown a blank screen while the instructions were presented auditorily, and 1000 ms after the offset of the spoken instructions, the visual display appeared. We used a 24-inch widescreen LCD monitor in both experiments.

Design

Each experiment used a 2 (Color inclusion: pattern-color vs. pattern-only) × 2 (Color discriminability: full vs. partial) repeated measured design, resulting in four lists, including 48 experimental items, 12 items per condition, and 84 filler items.

Results and Discussion

Experiment 2A: Concurrent presentation of visual display and auditory instructions

Mouse-clicking latency. Fig.7 reports the mean mouse-clicking latencies. The analyses were carried out as before. The initial main analysis included color inclusion (pattern-color vs. pattern-only) and color discriminability in the display (full vs. partial) as fixed effects, and planned comparisons examined the impact of color inclusion for each display separately. **Table 4** summarizes the results. Color inclusion delayed mouse-clicking times (2211 ms) relative to pattern-only descriptions (2019 ms) and the delay was significant both when color was fully discriminating and when it was partially discriminating. No effect of color discriminability or a color inclusion \times discriminability interaction was found.

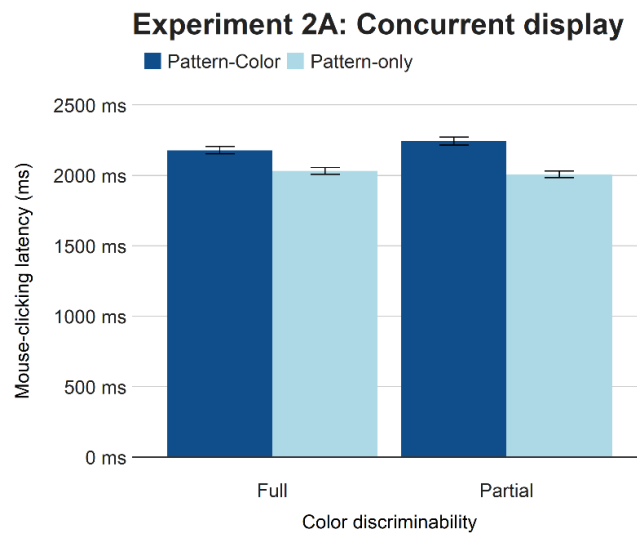


Fig.7. Mean mouse-clicking latencies (ms) for the target objects from the onset of the referring expression in Experiment 2A. Error bars represent standard errors.

Table 4

Analyses of mouse-clicking latencies in Experiment 2A

	β	SE	t	p
<i>Main analysis</i>				
Color inclusion	0.04	0.01	8.10	<.001
Color discriminability	< 0.01	< 0.01	0.56	.581
Color inclusion \times discriminability	0.01	0.01	1.61	.116
<i>Color fully discriminating</i>				
Color inclusion	0.03	0.01	4.45	<.001
<i>Color partially discriminating</i>				
Color inclusion	0.05	0.01	6.61	<.001

Eye movements. The mean fixation proportions for the target and competitor from the onset of the first adjective are shown in **Fig.8**. The graph shows increased looks to the target and slightly fewer looks to the competitor in response to the second-mentioned color adjective but only when color was fully discriminating (top panel). As in Experiment 1, for statistical analyses, we computed the log gaze probability ratios. Since the first adjective was identical across conditions, the analyses concentrated on fixations starting 200 ms after the onset of the second adjective (in pattern-color descriptions) or noun (in pattern-only descriptions) and ending with mouse-clicking, with a cut-off point of 1200 ms after the onset of the second adjective or noun (the average mouse-clicking latency was 1227 ms). **Fig.9** reports the means. The initial analyses included color inclusion (pattern-color vs. pattern-only), color discriminability (full vs. partial) and time as fixed effects, and planned comparisons examined simple effects of color inclusion for each level of color discriminability. **Table 5** summarizes the results. There was no significant main effect of color inclusion. However, the main effect of color discriminability was significant, with more looks to the target when color was fully discriminating in the display than otherwise. The effect marginally interacted with color inclusion. The planned comparisons confirmed that the mention of color increased fixations to the target (relative to pattern-only descriptions) when color was fully discriminating, whereas it had no effect when color was partially discriminating.

Experiment 2A: Concurrent display

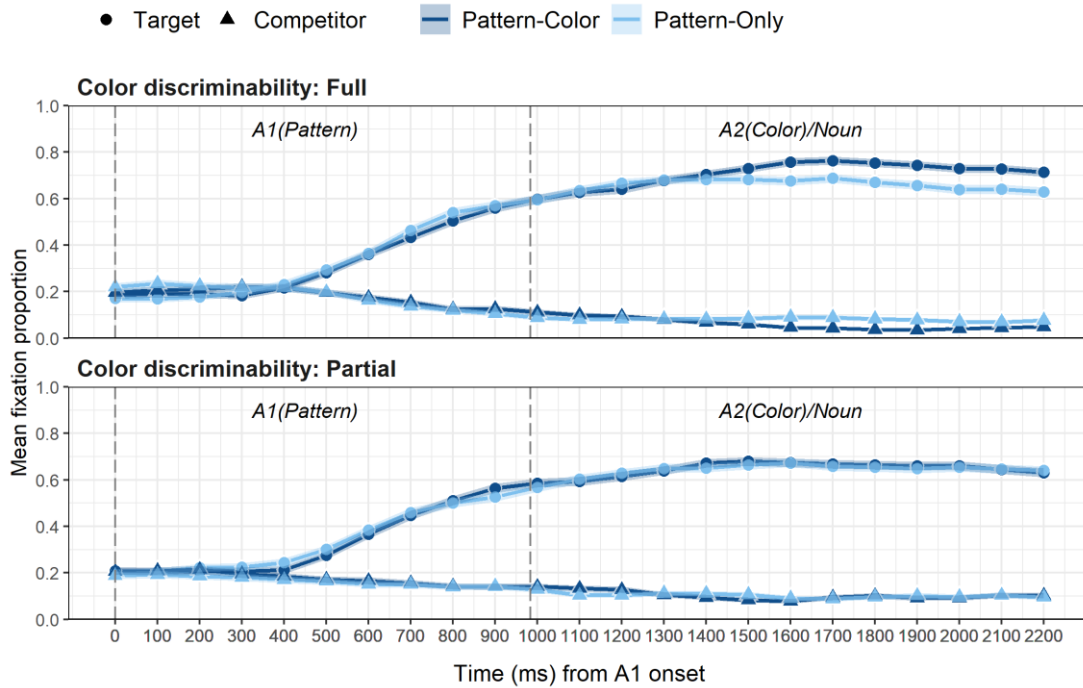


Fig.8. Mean fixation proportions for the target and competitor in Experiment 2A. Vertical dashed lines indicate the average onset of each region. A1 = First adjective, A2 = Second adjective.

Experiment 2A: Concurrent display

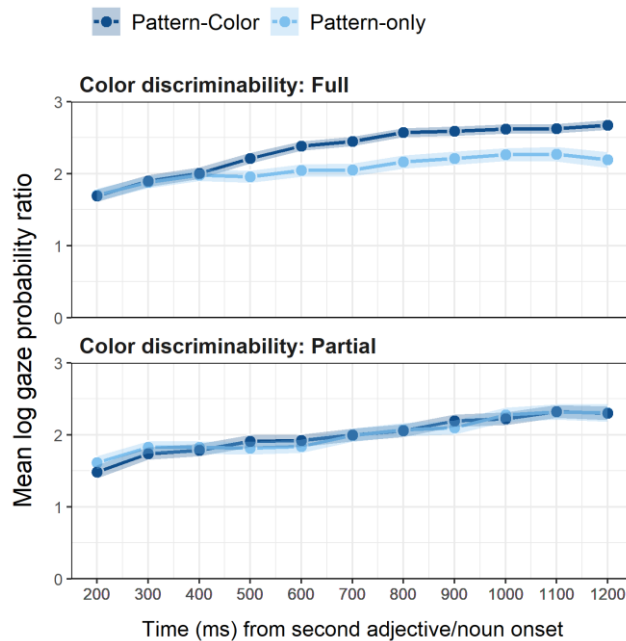


Fig.9. Mean log gaze probability ratios of fixations to the target relative to the competitor after the onset of the second adjective or the noun in Experiment 2A.

Table 5

Analyses of eye movements from second adjective/noun onset in Experiment 2A

	β	<i>SE</i>	<i>t</i>	<i>p</i>
<i>Main analysis</i>				
Color inclusion	0.05	0.03	1.47	.146
Color discriminability	-0.11	0.03	-3.66	.001
Time	0.23	0.04	5.94	< .001
Color inclusion \times discriminability	-0.06	0.03	-1.81	.075
Color inclusion \times Time	0.02	0.03	0.93	.360
Color discriminability \times Time	-0.02	0.03	-0.64	.523
Color inclusion \times discriminability \times Time	-0.03	0.03	-1.10	.279
<i>Color fully discriminating</i>				
Color inclusion	0.11	0.05	2.39	.019
Time	0.26	0.04	5.88	< .001
Color inclusion \times Time	0.06	0.03	1.64	.106
<i>Color partially discriminating</i>				
Color inclusion	-0.01	0.05	-0.24	.813
Time	0.21	0.05	3.93	< .001
Color inclusion \times Time	< -0.01	0.04	-0.09	.927

Summary of Experiment 2A

When color was fully discriminating, the additional mention of color as a second adjective boosted looks to the target relative to pattern-only descriptions, whereas it did not result in fewer looks to the target when color was only partially discriminating (so it was compatible with a competitor). Hence, under incremental processing, the second adjective can *reinforce, but not alter*, the referential processes triggered by a fully discriminating first adjective, which restricts the set of possible referents for the second adjective. Importantly, although this increase in target looks indicates an advantage of over-specification, the mention of color after pattern delayed mouse-clicking regardless of color discriminability, which clearly points to over-specification as a cost. Under incremental processing, word order determines which attribute is used first for referent search. When pattern was mentioned first and fully discriminating, listeners were able to launch a

search based on pattern; hence, even though the later-mentioned color attribute reinforced their interpretation, it did not offset the additional processing time for that color attribute. The question is whether the second-mentioned color adjective could speed up referent identification under parallel processing, where color can be used as early as pattern. Experiment 2B investigated this.

Experiment 2B: Auditory first display

Mouse-clicking latency. Fig.10 reports the means. The analyses were carried out as before.

Table 6 summarizes the results. The main effect of color inclusion showed that participants were faster mouse-clicking when color was mentioned (1196 ms) than when it was not (1502 ms). A significant main effect of color discriminability also revealed faster mouse-clicking when color discriminated fully (1320 ms) than when it only partially discriminated (1379 ms). The significant color inclusion \times discriminability interaction indicated that color inclusion sped up mouse-clicking more (235 ms) when color was fully discriminating compared to when it was only partially discriminating, though the effect of color inclusion was significant in both conditions.

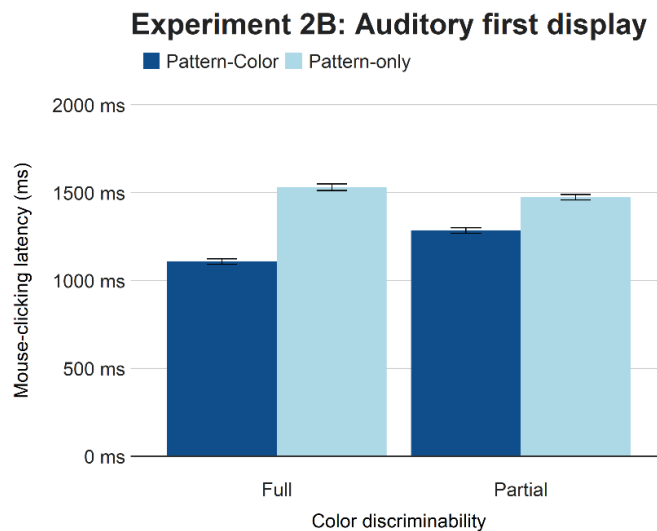


Fig.10. Mean mouse-clicking times (ms) from display onset in Experiment 2B. Error bars represent standard errors.

Table 6
Analyses of mouse-clicking latencies in Experiment 2B

	β	<i>SE</i>	<i>t</i>	<i>p</i>
<i>Main analysis</i>				
Color inclusion	-0.12	0.01	-19.12	<.001
Color discriminability	0.03	0.01	5.79	<.001
Color inclusion \times discriminability	0.05	0.01	9.46	<.001
<i>Color fully discriminating</i>				
Color inclusion	-0.16	0.01	-17.61	<.001
<i>Color partially discriminating</i>				
Color inclusion	-0.07	0.01	-11.13	<.001

Eye movements. **Fig.11** reports the mean fixation proportions for the target and competitor, showing a steep rise in target looks in pattern-color descriptions, particularly when color was fully discriminating (top panel); interestingly, though, even when color was only partially discriminating (bottom panel), as early as 200 ms after the display representations, looks to both the target and to the competitor were higher in pattern-color descriptions than in pattern-only descriptions, indicating that color mention immediately eliminated some referential alternatives. As before, for statistical analyses, we analyzed the log gaze probability ratios for the period from 200 ms until mouse-clicking, with the cut-off point set at 1300 ms after the display onset, since the average mouse-clicking latency was 1350 ms after the display onset. **Fig.12** reports the means, and **Table 7** summarizes the results. A main effect of color inclusion indicated more looks to the target when color was mentioned than when it was not. The main effect of color discriminability was also significant, with more target looks when color was fully discriminating than when it was only partially discriminating. The significant color inclusion \times discriminability interaction indicates that color inclusion increased looks to the target more when color was fully discriminating than when it was only partially discriminating. The simple effect of color inclusion was significant in both cases.

Experiment 2B: Auditory first display

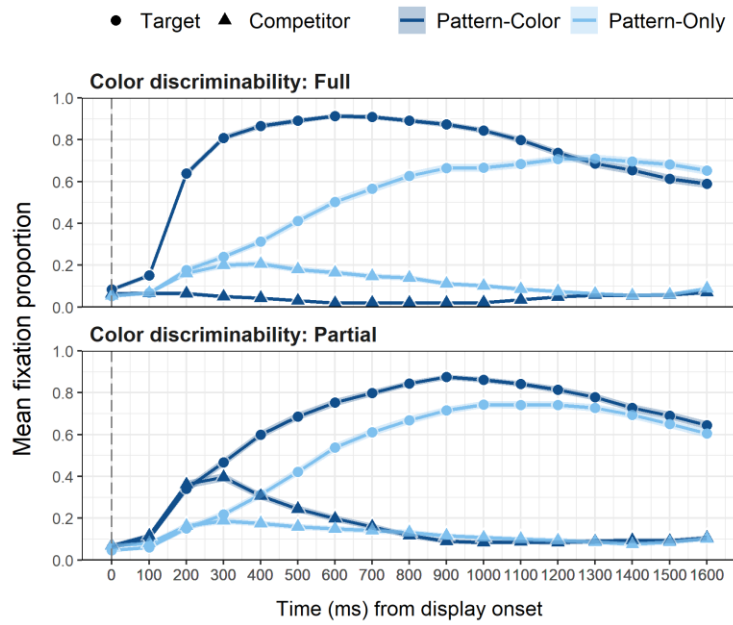


Fig.11. Mean fixation proportions for the target and competitor from display onset in Experiment 2B.

Experiment 2B: Auditory first display

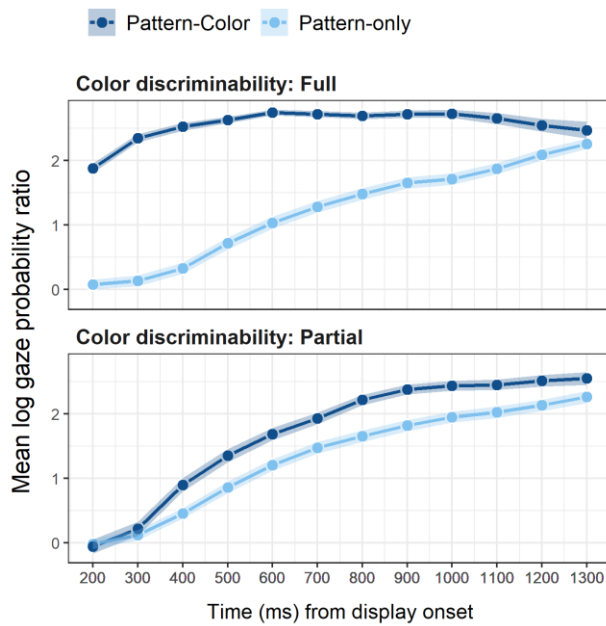


Fig.12 Mean log gaze probability ratios of fixations to the target relative to the competitor after display presentation in Experiment 2B.

Table 7

Analyses of eye movements from display onset in Experiment 2B

	β	<i>SE</i>	<i>t</i>	<i>p</i>
<i>Main analysis</i>				
Color inclusion	0.47	0.03	13.84	< .001
Color discriminability	-0.20	0.03	-7.83	< .001
Time	0.69	0.04	17.02	< .001
Color inclusion \times discriminability	-0.27	0.03	-9.60	< .001
Color inclusion \times Time	-0.07	0.03	-2.63	.011
Color discriminability \times Time	0.16	0.03	6.01	< .001
Color inclusion \times discriminability \times Time	0.17	0.02	7.68	< .001
<i>Color fully discriminating</i>				
Color inclusion	0.75	0.05	15.90	< .001
Time	0.52	0.04	12.22	< .001
Color inclusion \times Time	-0.25	0.03	-7.56	< .001
<i>Color partially discriminating</i>				
Color inclusion	0.22	0.04	5.48	< .001
Time	0.85	0.06	15.41	< .001
Color inclusion \times Time	0.09	0.04	2.31	.025

Summary of Experiment 2B

The additional inclusion of color as the second adjective immediately enhanced looks to the target and sped up mouse-clicking relative to a more concise, pattern-only description. Interestingly, this happened even when color was only partially discriminating. Although color was mentioned after pattern, because the referring expressions had been heard in advance of the visual display presentation in this experiment, both attributes were available from the outset of the search. Because color discriminates more quickly than pattern, as shown in Experiment 1, the mention of color immediately narrowed down the set of referential candidates for pattern, enhancing target looks and speeding up mouse-clicking by reducing the time for pattern to discriminate. These findings contrast with the results of Experiment 2A, where the second-mentioned color attribute delayed mouse-clicking, even though it enhanced looks to the target. Hence, Experiment 2 demonstrated that over-

specification affects referent search differently depending on whether the additional attribute can be used early or late in referent search.

Experiment 3

In Experiment 1, when the first-mentioned color attribute was fully discriminating, the discriminability of the second-mentioned pattern adjective had no effect on either mouse-clicking or eye movements. Hence, one possibility is that listeners ignore pattern altogether when it is over-specified after color, such that the additional mention of pattern is not detrimental relative to a color-only description. Alternatively, listeners may attend to all the attributes mentioned in the referring expression, so the additional mention of pattern would generally hinder referential processing relative to a color-only description. That is, pattern mention may delay referential processing, not only when it can only be used later than color (in incremental processing), but also when it can be used as early as color (in parallel processing).

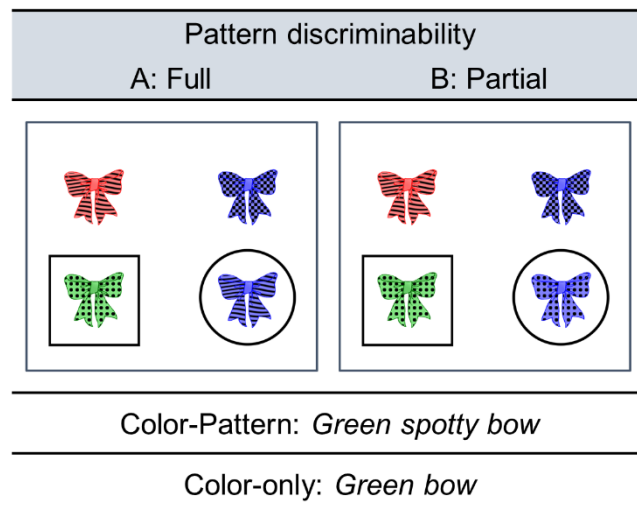


Fig.13. Example stimuli in Experiment 3. Box = Target, Circle = Competitor.

To test these hypotheses, Experiment 3 contrasted color-only descriptions (*green bow*) with color-pattern descriptions (*green spotty bow*) using the same design as Experiment 2. The first adjective was color and was always fully discriminating. By contrast, pattern was either mentioned or not mentioned and in the display, it was either fully (**Fig.13A**) or only partially discriminating

(**Fig.13B**). As in Experiment 2, the auditory referring expressions were presented concurrently with the visual display (Experiment 3A) or prior to the visual display presentation (Experiment 3B).

Method

Participants

Experiments 3A and 3B each recruited forty-eight participants from the same population as Experiment 2.

Materials

For the two-adjective conditions, we used the color-pattern descriptions from Experiment 1, while in the one-adjective conditions, we omitted the color adjective. In the display, pattern was either fully discriminating (Fig.13A) or only partially discriminating (Fig.13B).

Apparatus and procedure

These were the same as in Experiment 2. In Experiment 3A, the visual display was concurrently available when the auditory instructions were presented. In Experiment 3B, the visual displays were shown 1000 ms after the offset of the instructions.

Design

Each experiment used a 2 (Pattern inclusion: color-pattern vs. color-only) \times 2 (Pattern discriminability: full vs. partial) repeated measures design.

Results and Discussion

Experiment 3A: Concurrent presentation of visual display and auditory instructions

Mouse clicking latency. **Fig.14** reports the means by condition. The initial analyses contained pattern inclusion (color-pattern vs. color-only) and pattern discriminability (full vs. partial) as fixed effects, and planned comparisons examined the impact of pattern inclusion separately for each display. **Table 8** summarizes the results. The main effect of pattern inclusion indicated delayed mouse-clicking when pattern was mentioned (2020 ms) than when it was not (1619 ms), and the effect was significant both when pattern was fully discriminating and when it

was only partially discriminating. There was no significant main effect of pattern discriminability or interaction. These results are in line with the corresponding results from Experiment 2A.

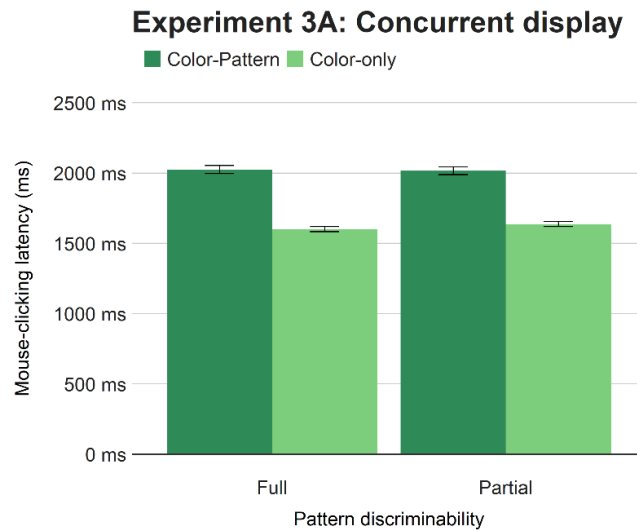


Fig.14 Mean mouse-clicking times (ms) from the onset of the referring expression in Experiment 3A. Error bars represent standard errors.

Table 8

Analyses of mouse-clicking latencies for Experiment 3A

	β	<i>SE</i>	<i>t</i>	<i>p</i>
<i>Main analysis</i>				
Pattern inclusion	0.10	0.01	9.80	< .001
Pattern discriminability	0.01	< 0.01	1.26	.214
Pattern inclusion \times discriminability	-0.01	< 0.01	-1.14	.253
<i>Pattern fully discriminating</i>				
Pattern inclusion	0.10	0.01	9.65	< .001
<i>Pattern partially discriminating</i>				
Pattern inclusion	0.09	0.01	8.57	< .001

Eye movements. **Fig.15** reports the proportion of fixations on the target and the competitor objects from the onset of the first adjective, illustrating a sharp rise in target looks in response to the first color adjective in all conditions. As before, statistical analyses were carried out on the log gaze probability ratios, the means of which are reported in **Fig.16**. The analyses were performed on

fixations starting 200 ms after the onset of the second adjective/noun and ending at mouse-clicking, with 1200 ms after the second adjective/noun onset as the cut-off point (the average mouse-clicking latencies were 1254 ms after the second adjective/noun onset). **Table 9** summarizes the statistical analyses. The analyses revealed no effect of pattern discriminability nor a pattern inclusion \times pattern discriminability interaction. However, a main effect of pattern inclusion indicated slightly but significantly fewer fixations on the target in color-pattern descriptions than in color-only descriptions.

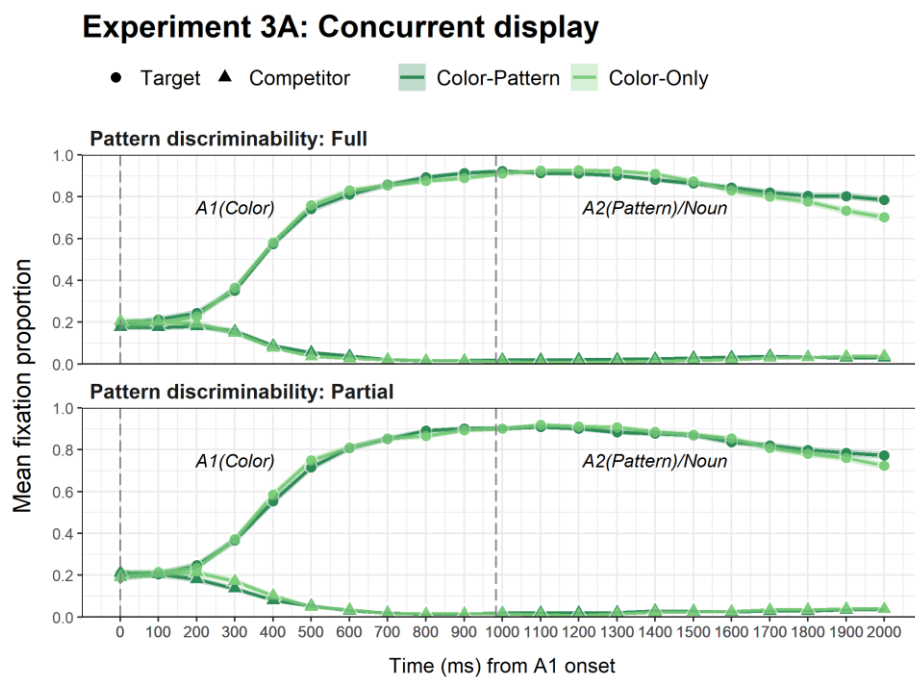


Fig.15. Mean fixation proportions for the target and competitor in Experiment 3A. Vertical dashed lines indicate the average onset of each region. A1 = First adjective, A2 = Second adjective.

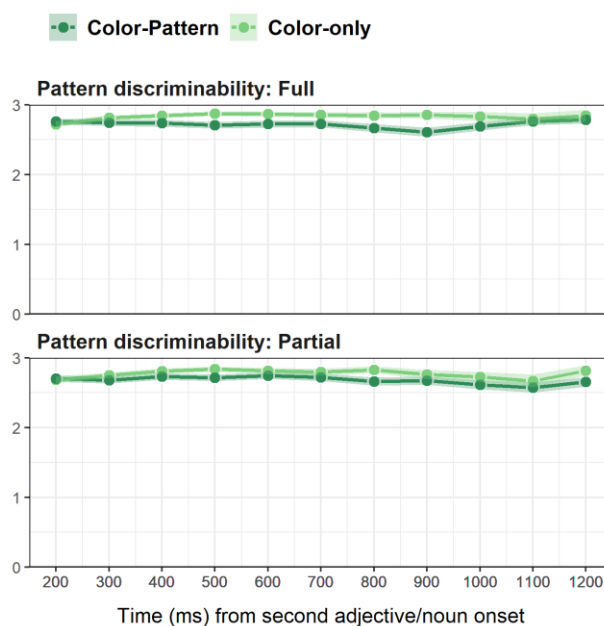
Experiment 3A: Concurrent display

Fig.16. Mean log gaze probability ratios of fixations to the target relative to the competitor (after the onset of the second adjective or the noun) in Experiment 3A.

Table 9

Analyses of eye movements in Experiment 3A

	β	<i>SE</i>	<i>t</i>	<i>p</i>
Pattern inclusion	-0.05	0.02	-2.59	.012
Pattern discriminability	-0.01	0.02	-0.65	.515
Time	< 0.01	0.02	0.19	.848
Pattern inclusion \times Pattern discriminability	< 0.01	0.02	0.20	.844
Pattern inclusion \times Time	-0.03	0.01	-1.89	.064
Pattern discriminability \times Time	< -0.01	0.02	-0.16	.872
Pattern inclusion \times Pattern discriminability \times Time	< -0.01	0.01	-0.21	.836

Summary of Experiment 3A

The mention of pattern as the second adjective delayed mouse-clicking and reduced looks to the target, regardless of its discriminability, providing evidence that the additional mention of less salient attributes hinders referential processing. The results contrast with those of Experiment 2A, where the mention of a fully discriminating color attribute boosted fixations to the target. One may wonder if the findings go against other eye-tracking studies showing that listeners identify the

referent as soon as they have sufficient information to do so (e.g., Eberhard et al., 1995; Rubio-Fernandez et al., 2020; Rubio-Fernandez & Jara-Ettinger, 2020). Note, though, that these studies relied on evidence from the eye movement data only. The initial eye movement pattern in our data is entirely consistent with the data from those studies in that we also observed a steep increase in looks to the target in response to the first-mentioned discriminating color adjective. Critically, referential processing does not end there: The subsequent mention of a pattern modifier generally delayed mouse-clicking and decreased target fixations (albeit only slightly). The effects did not interact with pattern discriminability, giving no indication that the mention of the pattern adjective altered the interpretations based on the earlier mentioned color adjective. Nevertheless, the reduced fixations and delayed mouse-clicking clearly indicate that referential processing was hampered by pattern mention and that listeners do attend to the second-mentioned, less salient information, rather than ignoring it. This is why over-specification can hamper referential processing. The question is whether this is specific to incremental processing, that is, it only arises because listeners “postpone” mouse-clicking until they have heard the second attribute. If so, the additional mention of a less salient attribute should not interfere with referent search if listeners have heard all the attributes in advance. We examined this possibility in Experiment 3B.

Experiment 3B: Auditory first display

Mouse-clicking latencies. Fig.17 reports the mean mouse-clicking latencies by condition. The analyses are summarized in **Table 10**. A significant main effect of pattern inclusion indicated slower mouse-clicking latencies when pattern was mentioned (1079 ms) than when it was not (1041 ms), though the difference (38 ms) was not as large as the one observed in Experiment 3A (401 ms). The inclusion of pattern delayed mouse-clicking both when pattern was fully discriminating and when it was only partially discriminating, and there was no significant main effect of pattern discriminability nor a significant interaction.

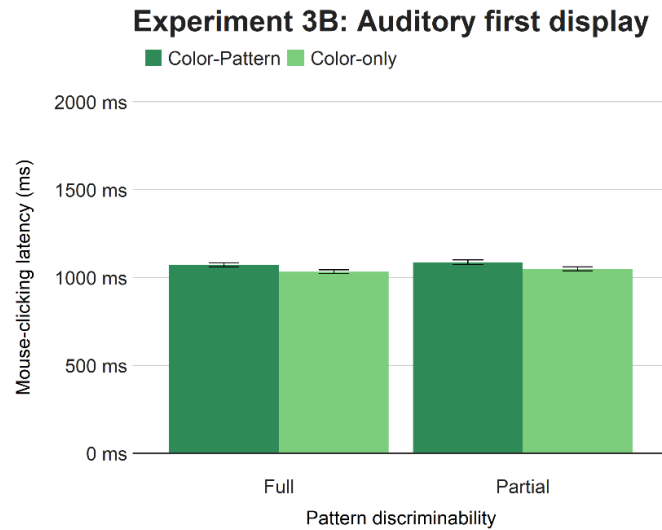


Fig.17. Mean mouse-clicking latencies (ms) from display onset in Experiment 3B.

Error bars represent standard errors.

Table 10

Analyses of mouse-clicking latencies for Experiment 3B

	β	<i>SE</i>	<i>t</i>	<i>p</i>
<i>Main analysis</i>				
Pattern inclusion	0.02	< 0.01	4.46	< .001
Pattern discriminability	0.01	< 0.01	1.61	.108
Pattern inclusion \times discriminability	< -0.01	< 0.01	-0.44	.661
<i>Pattern fully discriminating</i>				
Pattern inclusion	0.02	0.01	3.49	< .001
<i>Pattern partially discriminating</i>				
Pattern inclusion	0.02	0.01	2.44	.021

Eye movements. **Fig.18** reports the mean proportions of looks to the target and competitor objects from the display presentation. Again, the graph shows a steep rise in target fixations in all conditions, as in Experiment 3A and in some conditions in Experiment 2B. Statistical analyses were conducted on the log gaze probability ratios in a period starting 200 ms after display onset and ending at mouse-clicking, with 1000 ms after display onset adopted as the cut-off point, since the overall average mouse-clicking latency was 1060 ms. **Fig.19** presents the means. **Table 11** summarizes the results. A marginally significant main effect of pattern inclusion indicated slightly

fewer fixations on the target in color-pattern descriptions than in color-only descriptions. This did not reliably interact with pattern discriminability and there was no significant main effect of pattern discriminability.

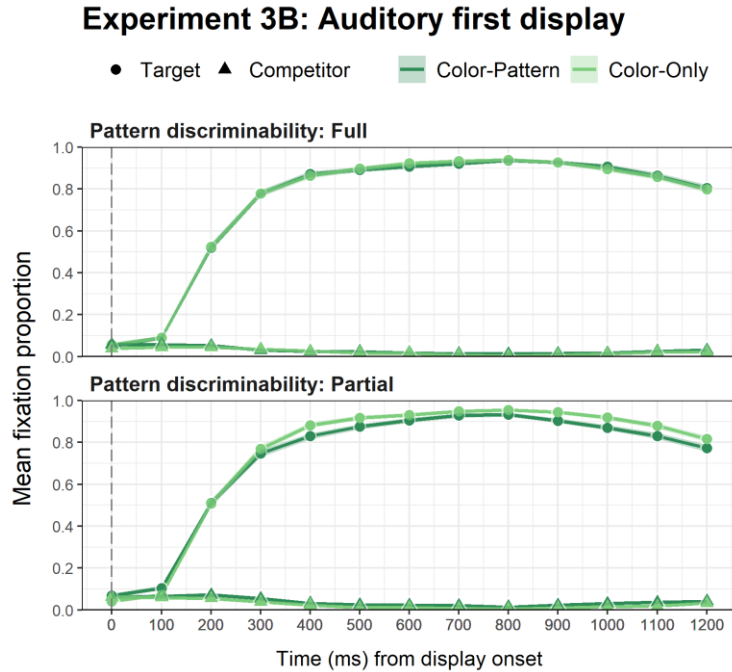


Fig.18. Mean fixation proportions for the target and competitor from display onset in Experiment 3B.

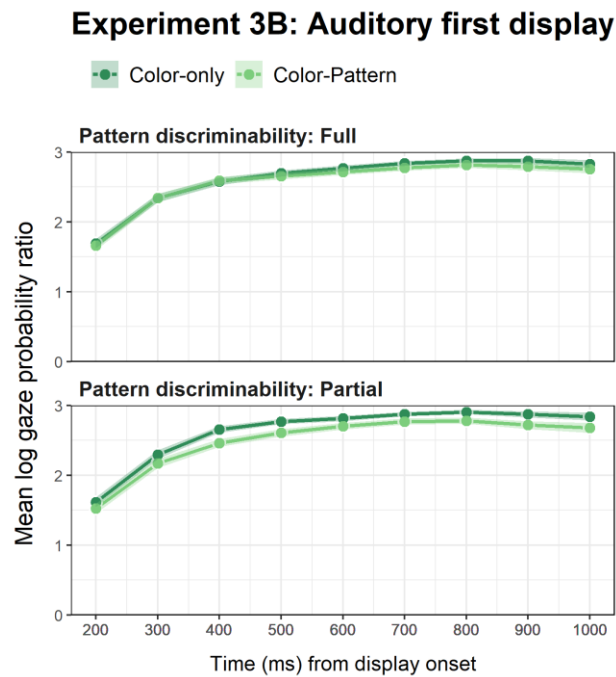


Fig.19. Mean log gaze probability ratios of fixations to the target relative to the competitor (after display presentation) in Experiment 3B.

Table 11
Analyses of eye movements for Experiment 3B

	β	<i>SE</i>	<i>t</i>	<i>p</i>
Pattern inclusion	-0.04	0.02	-1.75	.085
Pattern discriminability	-0.01	0.02	-0.70	.484
Time	0.33	0.03	11.17	< .001
Pattern inclusion \times discriminability	-0.02	0.02	-1.04	.303
Pattern inclusion \times Time	< -0.01	0.01	-0.32	.749
Pattern discriminability \times Time	0.02	0.01	1.14	.261
Pattern inclusion \times discriminability \times Time	0.01	0.02	0.41	.680

Summary of Experiment 3B

As in Experiment 3A, the additional mention of a pattern adjective after a fully discriminating color adjective delayed mouse-clicking and marginally reduced target fixations. Thus, in both incremental and parallel processing, the over-specification of pattern was detrimental relative to a color-only description. These results thus contrast with the findings of Experiment 2B, where, under parallel processing, the additional mention of color resulted in faster mouse clicking and enhanced target looks relative to a pattern-only description. This indicates that over-specified attributes cannot facilitate referent selection, even when they are available from the outset of the search, *unless* they are more salient and hence discriminate faster than attributes mentioned in the corresponding more concise description. In fact, our findings suggest something stronger: The over-specification of less salient or slow discriminating attributes such as pattern *delays* referent selection relative to a more concise description with highly salient attributes such as color. Less salient attributes discriminate more slowly, but listeners tend not to ignore those attributes when they are mentioned in the referring expressions. Hence even when the referring expression contains a salient or “fast” attribute, the inclusion of a less salient or “slow” attribute increases the overall processing time.

General discussion

Experiment 1 showed that when the visual display was the simultaneously present, the salience and discriminability of the attributes jointly facilitate incremental referential processes. Color discriminates much faster than pattern, such that when color is mentioned first and is fully discriminating, the referent is identified equally fast regardless of the discriminability of an over-specified pattern attribute. By contrast, when pattern is mentioned first, even when it fully discriminates, the discriminability of an over-specified color adjective facilitates referential processing. Experiments 2 and 3 then examined whether the additional mention of a second adjective hinders or facilitates referential processing relative to a shorter one-adjective only description. Under incremental processing, the additional mention of color as a second adjective delayed referent selection relative to a pattern-only description, even when color mention enhanced looks to the referent. By contrast, when the referring expressions were heard in advance so that the multiple attributes could be processed in parallel, additional color mention led to faster referent selection relative to a pattern-only description. The additional mention of pattern was detrimental in both incremental and parallel processing. Overall, our findings are consistent with the hypotheses, derived from findings in production (Fukumura, 2018): Whether over-specification facilitates comprehension depends not only on the salience or discriminability of the attributes, but also on whether listeners can use such information early.

These findings have important implications for theoretical accounts inspired by the Gricean (1975) Maxim of Quantity and its related notion of informativity (e.g., Dale, 1989; 1992; Reiter, 1990). According to these accounts and their interpretation of Grice's maxims, for optimal referential processing, referential expressions should contain the fewest possible attributes, and modifiers should be mentioned only when they are strictly necessary for discrimination. In some conditions our results appear to be in line with this length-driven optimization of the Gricean Maxim of Quantity in that descriptions with one adjective led to faster referent selection than those with two

adjectives when the referring expression was heard in the presence of the visual display. Crucially, this “brevity advantage” arose from the constraints of the incremental referential process, rather than from the listeners’ general preference for fewest possible attributes. Under incremental processing, the search for the referent begins as soon as the listener encounters the first-mentioned attribute, so if the first attribute fully discriminates, referent selection tends to occur faster when the listener does not have to process an additional attribute than otherwise. If speakers generally preferred the shortest possible referring expressions, then the inclusion of a second adjective should have always hindered comprehension. But this is not what we found: When the referring expressions were heard prior to the search, the additional mention of a second adjective did not hinder but sped up referent selection when it denoted a salient attribute such as color.

Consistent with this, we found no evidence that listeners *expect* speakers to use fewest possible attributes. Recall in Experiment 1 that, when the first adjective was partially discriminating, that is, when it was compatible with either the target or the competitor, there was no preference for either target or competitor, while when the first adjective was fully discriminating, there was a clear preference for the target. If listeners expected the first-mentioned attribute to be used optimally informatively (e.g., Sedivy et al., 1999), then when the first adjective was partially discriminating they would have preferred looking at the competitor rather than the target (the spotted blue bow rather than the spotted green bow in Fig.1B top display upon hearing *spotty* in *spotty green bow*), because the first-mentioned attribute was over-specified for the target (*green bow* would be sufficient in Fig.1B top display), whereas it was not for the competitor (*spotty* was necessary to discriminate the competitor from an alternative in Fig.1B top display). Instead, we found no preferential looks to either competitor or target before the disambiguation by the second adjective that was fully discriminating.

Our findings corroborate, for comprehension, results from production research pointing to the particularly strong discriminating property of color attributes. Although speakers’ tendency to

over-specify color (e.g., Engelhardt & F.Ferreira, 2014; Fukumura, 2018; Gatt et al., 2016; Schriefers & Pechmann, 1988; Tarenskeen et al., 2015; Van Gompel et al., 2019) may in part reflect their own tendency to mention the most easily recognisable (Pechmann, 1989) or preferred information (Dale & Reiter, 1995), our results show that in comprehension, color discriminates the referents much more quickly than pattern (Experiment 1). The over-specification of color (Experiment 2) also acted as a facilitating cue in referent search, as opposed to pattern, whereas the over-specification of pattern in the same conditions had the opposite effect (i.e., it caused a delay) (Experiment 3). Hence, it may be that speakers over-specify color more frequently and they tend to place it in an earlier position than pattern particularly when it discriminates fully (Fukumura, 2018), because salience and discriminability jointly determine how quickly an attribute discriminates (Experiment 1) and such a *fast discriminating* attribute tends to become available early during message encoding. That is, speakers preferentially choose an attribute that discriminates faster and produce it in an earlier position. This not only promotes incremental production processes but also facilitates incremental referential processing in comprehension.

Note, though, that these findings cannot be explained by the view that color over-specification facilitates comprehension because listeners take account of speakers' production preferences and thus they *expect* color to be over-specified in the speakers' utterances (e.g., Degen et al., 2020). This is because, like the brevity advantage, we show that the color advantage is affected by comprehension-internal processing constraints: In our study, the additional mention of color as a second adjective sped up referent selection only under parallel processing, where the second-mentioned color could be used as early as the first-mentioned pattern. Under incremental processing, the mention of color as a second adjective delayed referent selection relative to a more concise pattern-only description, even though it boosted fixations to the target. Hence, for the additional mention of color to speed up referent selection (relative to a more concise description),

the color attribute needs to be used early, either being mentioned earlier in the description under incremental processing or being used at least as early as other attributes under parallel processing.

Our finding that over-specification affects referential processing differently depending on whether the visual display is available to the listener at the time of speaking raises an interesting question for the issue of “audience design” in production. In the simultaneous presence of the visual display, because the attributes in the descriptions are incrementally processed in sequence, the later inclusion of additional attributes is likely to delay the execution of referent selection when earlier attributes alone fully discriminate relative to a more concise description without them. The question for future research therefore concerns whether speakers take this into account and adapt their descriptions accordingly to help their listener. For example, we may say that, when giving instructions for immediately available contexts (e.g., giving driving instructions), cooperative speakers should be concise, and they should mention the most salient distinguishing attribute first. By contrast, when the referent is located in a different place from where the instruction takes place (e.g., giving shopping instructions), speakers could over-specify, mentioning salient information that could eliminate alternatives more quickly.

Now, the current study focused on English, which is a prenominal modifier language, where the modifier precedes the noun. One may wonder whether over-specification might affect referential processing differently in post-nominal modifier languages, where the modifier follows the noun. Using eye-tracking, Rubio-Fernandez et al. (2020) showed that listeners fixated the target more when the modifier rather than the noun was discriminating in English, whereas the opposite was the case in Spanish, which is a post-nominal modifier language (see also Rubio-Fernandez & Jara-Ettinger, 2020). That is, in both English and Spanish, as soon as listeners heard the first attribute, whether it was the modifier (English) or noun (Spanish), they immediately launched a search for the referent; target fixations were boosted when the first-mentioned attribute, rather than the second-mentioned attribute, discriminated the referent. This is in line with findings from Eberhard et al.

(1995) and our current study, as well as with other studies showing that language processing proceeds incrementally across different languages (e.g., Kamide, Altmann, & Haywood, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Özge, Küntay, & Snedeker, 2019). With particular reference to the incremental processing of over-specification across different languages, based on our findings, we predict that in the simultaneous presence of the visual display, if an earlier-mentioned attribute fully discriminates, the late inclusion of an additional attribute is unlikely to speed up the execution of referent selection relative to a shorter description without the additional attribute, even when it increases target fixations (Experiment 2A). By contrast, when listeners search for the referent after having heard the entire referring expression so all the attributes could be used in parallel from the outset, an additional attribute could be advantageous, depending on how quickly the added attribute discriminates relative to alternative attributes mentioned in a shorter description (Experiment 2B).

In a cross-linguistic production experiment Rubio-Fernandez (2019) found that speakers of English over-specify color more often than speakers of Spanish when the noun fully discriminates. This group difference was interpreted by the author as arising from speakers' cooperativeness towards their addressees: Speakers of English over-specified color more than speakers of Spanish because color adjectives are more useful for referent search when occurring pre-nominally (in English) rather than when post-nominally (in Spanish). However, an alternative explanation is also possible: Color over-specification arises from production-internal constraints (e.g., Pechman, 1989), rather than speakers' cooperativeness towards their listeners. Findings by Brown-Schmidt and Konopka (2008) support this alternative explanation. Using eye-tracking, Brown-Schmidt and Konopka showed that Spanish-English bilinguals fixated a size contrast earlier when producing a size modifier pre-nominally in English rather than when producing it post-nominally in Spanish, demonstrating the highly incremental nature of message encoding: Speakers plan the size adjective earlier in English because in English adjectives precede the noun, and later in Spanish because

adjectives follow the noun in Spanish. In Rubio-Fernandez' study, the visual display was timed, but this does not rule out the possibility that, because of the cross-linguistic difference in noun-modifier order, speakers of Spanish focused on the nominal categories earlier than speakers of English, and this made it easier for Spanish speakers to realize, before starting to speak, that the noun alone was fully discriminating. Indeed, Rubio-Fernandez et al. (2020) reported that in both English and Spanish, the rates of color over-specification increased as the set size in the display increased. Such finding corroborates the view that color over-specification cross-linguistically arises from incremental production processes; it occurs more frequently when speakers are less certain as to whether the noun alone fully discriminates before starting to speak (cf. Pechman, 1989).

We should also note that the additional mention of color might not necessarily be advantageous relative to a shorter, noun-alone description. As discussed above, whether the additional mention of color speeds up referent selection relative to a noun-only description is dependent on how quickly the noun alone discriminates. In the current study, the additional mention of color led to faster referent selection relative to a pattern-only description under parallel processing (Experiment 2B) even when color was only partially discriminating. This was because color discriminated much faster than pattern, so the mention of color quickly eliminated some alternatives, thereby reducing the set size for pattern. Belke and Meyer (2002) found no evidence that object class (noun information) discriminates more slowly than color (though they did find that color discriminates faster than size). The current study showed that listeners tend to attend to all the attributes mentioned in the referring expression before finalizing referent selection, and this is one of the reasons why over-specification can hinder comprehension. The additional mention of color probably boosts fixations to the target but it also increases the number of attributes to be processed. Hence, if the noun alone discriminates fast in the context, additional color mention may not speed up referent selection compared to a shorter, noun-only description even under parallel processing, let alone under incremental processing.

Unlike the ordering of color and pattern, the relative ordering of other attributes has been shown to be semantically or syntactically restricted (e.g., Culbertson, Schouwstra, & Kirby, in press; Dixon, 1982; Hetzron 1978; Quirk, Greenbaum, Leech, & Svartik, 1985; Scontras, Degen, & Goodman, 2017; 2019; Scott, 2002; Trotzke & Wittenberg, 2019; Whorf, 1945). Thus, one may wonder to what extent the current findings depend on the fact that the ordering of color and pattern attributes can be varied more freely than the ordering of, for example, size and color. Specifically, during the reviewing process, a reviewer was concerned that in Experiment 2A, the inclusion of color after pattern did not facilitate referent selection because listeners might have inferred that the second-mentioned color was irrelevant given that color was not mentioned first. That is, if the color adjective was over-specified after a size adjective, could it have been facilitatory, knowing that color does not normally precede size in language with prenominal modification (e.g., Belke, 2001; Danks & Schwenk, 1972)? Although this is an interesting idea, what we found in Experiment 2A was both inhibition (mouse clicking) and facilitation (increased target fixations); moreover, in Experiment 2B, the late mention of color clearly facilitated referent selection; it sped up referent selection as well as boosted fixations to the target. Hence, we see no obvious reason to suspect that the current findings do not generalize across different adjective orders. We predict that, under incremental processing, if the size adjective alone clearly discriminates (e.g., there is only one large bow in the context of only small bows), the additional mention of color is likely to delay referent selection relative to the size-only description, whereas it can speed up referent selection under parallel processing, assuming that color discriminates faster than size (Belke & Meyer, 2002). Of course, we must add the caveat that there may be situations where over-specification could be helpful even under incremental processing. For instance, if the first-mentioned attribute is technically fully discriminating but does not discriminate clearly because the display contains more alternatives than in the current study and/or if the referential candidates are too similar to be discriminated by the first attribute, then the

additional mention of a salient discriminating attribute might not only enhance looks but also shorten the times for referent selection. But this is a question for further research.

To conclude, whether over-specification is “a waste of time” depends not only on the discriminability or salience or the relative order of the attributes, but also on the processing constraints that determine the way in which those attributes can be mapped onto the context. The additional inclusion of a highly salient or *fast discriminating* attribute can facilitate referential processes, but this critically depends on whether listeners are able to use such an attribute early, either because it is mentioned first for incremental processing or because the attributes can be used in parallel.

Supplementary materials

Analysis scripts, data, a list of materials and display images can be found here:

https://osf.io/vwse2/?view_only=ac5e339fb9064ee0b06a2d443800d52b

Acknowledgements

We thank all the students who contributed to the data collection for this study and Keith Edwards for his help with implementing the experiments. Kumiko Fukumura was supported by the Leverhulme Trust [RPG-2016-253].

References

- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.
- Arai, M., Van Gompel, R. G. P., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*, 218–250.
- Arts, A., Maes, A., Noordman, L. G. M., & Jansen, C. (2011a). Overspecification facilitates object identification. *Journal of Pragmatics*, *43*, 361-374.
- Arts, A., Maes, A., Noordman, L.G.M., & Jansen, C. (2011b). Overspecification in written instruction. *Linguistics*, *49*, 555-574.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.
- Baayen, R.H. (2008). *Analyzing Linguistic Data. A practical introduction to statistics*. Cambridge University Press.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.
- Bates, B., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.
- Bates, D., Kliegel, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. Unpublished manuscript. Downloadable at <https://arxiv.org/pdf/1506.04967.pdf>
- Belke, E. (2001). *On the time course of naming multidimensional objects in a referential communication task: Analyses of eye movements and processing times in the production of complex object specifications*. Doctoral dissertation, University of Bielefeld, Germany
- 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.
- Bierwisch, M. (1987). The semantics of gradation. In M. Bierwisch, & E. Lang (Eds.). *Dimensional Adjectives* (pp. 471–514). Berlin: Springer-Verlag.
- Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1–47.
- Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequenas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition, 109*(2), 274-280.
- Culbertson, J., Schouwstra, M., and Kirby, S. (in press). From the world to word order: deriving biases in noun phrase order from statistical properties of the world. *Language*.
- Dale, R. (1989). Cooking up referring expressions. In *Proceedings of the 27th annual meeting on Association for Computational Linguistics (ACL'89)* (pp. 68-75). Vancouver, BC: Association for Computational Linguistics.
- Dale, R. (1992). *Generating referring expressions: Building descriptions in a domain of objects and processes*. Cambridge: MIT Press.
- Dale, R., & Reiter, E. (1995). Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science, 19*, 233-263.
- Danks, J. H., & Schwenk, M. A. (1972). Pre-nominal adjective order and communication context. *Journal of Verbal Learning and Verbal Behavior, 11*, 183-187.
- Davies, C. & Katsos, N. (2013). Are speakers and listeners ‘only moderately Gricean’? A response to Engelhardt, Bailey and Ferreira (2006). *Journal of Pragmatics, 49*, 78-106.

- Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A Bayesian approach to “overinformative” referring expressions. *Psychological Review*. Advance online publication.
- Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, *11*, 159-184.
- Dixon, R. (1982). *Where have all the adjectives gone? And other essays in semantics and syntax*. Amsterdam: Mouton de Gruyter.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*, 409-436.
- Engelhardt, P. E., Demiral, Ş. B., & Ferreira, F. (2011). Over-specified referring expressions impair comprehension: An ERP study. *Brain and Cognition*, *77*, 304-314.
- Engelhardt, P.E., & Ferreira, F. (2014). Do speakers articulate over-described modifiers differently from modifiers that are required by context? Implications for models of reference production, *Language and Cognitive Processes*, *29*, 975-985.
- Engelhardt, P.E., Bailey, K., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*, 554-573.
- Field, A., Miles, J. & Field, Z. (2012) *Discovering Statistics Using R*. London: Sage Publications.
- Frank, M., & Goodman, N. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998.
- Fukumura, K. (2018). Ordering adjectives in referential communication. *Journal of Memory and Language*, *101*, 37-50.
- Fukumura, K., & Van Gompel, R. P. G. (2017). How do violations of Gricean maxims affect reading? *Journal of Memory and Language*, *95*, 1–18.

- Gatt, A., Krahmer, E., Van Deemter, K., & Van Gompel, R.P.G. (2016). Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science*, *41*, 1457–1492.
- Gelman, A., & Hill, J. (2007). *Data Analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics: Vol. III. Speech acts* (pp. 41–58). New York: Academic Press.
- Hetzron, R. (1978). On the relative order of adjectives. In: H. Seiler (Ed.), *Language Universals* (pp. 165-184). Tübingen: Gunter Narr Verlag.
- Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133–156.
- Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The Effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, *37*, 395-411.
- Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, *95*, 95–127.
- Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition*, 1-71.
- Martin, J. E. (1969). Semantic determinants of preferred adjective ordering. *Journal of Verbal Learning and Verbal Behaviour*, *8*, 697–704.
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*, 372–380.
- McDonald, J., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, *25*, 188–230.

- Monroe, W., Hawkins, R. X., Goodman, N. D., & Potts, C. (2017). Colors in context: A pragmatic neural model for grounded language understanding. *Transactions of the Association for Computational Linguistics*, 5, 325-338.
- Özge, D., Küntay, A., & Snedeker, J. (2019). Why wait for the verb? Turkish speaking children use case markers for incremental language comprehension. *Cognition*, 183, 152–180.
- Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, 27, 89-110.
- Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29, 633–654.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- R Core Team (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Prentice-Hall, Inc.
- Reiter, E. (1990). The computational complexity of avoiding conversational implicatures. In *Proceedings of the 28th annual meeting on Association for Computational Linguistics* (pp. 97-104). Association for Computational Linguistics.
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. *Perception*, 33, 217- 236.
- Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, 7, Article 153
- Rubio-Fernandez, P. (2019). Overinformative speakers are cooperative: Revisiting the Gricean Maxim of Quantity. *Cognitive Science*, 43.
- Rubio-Fernandez, P., & Jara-Ettinger, J. (2020). Incrementality and efficiency shape pragmatics across languages. *Proceedings of the National Academy of Sciences*.

- Rubio-Fernandez, P., Mollica, F., & Jara-Ettinger, J. (2020, in press). Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. *Journal of Experimental Psychology: General*.
- Schriefers, H., & Pechmann, T. (1988). Incremental production of noun-phrases by human speakers. In M. Zock & G. Sabah (Eds.), *Advances in natural language generation* (pp. 172-179). London: Pinter.
- Scontras, G., Degen, J., & Goodman, N. D. (2017). Subjectivity predicts adjective ordering preferences. *Open Mind, 1*, 53-66.
- Scontras, G., Degen, J., and Goodman, N.D. (2019). On the grammatical source of adjective ordering preferences. *Semantics & Pragmatics, 12*, pp. 1-21.
- Scott (2002). Stacked adjectival modification and the structure of nominal phrases. In G. Cinque (Ed.), *Functional structure in DP and IP: The cartography of syntactic structures* (Vol.1, pp.91-120). New York: Oxford University Press.
- 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., Tanenhaus, M., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*, 109–147.
- Singmann, H., & Kellen, D. (2020). An Introduction to Mixed Models for Experimental Psychology. In D. H. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology* (pp. 4–31). New York: Routledge.
- Sonnenschein, S., & Whitehurst, G.J. (1982). The effects of redundant communications on the behaviour of listeners: does a picture need a thousand words? *Journal of Psycholinguistic Research, 11*, 115–125.
- Spivey, M. J., Tyler, M. J., Eberhard, K.M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science, 12*, 282–286.

- Tanaka, M. N., Branigan, H. P., Mclean, J. F., & Pickering, M. J. (2011). *Conceptual influences on word order and voice in sentence production: Evidence from Japanese. Journal of Memory and Language, 65*, 318-330.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632.
- Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: Saliency, absoluteness, and consistency. *Frontiers in Psychology, 6*, 1-16.
- Tourtour, E., Delogu, F., Sikos, L., & Crocker, M. W. (2019). Rational over-specification in visually-situated comprehension and production. *Journal of Cultural Cognitive Science, 3*, 175–202.
- Trotzke, A., & Wittenberg, E. (2019). Long-standing issues in adjective order and corpus evidence for a multifactorial approach. *Linguistics, 57*, 273-282.
- Van Gompel, R.P.G., Van Deemter, K., Gatt, A., Snoeren, R., & Krahmer, E. (2019). Conceptualisation in reference production: Probabilistic modelling and experimental testing. *Psychological Review, 126(3)*, 345-373.
- Westerbeek, H., Koolen, R., Maes, A. (2015). Stored object knowledge and the production of referring expressions: the case of color typicality. *Frontiers Psychology, 6*, 1–12.
- Whorf, B.L. (1945). Grammatical categories. *Language, 21*, 1-11.