Color interpretation is guided by informativity expectations, not by world knowledge about colors

Hannah Rohde[a,d], Paula Rubio-Fernandez[b,c,d]

[a] *Linguistics & English Language, University of Edinburgh*

[b] *Philosophy, Classics, History of Art & Ideas, University of Oslo*

[c] *Brain & Cognitive Sciences, Massachusetts Institute of Technology*

[d] *Both authors contributed equally to this work.*

ARTICLE INFO

Keywords:

language comprehension

overspecification

color typicality

Bayes Rule

eye tracking

ABSTRACT

When people hear words for objects with prototypical colors (e.g., 'banana'), they look at objects of the same color (e.g., lemon), suggesting a link in comprehension between objects and their prototypical colors. However, that link does not carry over to production: The experimental record also shows that when people speak, they tend to *omit* prototypical colors, using color adjectives when it is informative (e.g., when referring to clothes, which have no prototypical color). These findings yield an interesting prediction, which we tested here: while prior work shows that people look at yellow objects when hearing 'banana', they should look *away* from bananas when hearing 'yellow'. The results of an offline sentence-completion task (N=100) and an online eye-tracking task (N=41) confirmed that when presented with truncated color descriptions (e.g., 'Click on the yellow...'), people anticipate clothing items rather than stereotypical fruits. A corpus analysis ruled out the possibility that this association between color and clothing arises from simple context-free co-occurrence statistics. We conclude that comprehenders make linguistic predictions based not only on what they *know* about the world (e.g., which objects are yellow) but also on what speakers tend to *say* about the world (i.e., what content would be informative).

1. Introduction

Finishing each other's sentences is often taken as a sign of knowing each other well. However, interlocutors' synchronization applies not only to people who are particularly close: decades of psycholinguistics research reveals that language processing is, by its very nature, *predictive* (Kuperberg and Jaeger, 2016). Language comprehenders demonstrate remarkable expectation-driven processing, as revealed by a range of methodologies, from priming (McNamara, 2005) and reading (Van Berkum,

*Corresponding author: Hannah Rohde, Hannah.Rohde@ed.ac.uk

For this study's pre-registrations, as well as materials, data and analysis code, see:
Pre-registration Experiment 2: `https://osf.io/djzwc`
Pre-registration Experiment 3: `https://osf.io/zcwvs`
All materials, data, analysis: `https://osf.io/sja5b`

Brown, Zwitserlood, Kooijman and Hagoort, 2005) to brain responses (DeLong, Urbach and Kutas, 2005) and anticipatory looking (Kamide, Altmann and Haywood, 2003).

Comprehenders use prediction at all levels of linguistic structure, relying on dependencies between sounds (DeLong et al., 2005), words (Kamide et al., 2003), syntactic structures (Levy, 2008), and propositions (Rohde and Horton, 2014). At the word level, comprehenders are particularly adept at using their world knowledge to activate relevant concepts. For example, people predict that 'the mechanic checked...' is more likely to be followed by 'the brakes' than 'the spelling', whereas they make the reverse prediction if the agent is 'the journalist' (Bicknell, Elman, Hare, McRae and Kutas, 2010).

However, in production, speakers do not necessarily use language to talk about predictable content. On the contrary, speakers are known to *omit* content that is prototypical or inferable in a situation. What speakers favor instead is content that is non-inferable (e.g., an atypical instrument like 'sweeping with a tree branch'; Brown and Dell, 1987; Lockridge and Brennan, 2002). This pattern of omission aligns with information-theoretic accounts that link predictability to reduction (Aylett and Turk, 2004; Levy and Jaeger, 2007) and more broadly to expectations of speaker cooperativity (Grice, 1975; Sperber and Wilson, 1986). Here we investigate comprehenders' sensitivity to this informativity-driven bias in production: Do comprehenders go beyond their knowledge of the world (i.e. what is typical) to anticipate upcoming words that reflect their knowledge of informativity (i.e. what is newsworthy)?

We focused on color words as a test case, and used truncated instructions to probe comprehenders' expectations. Consider (1).

(1) Can you pass me the yellow ...

To anticipate the next word, comprehenders may rely on their knowledge about the color of objects in the world: Eye-tracking studies show that comprehenders look at yellow things when hearing a word for a typically yellow object (e.g., a lemon when hearing 'banana'; Huettig and Altmann, 2011; see also Naor-Raz, Tarr and Kersten, 2003), suggesting access to probabilities that are determined by real-world knowledge (e.g., about the color of fruits).[1] Then again, if comprehenders hear a color word and assume the speaker is being informative, they may anticipate that the speaker is talking about an

[1] It should be noted that the studies we cite here do not specifically test anticipatory processing; rather, they establish comprehenders' awareness of prototypical colors associated with certain types of objects. Such studies aim to model comprehenders' conceptual representation by testing the accessibility of such features during processing. We cite them here as examples that demonstrate the availability of color information when a comprehender hears the mention of a noun – information which we speculate may serve additionally as the kind of knowledge available for generating predictions.

object that has low probability of possessing that color. This yields an interesting empirical prediction: While prior work has shown that comprehenders who hear the word ‘banana’ look at yellow objects, comprehenders who hear the word ‘yellow’ may instead look away from bananas and towards non-prototypically-yellow objects in the scene.

2. World knowledge and contrastive uses of color adjectives

If language comprehension and production are simply grounded in real-world plausibility, then speakers would be expected to produce utterances about plausible scenarios and comprehenders would anticipate content along those same lines. Such a transparent emphasis on real-world knowledge could be represented via the formula in (2). Here the left side of the formula captures a comprehender’s problem of guessing what object is being referred to given that they have heard a speaker mention the adjective ‘yellow’; this conditional probability is estimated as the raw probability that a candidate $object_i$ is typically yellow. For example, the comprehender’s guess that the yellow object being referred to is a banana would be higher than their guess that the object is a strawberry because $p(object_{banana,\text{COLOR}=yellow})$ is higher in the real world than $p(object_{strawberry,\text{COLOR}=yellow})$.

(2) Hypothesis: Comprehension reflects real-world priors alone

$$p(referent = object_i | \text{‘yellow’}) \propto p(object_{i,\text{COLOR}=yellow})$$

We call this the World Knowledge Hypothesis, whereby comprehenders use their experience with the world directly to anticipate what speakers will talk about. Indeed, there is evidence that comprehenders do rely on real-world, plausibility-driven priors during general language interpretation (Warren and Dickey, 2021). Under the account sketched in (2), comprehenders estimate that speakers use language to talk transparently about the world: Situations that are frequent in the world give rise to frequent utterances about those situations in speech. Bananas are frequently yellow and so speakers will encounter many yellow-banana situations and produce utterances describing those situations, and comprehenders will come to anticipate such utterances.

However, we know that such a model of transparent language production does not hold, at least not for speakers’ use of color adjectives. Production studies show that while speakers do often use color redundantly (i.e., they specify an object’s color even when color is unnecessary for disambiguating the intended referent; Pechmann, 1989), they rarely do so for objects with prototypical colors. Bananas are prototypically yellow but their inherent yellowness is rarely mentioned descriptively, whereas objects that lack an inherent color are described with redundant color modifiers quite often (e.g., a yellow notebook, Sedivy, 2003; Rubio-Fernandez, 2016; Westerbeek, Koolen and Maes, 2015). Given this production behavior on the part of speakers, a question arises as to whether comprehenders are aware of this speaker behavior and make use of it when anticipating an intended referent.

A partial answer to this question comes from comprehension studies on the contrastive use of color adjectives. In contexts with two objects of the same category (known as competitors; e.g., two bananas), comprehenders are able to make use of a speaker's mention of a color modifier in a way that suggests they are drawing what is called a *contrastive inference*, and the strength of this inference depends on color typicality (Sedivy, Tanenhaus, Chambers and Carlson, 1999; Sedivy, 2003). A contrastive inference can be illustrated as follows: In a scene containing two bananas (one yellow and one green) and a notebook (yellow), a speaker who intends to refer to the yellow banana would be expected to use a color modifier to disambiguate which banana they mean. Therefore, a comprehender who hears a speaker start the instruction 'Pick up the yellow...' may anticipate that the speaker is using the color adjective contrastively in referring to the yellow banana, rather than descriptively in referring to the yellow notebook, even though the instruction is compatible with a mention of either object. This contrastive inference (whereby comprehenders look anticipatorily to the colored object that has a category competitor rather than the one that is a singleton) is strongest for color modification that is prototypical of the object (e.g., 'yellow banana'). The comprehender appears to be aware that speakers are generally disinclined to include color modifiers for such objects unless there is a cooperative communicative reason (such as disambiguation), and this awareness permits an inference regarding the intended referent. The same instruction 'Pick up the yellow...' does not trigger anticipatory looks when neither yellow object has a prototypical color (e.g., a singleton mug and a ruler with a competitor; Sedivy, 2004; Rubio-Fernandez, Mollica and Jara-Ettinger, 2021).

These results suggest that comprehenders expect prototypical colors to be used contrastively (why else would a speaker mention a banana's yellowness?), whereas non-prototypical colors are expected to be used descriptively (Sedivy, 2003, 2004). In line with this eye-tracking literature, a recent study using a sentence-completion task showed that comprehenders expect color adjectives to be used descriptively rather than contrastively when color is atypical (e.g., yellow strawberry; Kreiss and Degen, 2020). These studies suggest that comprehenders are aware not only of objects' real-world color priors but also of speakers' likelihood of color mention. However, the above studies used contexts in which contrastive inferences were available. It is therefore possible that comprehenders only appeal to their knowledge of color mention likelihoods in contexts that necessitate disambiguation between two objects of the same category. The question therefore remains as to whether comprehenders make use of their knowledge of color mention likelihoods more generally, even in contexts where color modification is redundant and no contrastive inference is at stake.[2]

3. Informativity and redundant uses of color

To formally capture that comprehenders make use of their knowledge of color mention likelihoods, an adaptation of (2) is shown in (3) with an additional term for the production likelihood. In (3), the left side of the formula again captures the comprehender's problem of guessing what object is being referred to given that they've heard the adjective 'yellow'; in this case, however, the estimate on the right side consists of terms for both the prior (the probability of a particular object being yellow) and the likelihood (the probability that a speaker would include a color modifier for 'yellow' when this particular object is yellow).

(3) Hypothesis: Comprehension reflects real-world priors and speaker production likelihoods

$$p(\textit{referent} = \textit{object}_i \,|\, \text{`yellow'}) \propto p(\textit{object}_{i,\text{COLOR}=\textit{yellow}}) * p(\text{`yellow'} \,|\, \textit{object}_{i,\text{COLOR}=\textit{yellow}})$$

In contrast to the World Knowledge Hypothesis depicted in (2) where hearing 'yellow' yields an expectation for a mention of a prototypically yellow object, (3) can be said to capture an Informativity Hypothesis whereby comprehenders make use of their knowledge of speakers' color modification behaviors – namely that speakers' rate of modification varies in systematic ways depending on the inferability of the color and the need for referent disambiguation.

Crucially, comprehenders need to estimate speakers' color modification likelihoods with an awareness of the objects' color prototypicality and the context-specific redundancy/disambiguation associated with the use of a color modifier in a given scene. In a context with two objects of the same kind that differ in color (e.g., a yellow banana and a green banana), color modification is highly likely. In a context in which color modification is redundant (e.g., one with a yellow banana and a yellow notebook), color prototypicality would be expected to play a larger role.

[2] Sedivy reports a control condition where neither the yellow banana nor the yellow notebook had a competitor. In that condition, the color adjective 'yellow' is redundant and its presence did yield a marginal effect in favor of the non-prototypically yellow object (Sedivy, 2003, Footnote 5, p.17), suggesting that comprehenders are potentially sensitive to the different production likelihoods of prototypical and non-prototypical colors even when no contrastive inference is at play. However, this reported finding may still reflect participants' strategies in a task in which contrastive inferences were at play in many trials.

In addition, even among objects that lack a prototypical color, speakers tend to overspecify color more often when color is a feature that is strongly associated with that category, than when it is not (e.g., compare how important a property color is for clothing vs appliances; Rubio-Fernandez, 2016, 2019). Language comprehension studies have confirmed that descriptive color adjectives (i.e., those that are redundant) can facilitate comprehenders' visual search for the referent (Paraboni, Van Deemter and Masthoff, 2007; Arts, Maes, Noordman and Jansen, 2011; Tourtouri, Delogu, Sikos and Crocker, 2019; Rubio-Fernandez, in press), suggesting their use is cooperative and supports efficient communication (Rubio-Fernandez, 2016, 2019; Long, Rohde and Rubio-Fernandez, 2020). However, the tendency to mention the color of clothes seems so strong as to override efficiency considerations: When presented with a monochrome display of clothes, speakers produced expressions like 'the yellow shirt' around 40% of the time (despite color being a useless visual cue; Rubio-Fernandez, 2016), whereas they never produced 'the yellow triangle' in a monochrome display of shapes, and did so 40% of the time in polychrome displays (where color was an effcient visual cue, despite being formally redundant; Rubio-Fernandez, 2019). These speaker preferences can be captured in the likelihood term of the Informativity Hypothesis depicted in (3).

For example, under (3), a comprehender who sees a yellow banana and a yellow notebook and hears the speaker mention 'yellow' will estimate that the speaker is more likely to be talking about the notebook: Such a computation reflects the knowledge that the prior probabilities of yellow bananas and yellow notebooks differ (with a higher probability of a banana being yellow than a notebook) and the knowledge that the likelihood of mentioning 'yellow' for these objects also differs (with speakers almost never describing bananas as yellow redundantly, but doing so quite frequently for yellow notebooks). Note that (3) is not specifically a model of language production; it is simply Bayes' Rule, a theorem that establishes that the conditional probability on the left is proportional to the combination of the prior and likelihood on the right (for other applications in linguistics, see Kehler and Rohde, 2013; Franke and Jäger, 2016; Goodman and Frank, 2016; Rohde and Kurumada, 2018).

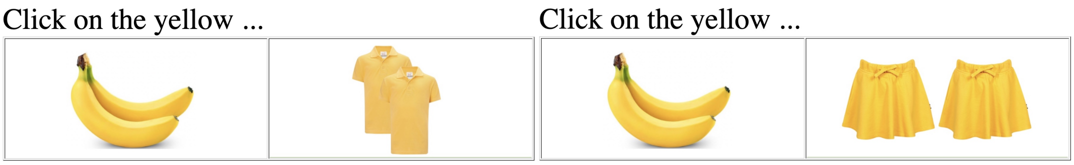
If predictive language processing depends in part on comprehenders' ability to reverse engineer what utterance a speaker is likely to be producing, casting the interpretation problem as in (3) highlights the two factors which determine the probability that a speaker is talking about a particular object in a context in which the mention of color would be purely descriptive. In that case, these two terms are at odds if a speaker starts an utterance 'Can you pass me the yellow...'. When estimating what object the speaker is requesting, a yellow strawberry is improbable (low situation prior) but if there were a yellow strawberry, that is the type of newsworthy color that a speaker might be likely mention (high production likelihood) (for related models, see Kreiss and Degen, 2020; Rohde, Futrell and Lucas, 2021).

If it is the case that comprehenders are aware of speakers' context-specific and category-specific color modification behavior and if they estimate which object will be mentioned via a Bayesian formulation like (3), we would expect to see evidence that color adjectives like 'yellow' can yield anticipation for upcoming mention of an object for which color is a strongly associated feature but has no prototypical value (e.g., clothing) over one for which color has a prototypical inherent value (e.g., fruit). If, on the other hand, comprehenders rely primarily on their real-world knowledge, as formulated in (2), we would expect adjectives like 'yellow' to yield anticipation for upcoming mention of a prototypically yellow object.

4. Outline of the study

We present two comprehension studies: a web-based offline experiment and an online eye-tracking experiment. In both tasks, participants indicated which object they anticipated after a truncated instruction like 'Click on the yellow...' when presented with prototypically colored foods (e.g., two yellow bananas) alongside non-inherently colored clothing items (e.g., two yellow shirts or two yellow skirts). We manipulated the redundant word in the instruction in order to compare responses to a color adjective ('yellow') and a baseline number ('two'). In addition to this within-subject Instruction manipulation, a between-subjects manipulation varied the assignment of non-prototypical colors to objects (see Figure 1). This Pairing manipulation was included in Experiment 1 (and extended in Experiment 2 with the use of new pictures) to better ensure that our findings were not attributable to the specific combinations of foods and clothes, or colors and clothes. To foreshadow our findings, across both experiments, the color adjective yielded a preference for the clothing over the prototypically colored food, in keeping with the Informativity Hypothesis.

Lastly, we provide a corpus analysis to addresses a possible concern that the observed behavior in Experiments 1 and 2 may be driven by simple co-occurrence statistics rather than by comprehenders' pragmatic awareness of speakers' context-specific and category-specific production behaviors. This alternative explanation is ruled out by showing, via the Google ngram corpus (Brants, Franz et al., 2006), that reliance on context-free co-occurrence statistics would predict behavior more in line with the World Knowledge Hypothesis.



(a) Screenshot of pairing 1

(b) Screenshot of pairing 2

Figure 1: Sample item displays with pairs of yellow objects. The pairing manipulation varies which clothing item is presented with which food (banana~shirts versus banana~skirts). The instructions manipulation varies the type of redundant word (color versus number); both screenshots (a) and (b) show the color instruction (number instruction would appear as 'Click on the two ...').

Our findings go beyond previous work on comprehenders' abilities to make linguistic predictions based on their conceptual knowledge (i.e. what they know about the world) to show an additional ability to track speakers' use of language (i.e. what speakers tend to say about the world), and to do so even in contexts in which no contrastive inference is triggered by the need for disambiguation via color modification.

5. Experiment 1: Anticipation of object mention [pre-registered]

Experiment 1 tested the World Knowledge and Informativity Hypotheses on color interpretation through a web task that elicited sentence continuations following an ambiguous color adjective (e.g., in Pairing 1, 'Click on the yellow...' was presented together with two yellow bananas and two yellow shirts and participants had to select the intended referent; see Figure 1).

5.1. Data Availability

All materials, data and analysis scripts for the studies reported here are available at an Open Science Framework repository: `https://osf.io/sja5b`

5.2. Methods

Participants

112 participants were recruited from the crowdsourcing platform Prolific and paid £3. Using Prolific's demographic selections, we recruited participants who were over 18, of US/UK Nationality, and monolingual English speakers. We excluded 12 participants who were not above chance on filler trials that had a correct response (n=4) or who did not complete the task (n=8). This yielded a dataset of 100 participants, 50 each for Pairings 1 and 2. Target participant numbers were estimated from an earlier pilot study.

Materials

Each item consisted of a visual scene with two pairs of objects and a truncated text instruction containing either a color word or the number word 'two' (see Figure 1). The 20 experimental items depicted familiar foods and clothing items; all images for the study are available at an Open Science Framework repository (see `https://osf.io/sja5b`). The objects appeared in pairs in order to make the instructions sound natural in both the color and number conditions (e.g., the formulation 'the one banana' may have raised the question of why the speaker didn't use the simpler formulation 'the banana'). The instructions for the experimental items contained 10 color words, each used twice within each Pairing. The food∼clothing pairings were selected semi-randomly, with replacements when a particular color∼clothing pairing corresponded to a brand name or fashion term that might make the clothing item more predictable (e.g., 'Red Hat Linux', 'black tie'). For the experimental items, the position of the two pictures in each trial was counterbalanced (e.g., whether the bananas appeared on the left and the shirts on the right, or vice-versa). The color∼food combinations were dictated by the foods themselves (e.g., purple eggplants, orange carrots). Counterbalancing of the visual presentation and the Instruction and Pairing manipulations resulted in 8 lists of materials.

The 20 experimental items were interleaved with an additional 35 filler items. Ten of these were similar to the experimental items with instructions that contained an ambiguous color word with no correct response ('Click on the yellow...'), but the images were either both food (5) or both clothing (5). The remaining 15 fillers had a correct response cued by a color word (5), a number word (5), or an object category word (5). The position of the correct image was balanced across the filler types.

For all items, the instructions appeared as text above the images.

Procedure

Forced-choice responses were collected via a web-based interface that participants accessed from their own computer. Participants were redirected from the Prolific platform to a custom-made website for collecting responses, which was hosted on a University of Edinburgh server. Participants were told that they would be given the first few words that someone said and that their task was to click which object they though the speaker was talking about. Each item was presented on a page by itself with the two images displayed side by side. Participants were randomly assigned to one of the eight lists of materials. Trials were presented in one of a large set of possible pseudo-random orders (no two adjacent experimental trials). No reaction times were collected.

Predictions

Under the World Knowledge Hypothesis, the color instruction should yield a food bias (compared to the baseline number condition) because of the stable conceptual associations between foods and their prototypical colors. Under the Informativity Hypothesis, the color condition should instead yield a clothing bias, due to comprehenders' awareness of speakers' preference to produce redundant color adjectives for clothing items. Neither hypothesis predicts a main effect or interaction with Pairing since that manipulation is a control to vary the assignment of colors to clothing (i.e., the clothing images are the only images that change between Pairing 1 and Pairing 2). This manipulation was included to reduce the possibility that an observed bias towards clothing in the color condition could be attributed to the experimenters' selection of a particularly color-compatible clothing image (e.g., a particularly yellow pair of yellow shirts).

5.3. Results

The binary outcome of clothing versus food was analysed using a logistic mixed effects model with the lme4 package (Bates, Kliegl, Vasishth and Baayen, 2015) in R (R Core Team, 2016). The model contained fixed effects for Instruction and Pairing and their interaction, and random effects for participants and items (where 'item' corresponded to the food image that was common across all Instructions and Pairings). Maximal random effect structure was used according to the experimental design, with random intercepts for participants and items, random by-item slopes for Pairing, and random by-participant and by-item slopes for Instruction. We centered the factors Instruction (number -.5, color +.5) and Pairing (pairing1 -.5, pairing2, +.5).

As can be seen in Figure 2, participants show no response preference in the baseline number condition (56% clothing), whereas they favor clothing in the color condition (68% clothing). This main effect of Instruction (β=0.853, SE=0.232, Z=3.671, $p<0.001$) is expected under the Informativity Hypothesis. There was no main effect of Pairing (β=-0.364, SE=0.283, Z=1.288, p=0.20), nor an Instruction $\times$ Pairing interaction (β=0.168, SE=0.426, Z=0.394, p=0.69), thereby confirming that participants' click preferences were not driven by one particular set of pictures or assignment of colors to clothing.[3]

While the results of Experiment 1 offered support to the Informativity Hypothesis via the partic-

[3]Following a reviewer's suggestion, we also checked whether participant behavior varied across the experiment, given that the truncated instruction 'Click on the yellow...' isn't felicitous as a stand-alone expression and repeated exposure to it might have altered their responses. We constructed a mixed effects logistic regression as in the main text, with the addition of trial number as a continuous factor. In that model, the main effect of Instruction is still significant; we also see a marginal Instruction $\times$ Pairing $\times$ Trial Number interaction, whereby the main effect of Instruction is slightly stronger across trials in Pairing 1 and is slightly attenuated across trials in Pairing 2. Even at its lowest in the last tertile of Pairing 2, the bias to clothing in the color condition is still 61% compared with 53% in the number condition. This finding should not undermine our claim that the color condition favors upcoming mentions of clothing, but the results suggest that shorter tasks may help avoid the development of participant strategies across items and of course more natural materials are also preferable.

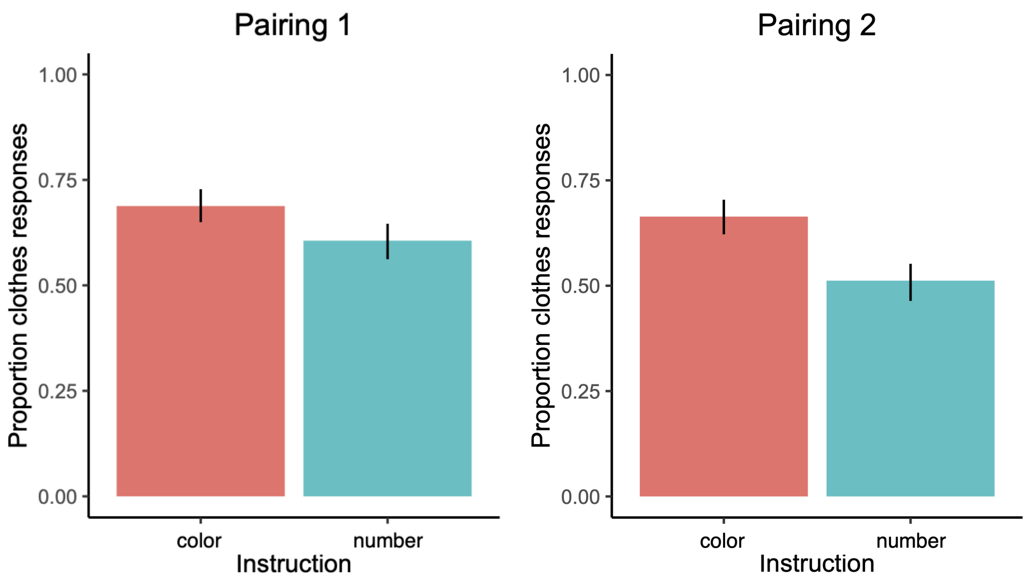


Figure 2: Experiment 1 clothing responses by Pairing and Instruction.

ipants' eventual response clicks, it is nonetheless possible that participants may have experienced an early bias towards the food items with prototypical colors, which we failed to detect with the offline task and which would have confirmed the predictions of the World Knowledge Hypothesis during earlier stages of processing. Eye-tracking studies showing that comprehenders make early fixations to color competitors when presented with objects with prototypical colors (e.g., hearing 'banana' and looking at a lemon; Huettig and Altmann, 2011) suggest that world knowledge about colors can have an early effect on processing. If this early world-knowledge bias was to be observed in our paradigm (followed by a later informativity bias, as shown in Experiment 1), our results would suggest that different information sources affect language processing at different stages. This possibility was investigated in Experiment 2 with the use of eye-tracking in the same sentence-completion task.

6. Experiment 2: Eye-tracking anticipation of object mention [pre-registered analyses]

To test whether comprehenders' early processing may conform to the predictions of the World Knowledge Hypothesis, the current experiment adopts the design and conditions from Experiment 1 in an eye-tracking paradigm that allows for detection of early effects in participants' real-time processing.

6.1. Methods

Participants

41 participants were recruited from the Subject Pool at MIT Brain and Cognitive Sciences Department and paid 10 USD. All participants were native speakers of English and reported having normal color vision. Target participant numbers were estimated based on similar eye-tracking studies on color inferencing.

Materials

The same food-clothing pairs used in Experiment 1/Pairing 1 were used in Experiment 2. However, while the depicted objects were the same, the photographs were different to avoid unwanted effects of the specific photographs employed in the study. As in Experiment 1, the materials included 20 critical items and 35 fillers. However, in order to test real-time language processing, we presented each instruction sentence in the auditory modality. The instructions were recorded by a female native speaker of American English. As in Experiment 1, instructions included color descriptions in the critical condition (e.g., 'Click on the yellow…') and a number description in the baseline condition ('Click on the two…'). The position of the two images in each pair was counterbalanced and each item was matched with two different instructions (one including a color adjective and the other including the number 'two'), which were fully crossed in 4 lists of materials. There was a 300ms preview window before the instructions started. Each pair of pictures was visible for 4000 ms (independent of participants' clicks).

Procedure

The procedure in Experiment 2 was the same as in Experiment 1. Participants were asked to complete a series of truncated instructions by clicking on one of two images on the screen. Eye movements were recorded with a portable eye-tracking system (RED-m by SMI) that measured eye position at a sampling rate of 120 Hz and had a spatial resolution (RMS) of 0.1° and an accuracy of 0.5°. Participants were seated about 60cm from the computer screen. The contact-free set up of the system allowed free head-movement during eye tracking. The experiment lasted 15 minutes.

Three measures were collected: responses (i.e. clicks on foods vs clothes), response times (RTs; measured from the onset of the instructions) and proportion of fixation time on the two images (foods vs clothes). Areas of Interest (AOIs) were delimited by applying the same square contour around all images in the experiment.

Predictions

Experiment 2 tested the same predictions as Experiment 1: Under the World Knowledge Hypothesis, the color instruction should yield a food bias relative to the baseline number condition, whereas under the Informativity Hypothesis, the color condition should instead yield a clothing bias. In addition, we use an online measure of language processing to assess whether in the color condition, participants' eye movements might reveal an initial food bias (driven by a bottom-up effect of world knowledge about colors) before they complete the instruction by selecting the clothes (as they did in their offline responses in Experiment 1). Such a pattern of results would suggest that world knowledge and informativity expectations (i.e., the prior and the likelihood terms in (3)) affect language prediction at different points during processing and that there is an early window in which the link between color and food (the World Knowledge Hypothesis) is present even if participants' eventual click response shows the pragmatic awareness that we hypothesize (the Informativity Hypothesis). Alternatively, participants' responses may be driven by their informativity expectations (i.e. redundant color adjectives are used to refer to clothes more often than to foods) from the earliest stages of processing, without suffering interference from the conceptual association between foods and prototypical colors (Huettig and Altmann, 2011; Naor-Raz et al., 2003).

In addition to eye movements, RTs also tap real-time processing. For example, participants may click on clothes more often in the color condition than in the number condition, but may also respond more slowly. Such a pattern of results would also suggest that estimating color informativity is cognitively costly relative to a neutral baseline.

6.2. Results

Responses

As in Experiment 1, participants' binary choices between foods and clothes were analyzed using a logistic mixed-effects model with a fixed effect of Instruction (Color vs Number) and the maximal random effect structure for Participants and Items, including random intercepts for Participants and Items and random by-participant and by-item slopes for Instruction. Replicating the results of Experiment 1, and supporting the Informativity Hypothesis, there was a significant main effect of Instruction, with more clicks recorded on clothes in the Color condition (75% clothing responses) than in the Number condition (55% clothing responses; β=-1.067, SE=0.297, Z=-3.590, p<0.001; see Panel A in Figure 3).

Response Times

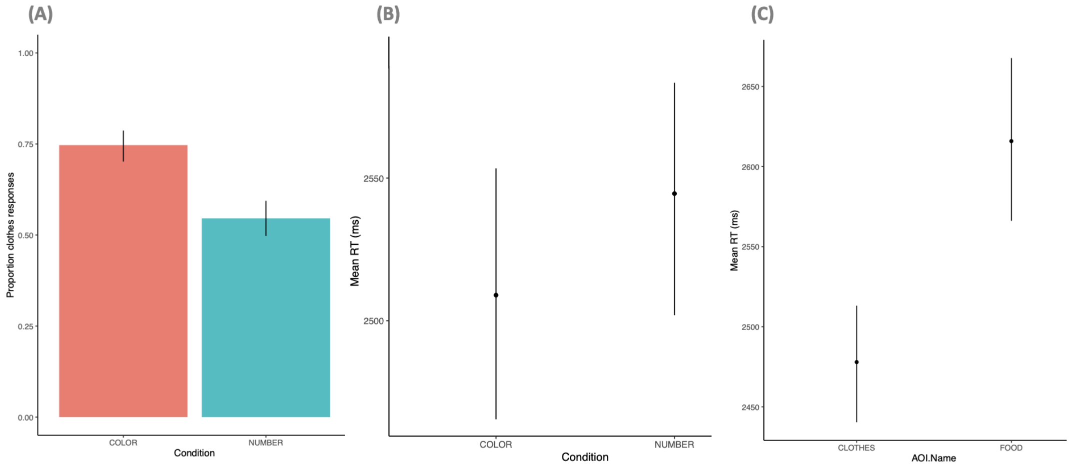


Figure 3: Experiment 2 clothing responses by Instruction (A,) RTs by condition (B) and RTs by object selected (C).

An analogous linear mixed-effects model was conducted on participants' RTs, but here the main effect of Instruction was not significant (β=39.79, SE=32.79, df=32.63, t=1.214, p=0.234), suggesting that participants made their picture selection at a comparable speed when they heard a color or a number description (see Panel B in Figure 3).

Following our pre-registration, we performed an exploratory RT analysis modelling the effect of Instruction and Response. That is, we investigated whether participants were faster or slower depending on what type of Instruction they heard (color vs number) and which picture they selected (clothes vs foods). For this analysis we used a linear mixed-effects model with fixed effects of Response and Instruction plus their interaction and the maximal random effect structure (because of model convergence issues, the interaction term between Instruction and Response was removed from the random effect structure of Item). The only significant result was a main effect of Response (β=110.30, SE=50.34, df=41.01, t=2.191, p=0.034), with faster RTs observed for clothes than for foods. The RT analysis does not suggest that participants' preference for clothes in the color condition is cognitively costly. Instead, they were generally faster to click on clothes than on foods see Panel C in Figure 3).

Fixation time

Three time windows were pre-registered for the analysis of looking data: an early window corresponding to the critical word in the instructions (i.e. from the onset to the offset of the color adjective or the numeral, depending on the condition; mean color adjective duration: 572 ms; mean number word duration: 489 ms), an intermediate window ranging from the offset of the critical word until

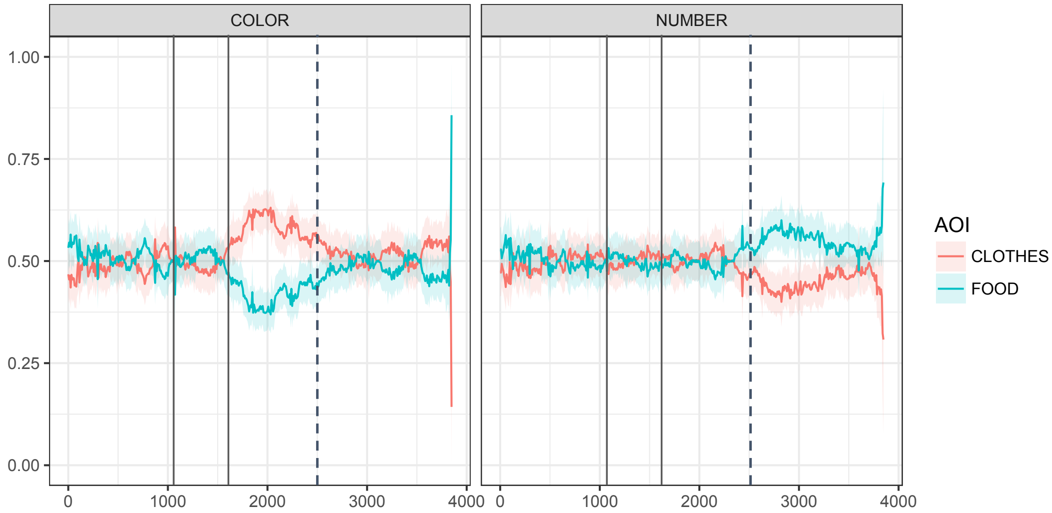


Figure 4: Experiment 2 proportion looks to clothing and food areas of interest in the Color condition (left) and Number condition (right). Zero on the timeline corresponds with the onset of the instruction. Solid vertical lines show the onset and average offset of the color/number words (uncorrected). Dashed vertical lines show the average click time.

the participant's response, and a late window ranging from the participant's response until the end of the trial. Following the standard convention in visual world studies, time windows were adjusted by +200ms to account for the time it takes to launch a saccade. Proportion of fixation time were extracted for the two AOIs (clothes vs fruits) for each trial and participant in the three time-windows (early, intermediate and late). Separate analyses were performed for each time window, as pre-registered. For visualizations of the proportion of fixation time during real-time processing, see Figure 4.

Early time window

Following previous analyses, proportions of fixation time on the Clothes AOI were analyzed using a linear mixed-effects model with a fixed effect of Condition (Color vs Number) and the maximal random effect structure for Participants and Items, including random intercepts for Participants and Items and random by-participant and by-item slopes for Condition. The model did not reveal significant results in the early time window (β=1.496, SE=2.369, df=36.271, t=0.631, p=0.532), suggesting that when comprehenders processed color adjectives, they did not have an early food bias driven by world knowledge, which was then overridden by an informativity bias. Instead, in the earliest stages of processing, participants did not show a preference for either clothes or foods, failing to reveal either type of bias.

Intermediate time window

The same model was constructed for the fixation-time data for clothes during the second time window (because of model convergence issues, the random intercept and random slope effects for Participant were uncorrelated). The model revealed a significant effect of Condition (β=-8.907, SE=2.360, df=31.954, t=-3.774, p<0.0007), with participants fixating longer on the clothes in the Color condition than in the Number condition. The results of the intermediate time window therefore confirm that starting from the offset of the color adjective, participants showed a preference for clothes driven by informativity expectations. There was therefore no evidence of an early stage in color word processing that is driven by world knowledge about color typicality.

Late time window

Analyses of the fixation data for clothes in the last window revealed a significant effect of Condition (β= -6.949, SE=2.157, df=22.811, t=-3.222, p<0.0039), resulting from participants fixating more on the foods than the clothes once they had selected a picture. The Number condition was included as a baseline against which to compare behavior in the Color condition, for which the Informativity and World Knowledge Hypotheses make competing predictions. However, we speculate that participants' behavior in the Number condition —i.e., their late preference for foods over clothes — could reflect the flip side of those same predictions. If speakers tend to omit color words when talking about food, there may be other words (including numbers) that appear with food more frequently, a point we revisit in the corpus study below.

7. Corpus analysis of co-occurrence statistics

The pattern we observe in Experiments 1 and 2 whereby color words bias participants towards anticipating a clothing word is predicted under the Informativity Hypothesis and is based on the assumption that comprehenders are aware that speakers produce redundant color adjectives at higher rates for objects that lack a protoypical color than for those with a protoypical color, and they do so in category-specific and context-specific ways such that color modification is particularly high for clothing even in contexts that do not require contrastive inferences. However, it is also possible that a clothing bias in the color condition could emerge if color+clothing collocations are simply more frequent in comprehenders' linguistic input overall and if comprehenders track those context-independent frequencies. In other words, rather than tracking speakers' descriptions of particular object categories (e.g., clothes vs foods) in the kinds of contexts we use here (i.e., contexts that lack a competitor object for which color might be discriminating), comprehenders could simply track the frequency of the phrases ‘the yellow shirt’ versus ‘the yellow banana’ in a context-independent way.

To rule out this alternative explanation for our findings, we extracted corpus counts from the Google ngram corpus (Brants et al., 2006), whose contexts of use are unknown but whose ngram counts nonetheless provide an estimate of word collocation frequencies in comprehenders’ general input. We use the Google ngram corpus because it provides counts of approximately 1 trillion words from publicly-accessible web pages containing English text. The ngrams we target correspond to the Experiment 1 linguistic materials (see Supplemental Information). We of course have no information about the contexts in which the corpus instances were produced (whether the speaker was mentioning color redundantly or for disambiguation) or about other formulations which speakers might use to mention color beyond prenominal modification. This context-free analysis, however, is exactly what we need in order to test whether our participants’ responses in Experiments 1 and 2 could have arisen simply from their tracking context-free co-occurrence statistics.

We use two corpus measures: First we report how often clothing versus food words appear in the trigrams that comprise our linguistic materials and, second, we compute Point-wise Mutual Information (PMI), an information theoretic measure that takes into account the relative frequencies of the words involved. The first metric uses raw frequencies; we compare the trigram counts for clothing and food words in each of our experimental items in the color condition (‘the yellow shirts’ versus ‘the yellow bananas’) and the number condition (‘the two shirts’ versus ‘the two bananas’). The second metric incorporates the relative frequency of the words themselves (to account for the fact that clothing words are simply more frequent overall than food words in the Google ngram corpus); we compute Point-wise Mutual Information (PMI) to test how close the association is between color and food words and between color and clothing words, and likewise between number and food and between number and clothes (for a related analysis, see Culbertson, Schouwstra and Kirby, 2020). To foreshadow our findings, the corpus results rule out the alternative explanation of our Experiment 1 and Experiment 2 findings: Neither the trigram analysis nor the PMI analysis showed evidence that the pattern predicted under the Informativity Hypothesis could emerge from context-independent frequencies in comprehenders’ linguistic input for the expressions used in our experimental items. The corpus results thereby lend support to our claim that comprehenders in our task are showing pragmatic awareness of context-sensitive and category-sensitive production likelihoods of color words, which have been documented in prior work on speakers’ redundant color modification (Sedivy, 2003; Rubio-Fernandez, 2016; Westerbeek et al., 2015).

Trigram	Experimental item usage	Clothing	Food
'the [color] [object]s'	Pairing 1 colors	69.8	355.5
'the [color] [object]s'	Pairing 2 colors	176.8	355.5
'the two [object]s'	Pairing 1 & Pairing 2	425.7	94.4

Table 1
Mean frequencies for Experiment 1 linguistic materials, extracted from Google trigram counts where [color] is one of 10 colors and [object] is one of the 20 clothing and food objects in Table A.1. The first two rows show the mean frequencies for the color condition. For clothing, the color∼object pairs change across pairings (e.g., Pairing 1 skirts were yellow in order to be paired with bananas; Pairing 2 skirts were white in order to be paired with marshmallows), whereas for food, the color does not change across pairings (e.g., bananas were always yellow). The last row shows the number condition, which does not vary across pairings since the form was always 'the two [object]s.'

7.1. Trigram counts

Table 1 shows the mean counts for the trigram sequences in the Experiment 1 materials ('the two shirts', 'the two bananas', 'the yellow shirts', 'the yellow bananas', etc.). All counts represent the sum over lowercase and uppercase variants and were computed as by-object means for the trigrams 'the two [object]s' and 'the [color] [object]s' for all objects in the Experiment 1 materials with the specific color object pairings that we used (e.g., 'the yellow shirts' and 'the yellow sandals' but not 'the orange shirts').

For the baseline condition 'the two [object]s', trigrams mentioning number and clothing are more frequent (425.7) than trigrams mentioning number and food (94.4), whereas for the critical color condition 'the [color] [object]s', this pattern is reversed such that the color words are followed more often by food words (355.5) than clothing words (69.8 for Pairing 1; 176.8 for Pairing 2). If comprehenders track context-free frequencies like those in Table 1 and use those generalized frequencies to anticipate upcoming words in the specific contexts of the Experiment 1 and Experiment 2 items, we would have expected to see behavior in keeping with the World Knowledge Hypothesis because our specific color words are followed more often by food words ('the yellow bananas') than clothing words ('the yellow shirts').

Another way of thinking about the comparison of trigram frequencies across clothing and food is to think of the situation (as in our materials) where a comprehender has encountered the beginning of a referring expression (e.g., 'the yellow') and is trying to guess what will follow – i.e., what is the conditional probability of 'bananas' given 'the yellow' and how does that differ from the conditional probability of 'shirts' given 'the yellow'. Those conditional probabilities would depend directly on the raw trigram frequencies represented in Table 1, and crucially their relative size in Table 1 directly predicts the relative size of the conditional probabilities. For example, to compare the probability of encountering ‘bananas’ or ‘shirts’ after ‘the yellow’, one would compute *p*(‘bananas’ | ‘the yellow’) and *p*(‘shirts’ | ‘the yellow’). These conditional probabilities would be computed as in (i) and (ii): Of all the times an expression started ‘the yellow’, how often did the expression end with ‘shirts’ versus ‘bananas’? Given that the denominator is the same across (i-ii) and the only difference is the trigram count in the numerator, one can see the trigram counts in Table 1 as indicative of the difference in predictability of the clothing word versus the food word in an expression ‘the yellow...’.

$$\text{(i)}\quad \text{p(‘bananas’ | ‘the yellow’)} = \frac{\text{count(‘the yellow bananas’)}}{\text{count(‘the yellow’)}}$$

$$\text{(ii)}\quad \text{p(‘shirts’ | ‘the yellow’)} = \frac{\text{count(‘the yellow shirts’)}}{\text{count(‘the yellow’)}}$$

However, one should be cautious in interpreting the values in Table 1 because these raw frequencies are not adjusted for the relative frequency of individual words in the Google ngram corpus. For example, clothing words are simply more frequent than food words in the corpus overall. Because of this, the chance of the words ‘the yellow’ being followed by a clothing word might be higher than the chance of those words being followed by a food word simply because clothing words appear more frequently in general.

7.2. Point-wise mutual information

To incorporate the relative frequency of the words in our critical expressions, we also compute PMI scores which measure the association between a pair of words: e.g., between the two words in ‘yellow shirts’ or between those in ‘yellow bananas’. PMI scores allow us to estimate whether the word ‘yellow’ appears more frequently with ‘shirts’ or with ‘bananas’, taking into account the frequency of the individual words. PMI is defined as the log ratio of the probability of a pair of words appearing together and the independent probabilities of each of those words appearing at all (Church and Hanks, 1990). The PMI calculation is shown in (iii), illustrated with the words ‘yellow shirts’ with the numerator containing the probability of the bigram ‘yellow shirts’ and the denominator containing the probabilities of each of those words alone. We estimate PMI as in (iv) where the probability of

Bigram	Experimental item usage	Clothing	Food
'[color] [object]s'	Pairing 1 colors	1.14	3.15
'[color] [object]s'	Pairing 2 colors	1.12	3.15
'two <objects>'	Pairing 1 & Pairing 2	0.51	1.26

Table 2
Point-wise Mutual Information for Experiment 1 linguistic materials, using Google bigram and unigram frequencies. As in Table 1, the first two rows show the color~object pairs for Pairings 1 and 2, with clothing color varying across pairings while food color stays constant. The last row shows the number condition, which was the same across pairings.

a word is calculated as the count of that word, divided by the total number of words in the corpus (N is the size of the Google ngram corpus, approximately one trillion); the probability of a bigram is calculated similarly (the total number of bigrams is also close to N for a large enough corpus).

$$\text{(iii)} \quad \text{PMI of the bigram 'yellow shirts'} = log \frac{p(\text{'yellow shirts'})}{p(\text{'yellow'})p(\text{'shirts'})}$$

$$\text{(iv)} \quad \text{Estimated PMI of the bigram 'yellow shirts'} = log \frac{\frac{\text{count('yellow shirts')}}{\text{N}}}{\frac{\text{count('yellow')}}{\text{N}} * \frac{\text{count('shirts')}}{\text{N}}}$$

Table 2 reports the PMI scores for the specific bigrams in Experiment 1. Those PMI scores show that the strength of association is higher between food words and their prototypical colors than between clothing words and the colors in which clothing was depicted in our materials. This presumably reflects the tight conceptual link between foods and their prototypical colors and the lack of a tight link between clothing and a specific color. The word 'two' in our materials is also more closely associated with food words than with clothing words, perhaps reflecting a property of foods (that they are counted and/or described with a low number) that differs from clothing (which are not necessarily quantified in pairs). We speculate that this number~food association could explain the late-emerging food bias in the Number condition of Experiment 2 (see Fig. 4).

If comprehenders track the PMI scores of the Pairing 1 and Pairing 2 colors and use them to anticipate upcoming words, we would expect to see behavior in keeping with the World Knowledge Hypothesis because the PMI scores indicate that color words are more closely associated with food than clothing. Note that the PMI values for bigrams that contain colors words are higher than for bigrams that contain ‘two’, a finding that is in keeping with prior work that likewise reports a tighter association between objects and their properties (like color) than between objects and numbers (Culbertson et al., 2020).

In sum, this analysis reports corpus frequencies in order to establish whether the behavior supporting the Informativity Hypothesis could arise via simpler word associations in comprehenders’ general input. For our experimental materials, neither the trigram nor PMI measures correspond to the observed pattern whereby color words create a stronger clothing bias than number words.

8. General Discussion

Across two behavioral experiments, we tested comprehenders’ expectations about upcoming words in contexts that pitted world knowledge against informativity expectations. Comprehenders consistently used a prenominal color adjective to anticipate the mention of clothing (‘the yellow...shirts’), which lacks prototypical color. This color∼clothing bias emerged even though the color word described the prototypical color of a depicted food. In offline picture selections and online eye movements, comprehenders showed no evidence of a color∼food bias, contra models in which next-mention expectations are primarily driven by world knowledge (‘the yellow... bananas’). Rather, the findings support our Informativity Hypothesis: Comprehenders use knowledge not only about the probability of different situations in the world, but also their pragmatic knowledge about speakers’ use of language to describe the world. A corpus analysis confirms that these findings are unlikely to have arisen from general word co-occurrence statistics abstracted from any specific context, lending further support to our conclusion that comprehenders are able to incorporate speaker production likelihoods in context-specific and category-specific ways.

Recent psycholinguistic studies have shown that comprehenders keep track of different sources of distributional information in language (e.g., words, morphemes, phonemes) in speakers’ productions, which is independent of their world knowledge. Arnon and Snider (2010) show that comprehenders are sensitive to the frequencies of compositional four-word phrases (e.g. ‘Don’t have to worry’ vs ‘Don’t have to wait’), with faster processing for phrases that speakers produce with higher frequency. The authors conclude that language users not only keep track of word-level usage but also store phrase-frequency information spanning various levels of abstraction. More recently, Morgan and Levy (2016) investigated multi-word phrases whose semantic content is explicitly held constant. They show that online processing of highly frequent binomials (e.g., ‘bread and butter’) is primarily driven by direct experience (as estimated from corpus frequency counts), whereas online processing of novel binomial expressions (e.g., 'bishops and seamstresses' vs 'seamstresses and bishops') is influenced by abstract knowledge of the ordering constraints in speakers' productions (as estimated by a probabilistic model). The authors interpret their results as supporting models of language processing including both compositional generation and direct reuse of multi-word expressions.

Our study contributes to the above literature by showing comprehenders' sensitivity to speaker production constraints: Alongside prior work showing that comprehenders can deploy their world knowledge of stereotypical colors in referential communication (e.g., looking at a lemon when hearing 'banana'; Huettig and Altmann, 2011; see also Naor-Raz et al., 2003), our results show that comprehenders are sensitive to informativity constraints in the use of color adjectives in non-contrastive contexts. Comprehenders' behavior suggests an awareness of speakers' production preferences, namely their tendency to use color words redundantly for referents without prototypical colors and for which color variability is a key feature, such as clothes. However, it must be noted that what counts as informative is context dependent, rather than categorical. For example, other things being equal, 'The mechanic checked... the brakes' is more predictable than 'the spelling' (Bicknell et al., 2010), but less predictable than yellow is for 'banana' Huettig and Altmann, 2011. Yet in a conversation about the benefits of ripe fruit, 'yellow bananas' may be optimally informative, rather than redundant.

The results of our eye-tracking experiment confirmed that informativity expectations constrain the interpretation of color adjectives relatively early in processing, rather than resulting from late top-down processes inhibiting world knowledge associations (e.g., between the color yellow and bananas). By contrast, when processing a numerical description ('Click on the two...'), participants' eye movements revealed hesitation between the clothes and foods in the pictures, in line with their chance performance when selecting a sentence continuation. If the color condition had induced hesitation, we expect we would have seen a pattern that resembles that in the baseline numerical condition.

These results contribute to the eye-tracking literature on adjective processing, which has shown that comprehenders are sensitive to the frequency with which different types of adjectives are used contrastively or redundantly, deriving contrastive inferences when interpreting scalar and material adjectives, but not always with color adjectives (Sedivy, 2003, 2004; Sedivy et al., 1999; Aparicio, Xiang and Kennedy, 2016; Rubio-Fernandez et al., 2021). A seemingly discrepant finding is the delayed preference for clothes over foods in the color condition, which did not emerge until the intermediate time window (i.e. after the offset of the color adjective). However, here it is important to bear in mind that the above visual-world studies used polychrome displays where color was an efficient cue for visual search. As shown in another recent eye-tracking study (Rubio-Fernandez, in press), color contrast facilitates visual search by color relative to similar displays with fewer colors. Since the critical displays in our study were monochrome, color processing may have resulted in protracted referent identification relative to earlier eye-tracking studies with polychrome displays.

Our results further show that comprehenders are sensitive to the kinds of lexical categories that tend to be overspecified by color (at least when the choice is between clothes and foods). Future studies should investigate how fine-grained comprehenders' sensitivity to color overspecification may be depending on the lexical categories contrasted (e.g., when selecting between a white car and a white fridge). In addition, future eye-tracking studies should investigate these questions using anticipatory looking during natural sentence processing, rather than sentence completion. It must be noted, however, that color adjectives are often monosyllabic in English (e.g., blue, red, black, pink), making it difficult to reliably tap anticipatory looking ahead of the noun (Rubio-Fernandez, Terrasa, Shukla and Jara-Ettinger, 2019).

Here we focus on color because its non-contrastive use is well-documented (Pechmann, 1989) and its inclusion/omission is strongly influenced by object category (Sedivy, 2003; Rubio-Fernandez, 2016; Westerbeek et al., 2015; Kreiss and Degen, 2020). However, the results should extend to descriptors with weaker color biases or those outside the visual domain. For example, both vehicles and appliances lack inherent color, but speakers mention color more for cars than for appliances. Likewise, some objects have inherent properties (<u>old</u> ruins, <u>modern</u> smartphones), and speakers may thus prefer to use those descriptors informatively for other objects (old/new cars, modern/antique furniture). If speakers reliably show such production biases and if comprehenders can effectively track them, new questions emerge regarding the domains where informativity expectations might arise.

9. Summary and conclusions

Ample empirical evidence shows that language comprehension is predictive in nature. For example, people predict that 'the mechanic checked...' is more like to be followed by 'the brakes' than by 'the spelling', whereas they make the reverse prediction if the agent is 'the journalist'. Here we use Bayes' Rule to formalize a related hypothesis: When language comprehenders anticipate a noun following a redundant color adjective (e.g., 'Can you pass me the red…'), they rely not only on their world knowledge about colors (e.g., which objects are typically red), but also on their experience of speakers' use of color adjectives to refer to different objects (e.g., the frequency of ‘the red strawberries’ vs ‘the red shoes’) and they do this in a context-specific and category-specific way. Our findings go beyond previous work on people’s ability to make linguistic predictions based on their world knowledge, and show how comprehenders can reverse engineer the way speakers tend to talk about the world.

10. Author Contributions

Both authors contributed equally to the study design, data collection, analysis, and manuscript drafting and revision.

11. Acknowledgements

This research was supported by a FRIPRO Grant from the Research Council of Norway (Ref. 275505) awarded to PRF and a Leverhulme Trust Prize in Languages and Literatures to HR (Ref. PLP-2016-128). The authors gratefully acknowledge this funding. Thanks also to Francis Mollica for help with data visualizations.

References

Aparicio, H., Xiang, M., Kennedy, C., 2016. Processing gradable adjectives in context: A visual world study. Semantics and Linguistic Theory 25, 413–432.

Arnon, I., Snider, N., 2010. More than words: Frequency effects for multi-word phrases. Journal of Memory and Language 62, 67–82.

Arts, A., Maes, A., Noordman, L., Jansen, C., 2011. Overspecification facilitates object identification. Journal of Pragmatics 43, 361–374.

Aylett, M., Turk, A., 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. Language and Speech 47, 31–56.

Bates, D., Kliegl, R., Vasishth, S., Baayen, H., 2015. Parsimonious mixed models. arXiv preprint arXiv:1506.04967 .

Bicknell, K., Elman, J.L., Hare, M., McRae, K., Kutas, M., 2010. Effects of event knowledge in processing verbal arguments. Journal of Memory and Language 63, 489–505.

Brants, T., Franz, A., et al., 2006. Web 1T 5-gram version 1. Philadelphia: Linguistic Data Consortium .

Brown, P.M., Dell, G.S., 1987. Adapting production to comprehension: The explicit mention of instruments. Cognitive Psychology 19, 441–472.

Church, K.W., Hanks, P., 1990. Word association norms, mutual information, and lexicography. Computational Linguistics 16, 22–29.

Culbertson, J., Schouwstra, M., Kirby, S., 2020. From the world to word order: deriving biases in noun phrase order from statistical properties of the world. Language 96, 696–717.

DeLong, K.A., Urbach, T.P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nature Neuroscience 8, 1117–1121.

Franke, M., Jäger, G., 2016. Probabilistic pragmatics, or why bayes' rule is probably important for pragmatics. Zeitschrift für Sprachwissenschaft 35, 3–44.

Goodman, N.D., Frank, M.C., 2016. Pragmatic language interpretation as probabilistic inference. Trends in Cognitive Sciences 20, 818–829.

Grice, H.P., 1975. Logic and conversation, in: Speech Acts. Brill, pp. 41–58.

Huettig, F., Altmann, G.T., 2011. Looking at anything that is green when hearing "frog": How object surface colour and stored object colour knowledge influence language-mediated overt attention. Quarterly Journal of Experimental Psychology 64, 122–145.

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.

Kehler, A., Rohde, H., 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. Theoretical Linguistics 39, 1–37.

Kreiss, E., Degen, J., 2020. Production expectations modulate contrastive inference, in: Proceedings of the 42nd Annual Meeting of the Cognitive Science Society, pp. 259–265.

Kuperberg, G.R., Jaeger, T.F., 2016. What do we mean by prediction in language comprehension? Language, Cognition and Neuroscience 31, 32–59.

Levy, R., 2008. Expectation-based syntactic comprehension. Cognition 106, 1126–1177.

Levy, R., Jaeger, T.F., 2007. Speakers optimize information density through syntactic reduction. Advances in Neural Information Processing Systems 19, 849.

Lockridge, C.B., Brennan, S.E., 2002. Addressees' needs influence speakers' early syntactic choices. Psychonomic bulletin & review 9, 550–557.

Long, M., Rohde, H., Rubio-Fernandez, P., 2020. The pressure to communicate efficiently continues to shape language use later in life. Scientific Reports 10, 1–13.

McNamara, T.P., 2005. Semantic priming: Perspectives from memory and word recognition. Psychology Press.

Morgan, E., Levy, R., 2016. Abstract knowledge versus direct experience in processing of binomial expressions. Cognition 157, 384–402.

Naor-Raz, G., Tarr, M.J., Kersten, D., 2003. Is color an intrinsic property of object representation? Perception 32, 667–680.

Paraboni, I., Van Deemter, K., Masthoff, J., 2007. Generating referring expressions: Making referents easy to identify. Computational Linguistics 33, 229–254.

Pechmann, T., 1989. Incremental speech production and referential overspecification. Linguistics 27, 89–110.

R Core Team, 2016. R: A language and environment for statistical computing.

Rohde, H., Futrell, R., Lucas, C.G., 2021. What's new? a comprehension bias in favor of informativity. Cognition 209, 104491.

Rohde, H., Horton, W.S., 2014. Anticipatory looks reveal expectations about discourse relations. Cognition 133, 667–691.

Rohde, H., Kurumada, C., 2018. Alternatives and inferences in the communication of meaning, in: Psychology of Learning and Motivation. Elsevier. volume 68, pp. 215–261.

Rubio-Fernandez, P., 2016. How redundant are redundant color adjectives? an efficiency-based analysis of color overspecification. Frontiers in Psychology 7, 153.

Rubio-Fernandez, P., 2019. Overinformative speakers are cooperative: Revisiting the gricean maxim of quantity. Cognitive Science 43, e12797.

Rubio-Fernandez, P., in press. Color discriminability makes over-specification efficient: Theoretical analysis and empirical evidence. Humanities and Social Sciences Communications .

Rubio-Fernandez, P., Mollica, F., Jara-Ettinger, J., 2021. Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. Journal of Experimental Psychology: General 150, 583–594.

Rubio-Fernandez, P., Terrasa, H.A., Shukla, V., Jara-Ettinger, J., 2019. Contrastive inferences are sensitive to informativity expectations, adjective semantics and visual salience URL: psyarxiv.com/mr4ah, doi:10.31234/osf.io/mr4ah.

Sedivy, J.C., 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. Journal of Psycholinguistic Research 32, 3–23.

Sedivy, J.C., 2004. Evaluating explanations for referential context effects: Evidence for gricean mechanisms in online language interpretation, in: Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions. MIT Press, pp. 345–364.

Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., Carlson, G.N., 1999. Achieving incremental semantic interpretation through contextual representation. Cognition 71, 109–147.

Sperber, D., Wilson, D., 1986. Relevance: Communication and cognition. volume 142. Harvard University Press Cambridge, MA.

Tourtouri, E.N., Delogu, F., Sikos, L., Crocker, M.W., 2019. Rational over-specification in visually-situated comprehension and production. Journal of Cultural Cognitive Science 3, 175–202.

Van Berkum, J.J., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from erps and reading times. Journal of Experimental Psychology: Learning, Memory, and Cognition 31, 443.

Warren, T., Dickey, M.W., 2021. The use of linguistic and world knowledge in language processing. Language and Linguistics Compass 15, e12411.

Westerbeek, H., Koolen, R., Maes, A., 2015. Stored object knowledge and the production of referring expressions: The case of color typicality. Frontiers in Psychology 6, 935.

Instruction	Pairing 1	Pairing 2
‘Click on the black/two...’	beans, aprons	beans, slippers
‘Click on the black/two...’	grapes, sweatshirts	beans, swimsuits
‘Click on the blue/two...’	plums, ties	plums, aprons
‘Click on the blue/two...’	bluecorn, swimsuits	corn, sweatshirts
‘Click on the brown/two...’	coconuts, gloves	coconuts, dresses
‘Click on the brown/two...’	onions, slippers	onions, ties
‘Click on the green/two...’	cucumbers, dresses	cucumbers, gloves
‘Click on the green/two...’	zucchini, boots	zucchini, hats
‘Click on the orange/two...’	carrots, hats	carrots, boots
‘Click on the orange/two...’	pumpkins, coats	pumpkins, blouses
‘Click on the pink/two...’	guavas, blouses	guavas, jackets
‘Click on the pink/two...’	sausages, belts	sausages, coats
‘Click on the purple/two...’	eggplants, jackets	eggplants, belts
‘Click on the purple/two...’	figs, high heels	figs, scarves
‘Click on the red/two...’	cherries, scarves	cherries, high heels
‘Click on the red/two...’	tomatoes, sweaters	tomatoes, bras
‘Click on the white/two...’	garlic, bras	garlic, shirts
‘Click on the white/two...’	marshmallows, skirts	marshmallows, sandals
‘Click on the yellow/two...’	lemons, sandals	lemons, sweaters
‘Click on the yellow/two...’	bananas, shirts	bananas, skirts

Table A.1
Materials for Experiment 2 (Pairing 1 and Pairing 2) and Experiment 3 (Pairing 1).

Supplementary Information

Table A.1 shows the food and clothing items that were paired in Experiments 1 and 2.