Title

On the role of object knowledge in reference production: effects of color typicality on content determination

Permalink

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN

1069-7977

Authors

Westerbeek, Hans
Koolen, Rudd
Maes, Alfons

Publication Date

2014

On the role of object knowledge in reference production: Effects of color typicality on content determination

Hans Westerbeek (h.g.w.westerbeek@tilburguniversity.edu)
Ruud Koolen (r.m.f.koolen@tilburguniversity.edu)
Alfons Maes (maes@tilburguniversity.edu)
Tilburg center for Cognition and Communication (TiCC), Tilburg University, The Netherlands

Abstract

In two language production experiments, we investigated whether stored knowledge of the typical color of objects affects spoken reference. In experiment 1, human speakers referred to objects with colors ranging from very typical (e.g., red tomato) to very atypical (e.g., blue pepper). The probability that speakers redundantly include color in their descriptions was almost linearly predicted by the degree of atypicality. In experiment 2, we extended this finding to references to objects for which color is inherently a less salient property in stored knowledge (i.e., objects with a highly characteristic shape, making color less important for recognition). Following these findings that typicality affects reference production, we conclude that speakers utilize stored knowledge about everyday objects they refer to. We discuss the implications of our findings for artificial agents that generate natural language, arguing that computational models fall short in capturing general knowledge about typical properties of objects.

Keywords: reference production; color typicality; content determination; visual saliency; AI models of reference production

Introduction

Reference production is the linguistic process of generating definite descriptions of objects, such as "the orange crocodile". The goal for human speakers is to refer to an object in such a way that an addressee can uniquely identify the target among distractor objects. Studying human reference production is essential for building artificial models (Van Deemter, Gatt, Van Gompel, & Krahmer, 2012b), as human-like reference production is an important predictor of naturalness in interaction between humans and artificial agents.

Central to reference production is *content determination*: the question which properties of the target object a speaker includes when referring to an object for the first time in conversation (e.g., Dale and Reiter, 1995; Van Deemter, Gatt, Van der Sluis, & Power, 2012a). One strategy is to only include properties that are necessary to rule out all distractor objects. In that sense, the expression "the orange crocodile" for the crocodile in Figure 1 contains a redundant color attribute (given that mentioning the type "crocodile" rules out all disctractors). However, human speakers often mention properties of objects that are not strictly needed for unique identification (e.g., Koolen, Gatt, Goudbeek, & Krahmer 2011; Pechmann, 1989) − especially color (e.g,. Viethen, Goudbeek, & Krahmer, 2012). Visual saliency is one reason for such redundancy: speakers base their selection of object properties on what they perceive as salient (e.g., Clarke, Elsner, & Rohde, 2013; Koolen, Goudbeek, & Krahmer, 2013). Properties can be regarded as salient for various reasons. For example, considering the example in Figure 1, it is reasonable to assume that the crocodile's color is salient (and therefore mentioned), because orange is an atypical color for crocodiles.

Figure 1: A crocodile, a frog, and a goldfish.

Visual saliency is generally characterized as a two-component process (Itti & Koch, 2000): a speaker's visual attention is guided by bottom-up and top-down factors. Bottom-up factors are perceptual, image-based cues that make areas in a visual scene 'pop out' pre-attentively, such as bright colors or strong contrasts. Top-down factors on the other hand are conceptual cues, guided by cognitive processing of the scene by the speaker. One top-down cue that seems to be largely ignored in models of reference production is the speaker's general knowledge about the type of object that is being referred to. For example, as noted above, the orange crocodile in Figure 1 has a color that is incongruous to general knowledge about crocodiles.

We argue that this knowledge should not be ignored in models and theories of content planning. When a speaker refers to an object, the process of content determination is essentially preceded by object recognition. In object recognition, a representation of the object in stored knowledge is accessed (e.g., Humphreys, Riddoch, & Quinlan, 1988). For objects that have one or more typical colors associated with them (i.e., color-diagnostic objects, for example tomatoes which are typically red), this knowledge contains color information (e.g., Tanaka, Weiskopf, & Williams, 2001). This is supported by experiments wherein people name visually presented objects. Recognition is slower when (color-diagnostic) objects are presented in atypical colors than when they are typically colored (e.g., Naor-Raz, Tarr, & Kersten, 2003; Tanaka et al., 2001; Therriault, Yaxley, & Zwaan, 2009). Furthermore, the contribution of color in object recognition is stronger for objects with a simple and uncharacteristic shape (e.g., oranges) than for objects with more complex shapes (e.g., fire trucks; Price & Humphreys, 1989). Uncharacteristically shaped objects are in particular natural objects such as fruits, with a simple shape (e.g., round, few protrusions) which cover most of the category members. For such objects, color is arguably more important for their recognition because it is more important in distinguishing category members.

The idea that stored representations of objects (which are accessed in object recognition) play a role in reference production gains support from a language production experiment by Sedivy (2003). In her experiment, participants referred to normally colored objects. These were either color-diagnostic objects, or objects that can have any color (e.g., cups). Speakers mentioned color significantly less often when referring to color-diagnostic objects. Sedivy (2003) attributes this to the fact that the colors of the color-diagnostic objects are more predictable than those of the any-color objects. This advocates that speakers decide on including a property (color) based on their stored knowledge about the type of the object they refer to. But Sedivy studied reference to normally colored objects, and the question remains whether properties that are rendered visually salient because they *deviate* from object knowledge are more likely to be encoded in the content determination process.

Irrespective of reference production, objects that have a color different from stored object knowledge are known to attract visual attention. Becker, Pashler, and Lubin (2007) eye-tracked participants who were presented with naturalistic scenes containing an object in an atypical color (a green hand), or in a typical color (a flesh-colored hand). Participants fixated earlier, more often, and longer on the green hand than on the normal hand. This result could not be ascribed to green being more salient than flesh-color, which Becker et al. controlled for by swapping the hand's color with a mug (which is equally typical in green or flesh-color). So, if the visual saliency of atypically colored objects is steered by a top-down process involving general object knowledge, it is likely that speakers mention their color when referring to such objects.

However, current studies in reference production have not yet focused on this proposed influence of the degree of atypicality on content determination. One study, by Mitchell, Reiter, and Van Deemter (2013), does investigate effects of atypicality, by showing that speakers prefer to mention the shape or material of objects when it is atypical (e.g., "octagonal mug", "wooden key"). However, Mitchell et al.'s results did not reveal how the degree of typicality of a certain shape or material affects content planning − does it matter just how atypical a property is for an object?

The current experiments

Based on the literature reviewed above, we expect content determination to be affected by the typicality of the objects that speakers refers to. Reference production incorporates object recognition, which addresses stored object representations. For color-diagnostic objects, these representations contain objects' typical colors. As atypically colored objects attract visual attention, this top-down saliency of the color of these objects may make it more likely that speakers include color in their referring expressions.

We test this expectation in two language production experiments. In these experiments, speakers produce spoken descriptions of typically and atypically colored objects that are embedded in simple visual scenes. They are instructed to do this in such a way that an addressee can identify the object among other (distractor) objects. In experiment 1, we manipulate the color of the described object in oder to investigate whether the degree of atypicality of the color (established by means of a pretest) affects the probability that speakers include it in their descriptions. In experiment 2, we extend these findings to objects for which color itself is a less salient property, by eliciting descriptions of objects with a fairly characteristic shape.

Experiment 1

Method

Participants Forty-two undergraduates (eleven men, mean age 22 years) participated for course credit. All were native speakers of Dutch (the language of the study). None were informed about the conditions in the experiment.

Materials pretest To determine the degree of typicality of objects in certain colors, we conducted a pretest. Sixteen color-diagnostic objects (mainly fruits and vegetables) were selected on the basis of stimuli used in object recognition studies (e.g., Therriault et al., 2009). For each object a high quality photo was obtained, and edited such that the object was seen on a plain white background. Further photo editing was done to make a red, blue, yellow, green, and orange version of each object. This resulted in a set of eighty different photos (sixteen object types × five colors).

This set was presented to forty participants in an on-line judgment task (thirteen men, mean age 26 years, none participated in any of the other experiments in this paper). To manage the length of this task, participants were randomly assigned to one of two halves of the photo set. Participants had to name the object and its color, and used a slider control to answer the question "how typical is this color for this object?", for each photo individually. The position of the slider was linearly converted to a typicality score ranging from 0 to 100, where 100 indicated that the color-object combination was judged as most typical. We also assessed whether the objects and colors were named correctly.

Materials Fourteen objects were selected to be used in the experiment (apple, banana, carrot, cheese, corn, grapes, lemon, lettuce, orange, pear, pepper, pineapple, pumpkin, tomato). This selection was based on their naming and typicality score in the pretest. Each object appeared in three of the five aforementioned colors in the final experimental stimulus set. The main selection criterion was that the final set consisted of objects and colors that together represented the whole spectrum of typicality ratings obtained in the pretest (scores 2−98, from very atypical to very typical, plus scores in between). As an illustration: the least typical objects were a blue pepper and red lettuce, among the most typical ones were yellow cheese and a red tomato, and a yellow apple and a green tomato fell somewhere in between the extremes.

The experimental materials consisted of forty-two scenes. Each scene contained six objects, positioned randomly in a three by two grid, in three different colors (each on two objects such that the target's color was never unique within a scene). We equalized typicality in the scenes by selecting two somewhat typical objects, two atypical ones, and two falling in-between typical and atypical, such that the mean

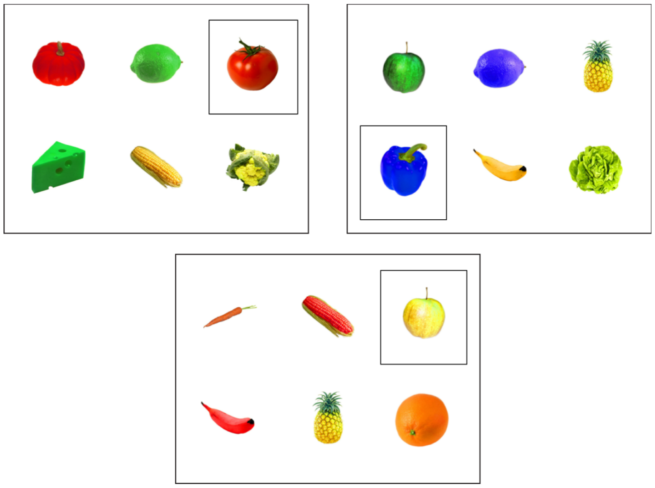

Figure 2: Examples of visual stimuli in experiment 1. From left to right: atypical, in-between, typical target.

typicality score of all scenes ranged between forty and sixty. One of the objects in each scene was the target object, which was clearly marked with a black square.

Crucially, the forty-two target objects differed in their degree of typicality, as established in the pretest. The target object was always of a unique type within the scene, so mentioning color was never needed to distinguish the target from the five distractors. Figure 2 presents three examples of these scenes, one with a highly typical target, one with an 'in-between' target, and one with an atypical target.

Procedure The experiment was performed at our university, and had an average running time of about twenty-five minutes. Participants sat at a table facing the experimenter, in front of a laptop. The participants were presented with the forty-two trials, one by one. Between each experimental trial, there was a filler scene. These filler scenes consisted of four hard-to-describe *greebles* (Gauthier & Tarr, 1997), all purple, so that participants were not primed with color in the other trials. Participants described the target objects in such a way that the experimenter would be able to uniquely identify them in a paper booklet. The instructions emphasized that it would not make sense to include location information in the descriptions, as the experimenter would see the objects in a different configuration. Participants could take as much time as needed to describe the target, and their descriptions were recorded with a microphone. The experimenter never asked the participants for clarification, so the data presented here are regarded as one-shot references.

There was one practice trial with six non-color-diagnostic objects (chair, marker, backpack, book, desk lamp, mug), and one practice trial with greebles. Once the experimenter identified a target, this was communicated to the participant, and the a button was pressed to advance to the next trial. The trials were presented in a fixed order; this order was reversed for half of the participants.

Results and discussion

In total, 1764 target descriptions were recorded in the experiment. Over 89% of these descriptions (n=1575) were intelligible and contained a correct type attribute, resulting in unique reference. Using the correct type was important, because otherwise we could not deduce whether the object's color was regarded as typical or atypical.

We administered whether color was mentioned in the referring expression, and analyzed the data using logit mixed models (Jaeger, 2008). Initial analyses revealed that stimulus order had no effects, so this was left out in the following analyses. In our model, color typicality (as scores on the pretest) was included as a fixed factor, standardized to reduce collinearity and increase comparability with experiment 2. Participants and target object types were included as random factors. The model had a maximal random effect structure: random intercepts and random slopes were included for all within participant and within item factors, to ensure optimal generalizability (Barr, Levy, Scheepers, & Tily, 2013).

Our analysis revealed a significant effect of color typicality on whether a target description contained a color attribute or not (β=–2.11, SE=0.28, p<.001). Figure 3 plots the typicality score of a target object in the pretest against the proportion of descriptions that mentioned color in the production experiment. Clearly, a higher typicality score in the pretest was associated with fewer speakers using color to refer to a target object in the experiment. An additional analysis by means of bivariate correlation reconfirmed that these two measures were significantly related (Pearson r=–.86, n=42 p<.001).

These results warrant the conclusion that content determination is affected by the degree of typicality of a target object's properties. When a property is more atypical for an object, this draws visual attention, and increases the probability that that property is included in a referring expression.

One might however point out that the objects used in this experiment mostly have simple, uncharacteristic shapes. This arguably makes color a very prominent feature in their recognition (cf. Price & Humphreys, 1989). Therefore color itself is an especially salient property of stored representations of these objects, and violations of typical color may be more conspicuous. So, in experiment 2 we test whether our findings generalize to color-diagnostic objects which have a more complex and characteristic shape than most of the objects used in experiment 1.

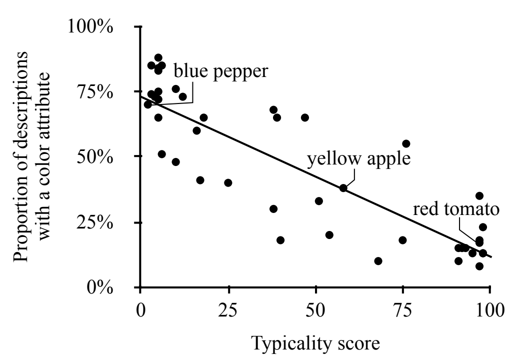

Figure 3: Typicality of colored objects (horizontal axis) and the proportion of descriptions of these objects that contain color (vertical axis) in experiment 1. Some illustrative objects are labeled in this plot; the line represents the correlation between the two variables.

Experiment 2

In this experiment, we test whether the color atypicality effect on content determination is modulated by the complexity of the shape of objects. As color is arguably a more salient feature in the stored representation of simple-shaped objects than it is for complex-shaped objects, we cross color typicality with shape complexity in a language production task similar to the one used in experiment 1.

We also introduce a number of methodological improvements. First, closer inspection of the results of experiment 1 revealed that color was most often mentioned for blue objects, which were all atypical. This may have lead to a perceptual (bottom-up) saliency effect, as blue may be more salient than other colors. To better control for such effects, we equally balanced colors over all conditions. We also equalized conditions on perceptual saliency estimated by a computational model (Erdem & Erdem, 2013). Second, to further assure that the colorful nature of our stimuli in experiment 1 has not boosted the overall probability that color was mentioned (Koolen et al., 2013), we inserted a relatively higher number of non-colorful filler scenes in experiment 2. We also distributed typically and atypically colored objects over two lists so that participants never saw one object in more than one color. Finally, addressees in experiment 2 were naive participants instead of a confederate (the experimenter), to improve ecological validity (cf. Kuhlen & Brennan, 2013).

Method

Participants Sixty-two undergraduates (Dutch; nine men, mean age 22 years) participated for course credit. Thirty-one acted as speakers, the others as addressees. None participated in experiment 1 or in any of the pretests.

Materials Similarly to experiment 1, high quality white-background photos of sixteen target objects were selected and edited, based on stimuli used in object recognition studies. Eight objects had a simple shape; the other objects had a more complex shape. Of each object, a typical and an atypically colored version was created.

The experimental materials consisted of sixteen scenes. Each scene contained six objects in three different colors (each on two objects), with three objects typically colored and the other three atypical. The colors of the objects were either red, green, yellow, or orange, and these colors were rotated across the target objects to create atypically colored versions. Therefore, all four colors were used equally often in both typicality conditions, in order to ensure that potential perceptual saliency effects caused by certain colors were minimized. A computational perceptual saliency estimation

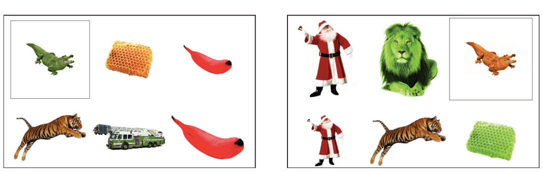

Figure 4: Examples of complex-shaped visual stimuli in experiment 2. Scenes with simple-shaped objects were comparable to the outermost panels in Figure 2.

(Erdem & Erdem, 2013) confirmed that typically and atypically colored objects were perceptually equally salient as the distractors within their scenes.

We manipulated whether the objects in a scene were objects with either simple, uncharacteristic shapes (targets were basketball, lemon, lettuce, orange, strawberry, tennis ball, tomato, watermelon), or with more complex, characteristic shapes (broccoli, carrot, cheese, chick, crocodile, goldfish, lobster, phone booth). We also varied whether the target object was either typically colored or atypically colored.

The same pretest procedure as in experiment 1 was used to obtain typicality scores of the target objects. The mean typicality, based on sixteen participants in this pretest (seven men, mean age 21 years, none participated in any of the other experiments and pretests) was 95/100 for typically colored objects and 4/100 for typical ones. The complexity of the objects did not interact with the typicality ratings of the pretest, so the difference between typical and atypical objects was not modulated by shape complexity, nor was there a main effect of complexity. Color typicality and shape complexity were crossed in our research design, resulting in scenes in four conditions. Figure 4 presents examples of critical trials in the complex-shape condition. As in experiment 1, the target object was always of a unique type within the scene, so mentioning color was never needed to distinguish the target from the distractors.

Procedure Each speaker described the sixteen critical scenes, as well as thirty-two filler scenes containing purple greebles. We made two lists containing the same critical trials, but with reversed typicality: target objects that were typically colored for one speaker were atypically colored for another. As such, color typicality and shape complexity were manipulated within participants, while ensuring that each target object appeared in only one typicality condition for each participant. The order of the scenes in each list was randomized for each participant, but there were always two filler trials between experimental ones.

The experiment was performed at our university, and had an average running time of about fifteen minutes. Participants took part in pairs. Who was going to act as the speaker and who as the addressee was decided by rolling a dice. Participants were seated opposite each other at a table, and each had their own computer screen. The screens were positioned in such a way that they did not obstruct the face of either participant, ensuring that eye contact was possible. Apart from these speaker-addressee arrangements, the procedure was identical to experiment 1.

The addressee was presented with the same forty-eight trials as the speaker, but with the objects in a different configuration, and without any marking of the target object. The addressee marked the picture that he or she thought the speaker was describing on an answering sheet. While the addressee was instructed that clarifications could be asked, there were no such requests during the whole experiment, so the data presented here are regarded as one-shot references.

There were two practice trials with greebles, plus the one practice trial used in experiment 1. Once the addressee had identified a target, this was communicated to the speaker, and a button was pressed to advance to the next trial.

Results and discussion

In total, 496 target descriptions were recorded in the experiment. Over 95% of these descriptions (n=472) were intelligible and contained a correct type attribute, resulting in unique reference. As in experiment 1, we analyzed the data using logit mixed models. Initial analyses revealed that stimulus order had no effects, so this was left out in the following analysis. In our model, color typicality and shape complexity were included as fixed factors, standardized to reduce collinearity and increase comparability with experiment 1. Participants and target object types were included as random factors. The model had a maximal random effect structure.

Our analysis, as shown in Figure 5, revealed a significant main effect of color typicality (β=−3.20, SE=0.32, p<.001): 75% of the references to an atypically colored target contained a color attribute, compared to 14% of the references to a typically colored target.

Furthermore, there was a main effect of shape complexity (β=−0.77, SE=0.33, p<.025), as 49% of the references to an object with a simple shape contained color, compared to 38% of the references to a target with a complex shape. Color typicality and shape complexity interacted (β=−0.67, SE=0.27, p<.025): the effect of color typicality on mentioning color was slightly larger for simple objects than for complex objects.

We replicated the color typicality effect found in experiment 1. The methodological differences between the two experiments did not influence the main result. More interestingly, we have also shown that this effect is (to a small degree) modulated by the importance of color in the object's representation in stored knowledge. Color is a more salient feature of stored representations of objects with a simple and uncharacteristic shape. It is for these objects that the color atypicality effect is slightly larger compared to objects with a more complex and characteristic shape.

General discussion

We report two language production experiments that show that atypicality of visually perceived objects affects content determination in reference production. When the color of an object is perceived as more atypical, speakers are more likely to redundantly include color when referring to it. This probability decreases linearly when the color of an object becomes more typical. Furthermore, when an object's shape is characteristic for the object's identity, its color becomes less essential for recognizing the object, and this (marginally) modulates the effect of color atypicality. The main effects of color typicality on content determination in our experiments are undoubtedly strong.

When producing referring expressions, human speakers utilize stored knowledge about the objects they refer to. Stored knowledge contains information about the typical color of objects (e.g., Naor-Raz et al., 2003), and when a property of an object in a visual scene contradicts this information speakers tend to include this property in an identifying description of that object. This is an effect of conceptual, or top-down, visual saliency on content determination.

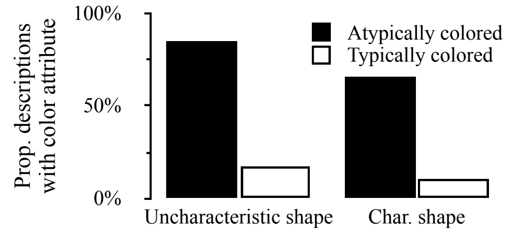

Figure 5: Proportion of descriptions containing color as a function of shape complexity and color typicality

Our findings resonate with other research on the influence of conceptual knowledge on content determination. It corroborates the findings of Mitchell et al. (2013), who show that atypical materials and shapes are preferred over typical ones in content determination, and the finding of Sedivy (2003) that decisions on mentioning an object's type and color are not taken independently of each other, but are indeed influenced by general knowledge about the referred-to object's type.

Our results suggest that the effect of color (a)typicality on content planning can be attributed to conceptual (i.e, top-down) visual saliency. Atypically colored objects attract visual attention in a scene (Becker et al., 2007), and we reason that because the color of the target object draws the speaker's visual attention, color is likely to be mentioned in a referring expression. The scenes in our experiments were designed in such a way that color was equally relevant in all our experimental conditions. So, our results strongly suggest that speakers do not always consider properties of a target object and distractors in terms of what is optimal with regard to informativeness (cf. following the Maxim of Quantity proposed by Grice, 1975), but that the visual attention drawn by certain properties (because they are atypical) also guides the decision to mention these properties in a description. This is arguably simpler to do than to consider the distinguishing value of properties, so it can be characterized as a speaker's decision that is based on a simple heuristic (i.e., a probabilistic judgmental operation; Tversky & Kahneman, 1974, p. 1124). This point of view is in line with other recent work on referring expressions (e.g., Koolen, 2013; Van Deemter et al., 2012b).

Implications for models of reference production

Being able to naturalistically refer to objects in everyday interaction is an important part of Natural Language Generation (NLG; a subfield of Artificial Intelligence). Cognitive scientists and computational linguists have made significant advances in modeling reference production in recent years (e.g., Dale & Reiter, 1995; Krahmer & Van Deemter, 2012; Van Deemter, et al., 2012a, 2012b; Frank & Goodman, 2012). However, considering general knowledge about the typical color of objects in content planning offers a challenge for current Referring Expression Generation (REG) algorithms. Perceptual (bottom-up) saliency is often incorporated in such models in some way (e.g., by considering salient properties such as color before less salient ones such as orientation; Van Deemter et al., 2012a). But top-down saliency based on object knowledge is generally ignored.

An obvious extension for such algorithms, in order to encompass the typicality of the color of objects referred to, is to feed these algorithms with knowledge about what typical colors of objects are. Assuming that object types are readily recognized by artificial agents (which works quite well in controlled environments nowadays, Andreopoulos & Ttsotsos, 2013), a knowledge base containing typical object information can be queried at runtime when a referring expression is generated (Mitchell, et al., 2013). However, for color, a simpler system without a dedicated knowledge base may be effective too. When the dominant color of the first *n* results of a web search for images is computationally determined, the typical color of an object should be derivable. In fact, we expect that this method can even generate the degree of atypicality of the color (cf. our results of experiment 1), by comparing the *n* search results showing the dominant color to the *n* results showing other colors.

Acknowledgments

We thank many of our colleagues and three anonymous reviewers for their valuable comments. We also thank Nanne van Noord for applying the computational saliency maps, and Eveline van Groesen for setting up and running experiment 1.

References

Andreopoulos, A. & Tsotsos, J. K. (2013). 50 Years of object recognition: Directions forward. *Computer Vision and Image Understanding, 117*, 827–891.

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*(3), 255–278.

Becker, M. W., Pashler, H., & Lubin, J. (2007). Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 20–30.

Clarke, A. D., Elsner, M., & Rohde, H. (2013). Where's Wally: The influence of visual salience on referring expression generation. *Frontiers in Psychology, 4,* 329.

Dale, R. & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*(2), 233–263.

Erdem, E. & Erdem, A. (2013). Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of Vision, 13*(4), 1−20.

Frank, M. C. & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998–998.

Gauthier, I. & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, *37*(12), 1673–1682.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 43−58). New York: Academic Press.

Humphreys, G. W. &, Riddoch, M. J., & Quinlan, P. T. (1988): Cascade processes in picture identification, *Cognitive Neuropsychology, 5*(1), 67−104.

Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics, 43*(13), 323−3250.

Koolen, R. (2013). *Need I say more? On overspecification in definite reference*. PhD dissertation, Tilburg University.

Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, *37*(2), 395–411.

Krahmer, E. & Van Deemter, K. (2012). Computational generation of referring expressions: A survey. *Computational Linguistics*, *38*(1), 173–218.

Kuhlen, A. K. & Brennan, S. E. (2013). Language in dialogue: when confederates might be hazardous to your data. *Psychonomic Bulletin and Review, 20*(1), 54−72.

Mitchell, M., Reiter, E., & Van Deemter, K. (2013). Typicality and object reference. In *Proceedings of the 35th annual meeting of the Cognitive Science Society (CogSci)*. Berlin, Germany.

Naor-Raz, G., Tarr, M. J., & Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception*, *32*(6), 667–680.

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

Price, C. J. & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. Q*uarterly Journal of Experimental Psychology A, 41*(4-A), 797−827.

Sedivy, J. (2003). Pragmatic versus form-based accounts of referential contrast: evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*(1), 3–23.

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, *5*(5), 211–215.

Therriault, D., Yaxley, R., & Zwaan, R. (2009). The role of color diagnosticity in object recognition and representation. *Cognitive Processing*, *10*(4), 335–342.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Van Deemter, K., Gatt, A., Van der Sluis, I., & Power, R. (2012a). Generation of referring expressions: Assessing the incremental algorithm. *Cognitive Science*, *36*(5), 799–836.

Van Deemter, K., Gatt, A., Van Gompel, R., & Krahmer, E. (2012b). Toward a computational psycholinguistics of reference production. *Topics in Cognitive Science*, *4*(2), 166–183.

Viethen J., Goudbeek, M, & Krahmer, E. (2012). The impact of colour difference and colour codability on reference production. In *Proceedings of the 34th annual meeting of the Cognitive Science Society (CogSci)*. Sapporo, Japan.