Recalibrating Color Categories Using World Knowledge

Holger Mitterer and Jan Peter de Ruiter

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ABSTRACT—*When the perceptual system uses color to facilitate object recognition, it must solve the* color-constancy *problem: The light an object reflects to an observer's eyes confounds properties of the source of the illumination with the surface reflectance of the object. Information from the visual scene (*bottom-up *information) is insufficient to solve this problem. We show that observers use world knowledge about objects and their prototypical colors as a source of top-down *information to improve color constancy. Specifically, observers use world knowledge to recalibrate their color categories. Our results also suggest that similar effects previously observed in language perception are the consequence of a general perceptual process.*

Observers can use color to identify objects (Brainard, 2004; Oliva & Schyns, 2000; Tanaka, Weiskopf, & Williams, 2001). Here, we show that the reverse is also true: Observers can use objects to identify color. Knowledge about an object's prototypical color influences perception of instances of that object. The same hue is categorized as yellow when viewed on a banana, but as orange when viewed on a carrot. More important, this categorization bias also affects the perception of other, color-neutral objects (i.e., artificial objects that can have different colors, such as socks and cars). Observers who see a banana with an ambiguous hue between yellow and orange later categorize this hue on a color-neutral object as yellow; observers who see a carrot with the same ambiguous hue later categorize the hue as orange.

This *top-down* flow of world knowledge influencing perception is one way in which observers might achieve *color constancy*, the ability to see an object as having the same color despite differing illumination conditions. Achieving color constancy is a perceptual challenge because the light that reaches the eye confounds illumination with the surface reflectance of an object.

Address correspondence to Holger Mitterer, Max Planck Institute for Psycholinguistics, P.O. Box 310, NL-6500 AH Nijmegen, The Netherlands, e-mail: holger.mitterer@mpi.nl.

To use color information for object recognition, observers need to compensate for the influence of illumination and see the color of an object as invariant over illumination conditions. Such compensation is made possible, in part, by using bottom-up information available from the image itself, such as global and local spectral means of the light, as well as the maximal intensities (Kraft & Brainard, 1999) and brightness-hue correlations in the image (Golz & MacLeod, 2002). Generally, color constancy is less than perfect because these sources of information are not always available (Brainard, 2004; Kraft & Brainard, 1999). The degree of color constancy achieved also depends on whether observers are explicitly instructed to focus on color constancy (Arend & Reeves, 1986; Troost & de Weert, 1991).

Color constancy could, in principle, be improved by a top-down flow of world knowledge. In the experiments we report here, we found evidence for such a mechanism: Knowledge about objects' natural colors is used to recalibrate color categories. That is, objects are used to identify colors.

Observers use world knowledge when judging the hue and brightness of an object (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Levin & Banaji, 2006). We replicated this finding in a pilot experiment, in which 12 members of the Max Planck Institute's subject pool participated for pay. All were native speakers of Dutch and had normal color vision. Participants were seated in a booth, where they viewed line drawings of a banana, a carrot, and a sock in the following seven hues (in International Commission on Illumination, CIE, *XYZ* values): Hue 1: 53.2, 42.1, 5.8; Hue 2: 54.3, 45.4, 6.3; Hue 3: 56.0, 49.0, 7.0; Hue 4: 57.8, 53.1, 7.7; Hue 5: 59.7, 57.4, 8.4; Hue 6: 61.8, 62.2, 9.3; Hue 7: 64.2, 67.3, 10.2. Participants were instructed that on each trial, they should press the right button of a two-button box if they perceived the object as orange and the left button if they perceived the object as yellow. Each of the three objects was presented 15 times in each of the seven hues, for a total of 315 trials. Analysis of the results revealed, in addition to the trivial effect of hue ($b_{hue} = 10.2$, $p_{rep} > .99$; see the Results section of Experiment 1 for the statistical method used), that there were significantly more "yellow" responses for the banana than for the sock ($b_{object=banana} = 0.88$, $p_{rep} > .99$, with the sock
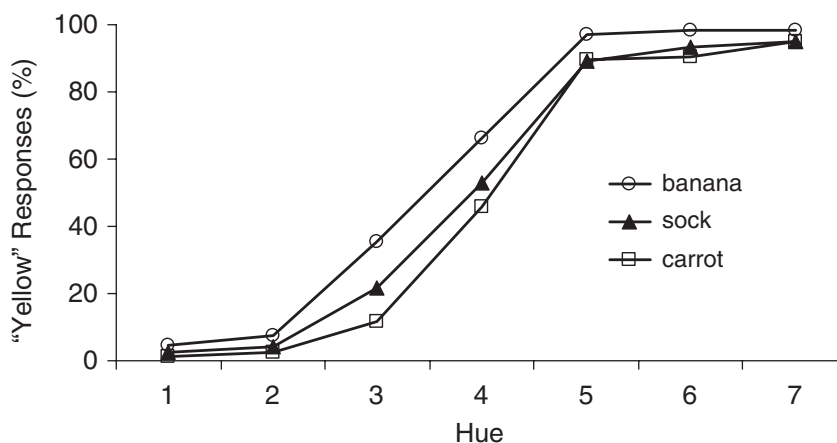
Authors' personal copy

Fig. 1. Results from the pilot experiment: percentage of trials on which the color of the line drawing was categorized as yellow, as a function of the object depicted (banana, sock, or carrot) and its hue (seven hues of the yellow-orange continuum).

condition mapped on the intercept) and fewer "yellow" responses for the carrot than for the sock ($b_{object=carrot} = -0.45$, $p_{rep} = .988$; see Fig. 1).

These results are, by themselves, not evidence for top-down processing; they show only that bottom-up information is merged with world knowledge to achieve an optimal perceptual decision (Massaro, 1998; Norris, McQueen, & Cutler, 2000). To be evidence for a true top-down effect, results would have to show that world knowledge about object A influences the perception of a neutral object B (Pitt & McQueen, 1998; Samuel & Pitt, 2003). For instance, is the likelihood that an observer will perceive a yellowish sock as yellow increased if the observer has recently seen a banana in the same yellowish hue?

EXPERIMENT 1

To find a true top-down effect, we designed an experiment with three phases. In the first phase (exposure phase), participants viewed colored line drawings of prototypically yellow and prototypically orange objects, presented in an achromatic context (see the left-hand side of Fig. 2): For half of the observers, the prototypically orange objects were shown in what our pilot experiment indicated was a good orange, and the prototypically yellow objects were shown in a hue that was ambiguous between yellow and orange; for the other half of the observers, the prototypically orange objects were shown in the same ambiguous hue, and the prototypically yellow objects were shown in a good yellow. Participants were asked to remember which objects they saw, so color was irrelevant to their task. In the second, and critical, phase (color-categorization phase), the participants' task was to judge whether a sock (an object without a prototypical color) was yellow or orange. This task was introduced as a mere filler intended to produce a time delay between the picture-encoding task and a later memory test; this should have discouraged observers from using explicit strategies relating the

stimuli in the two tasks. None of the participants questioned this cover story. In the third, and final, stage of the experiment (memory phase), object names were presented one at a time, and participants indicated whether each name corresponded to one of the pictures they had seen during the first phase.

Method

Participants

Fifty-two members of the Max Planck Institute's subject pool participated in the experiment for payment (€4). All were native speakers of Dutch and had normal color vision.
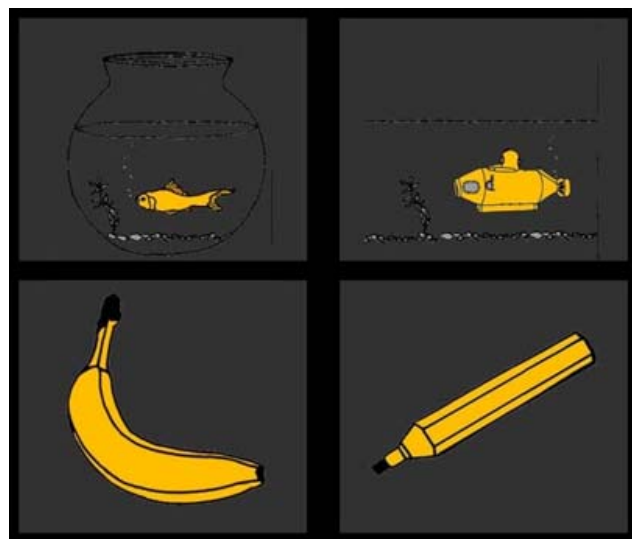

Fig. 2. Examples (in the ambiguous hue) of pictures used in the exposure phase of Experiments 1 and 2. The pictures on the left (Experiment 1) have a prototypical color (orange for a goldfish and yellow for a banana), whereas the pictures on the right (Experiment 2) have no prototypical color.

Stimuli and Procedure

We prepared line drawings of 24 objects: 11 prototypically orange objects, 12 prototypically yellow objects, and 1 object with both prototypically yellow and prototypically orange parts (a rubber duck). Of the 11 prototypically orange objects, 8 were natural (carrot, mandarin orange, pumpkin, lobster, goldfish, orangutan, egg yolk, beak of a swan), and 3 were artificial (traffic cone, EasyJet plane, basketball). Of the 12 prototypically yellow objects, 10 were natural (lemon, pineapple, banana, bee, chick, canary, wasp, daisy, sunflower, dandelion), and 2 were artificial (tennis ball, Dutch train). The prototypically yellow parts of the objects were presented in what the pilot experiment showed to be a good yellow (Hue 7) or in an ambiguous color (Hue 4). The prototypically orange parts of the objects were presented in what the pilot experiment showed to be a good orange (Hue 1) or in the same ambiguous color. The background and the nonprototypically yellow and orange parts of the objects were achromatic. Half of the participants saw the prototypically orange objects in the ambiguous hue and the prototypically yellow objects in the good yellow, and the other half saw the prototypically orange objects in the good orange and the prototypically yellow objects in the ambiguous hue. The first group, which saw the good yellow on yellow objects, was expected to develop an orange bias (i.e., a bias to categorize the ambiguous color as orange); the second group, which saw the good orange on orange objects, was expected to develop a yellow bias (i.e., a bias to categorize the ambiguous color as yellow).

Participants were seated in a sound-attenuated and shaded cubicle; the only light was provided by the computer monitor. They received instructions via the computer screen. Each of the 24 objects was presented three times in the exposure phase, and participants could view each image as long as they wanted while they were trying to memorize it. Instructions in the color-categorization phase were identical to those in the pilot experiment. Participants categorized each of five hues (shown on a sock) 20 times. These were the same hues as in the pilot experiment minus the two endpoints of the seven-step continuum. In the final, memory phase, the names of 24 objects and 24 related lures (e.g., *trolley* was a lure for *train*) were presented, and participants had to indicate whether or not they had seen each object before.

Results and Discussion

We analyzed the data using linear mixed-effects models (Baayen, Davidson, & Bates, in press) in which individual data points were predicted with fixed experimental effects that were crossed with a participants random effect. This approach made averaging the data superfluous, which increased the statistical power and allowed us to use trial-specific covariates, such as trial number. Moreover, by using a logistic linking function, we could accommodate the categorical nature of the dependent variable, as well as the typical sigmoid shape of categorization functions.

Hue was entered as a continuous predictor, and object identity was entered as a categorical predictor. For categorical predictors in such models, one level is mapped on the intercept, and other levels are coded as binary dummy variables. Continuous predictors were scaled to range from 0 to 1. The value of the coefficient thus indicates how the logarithm of the odds (log odds) for a "yellow" response changes with a change in condition for a categorical predictor and the degree to which the log odds for a "yellow" response changes within the range of a numerical predictor.

Exposure to the ambiguous hue on either prototypically yellow or prototypically orange objects affected the perception of hues on the color-neutral sock in the critical color-categorization task. Participants who had seen an ambiguous hue on prototypically yellow objects and had seen orange objects in a good orange identified the ambiguous hues in the middle of the test continuum as yellow more often than did participants who had seen the ambiguous color on prototypically orange objects ($b_{\text{yellow bias}} = 1.64$, $p_{\text{rep}} = -.980$, with the orange-bias, or good-yellow, group mapped on the intercept; see Fig. 3a). The overall effect of exposure got smaller over the course of the color-categorization task ($b_{\text{Yellow Bias} \times \text{Trial}} = -1.58$, $p_{\text{rep}} = .997$; see Fig. 3b, in which trials were grouped into five blocks) and was larger for participants who took longer to encode the objects in the exposure phase ($b_{\text{Yellow Bias} \times \text{Encoding Time}} = -6.09$, $p_{\text{rep}} = .957$) and for participants who better remembered which objects were shown ($b_{\text{Yellow Bias} \times \text{Recognition Accuracy}} = -8.67$, $p_{\text{rep}} > .969$).

As Figure 3a shows, group differences were found for Hues 3 and 4, but not Hue 5. Apparently, this latter hue is too close to the prototype of the category "yellow," so it is inoculated against top-down influences. This explanation is supported by the results from the pilot experiment, in which the effects of object identity were also strongest for Hues 3 and 4, and nearly absent for Hue 5.

In demonstrations of direct effects of world knowledge on color perception—as in our pilot experiment, when participants judged the hue of a banana and a carrot—the stimuli activate concepts such as "yellow" and "orange." The activation of these concepts can then influence the decision-making process—and hence the results—without influencing perception itself. Such an effect was not possible with the neutral stimulus used in the current experiment, because the picture of a sock does not differentially activate the concepts "yellow" and "orange." Thus, it is difficult to argue that the observed effect occurred at the level of decision making. The recalibration of the color categories is therefore most parsimoniously explained as a top-down recalibration of perception due to the prior exposure.

An alternative account would be that, on seeing a banana with an ambiguous hue, participants decided that they apparently had to call such a hue "yellow" in this experiment. We have two objections to such an account, a theoretical one and an empirical one. From a theoretical point of view, we argue that top-down recalibration helps to achieve perceptual invariance in the face
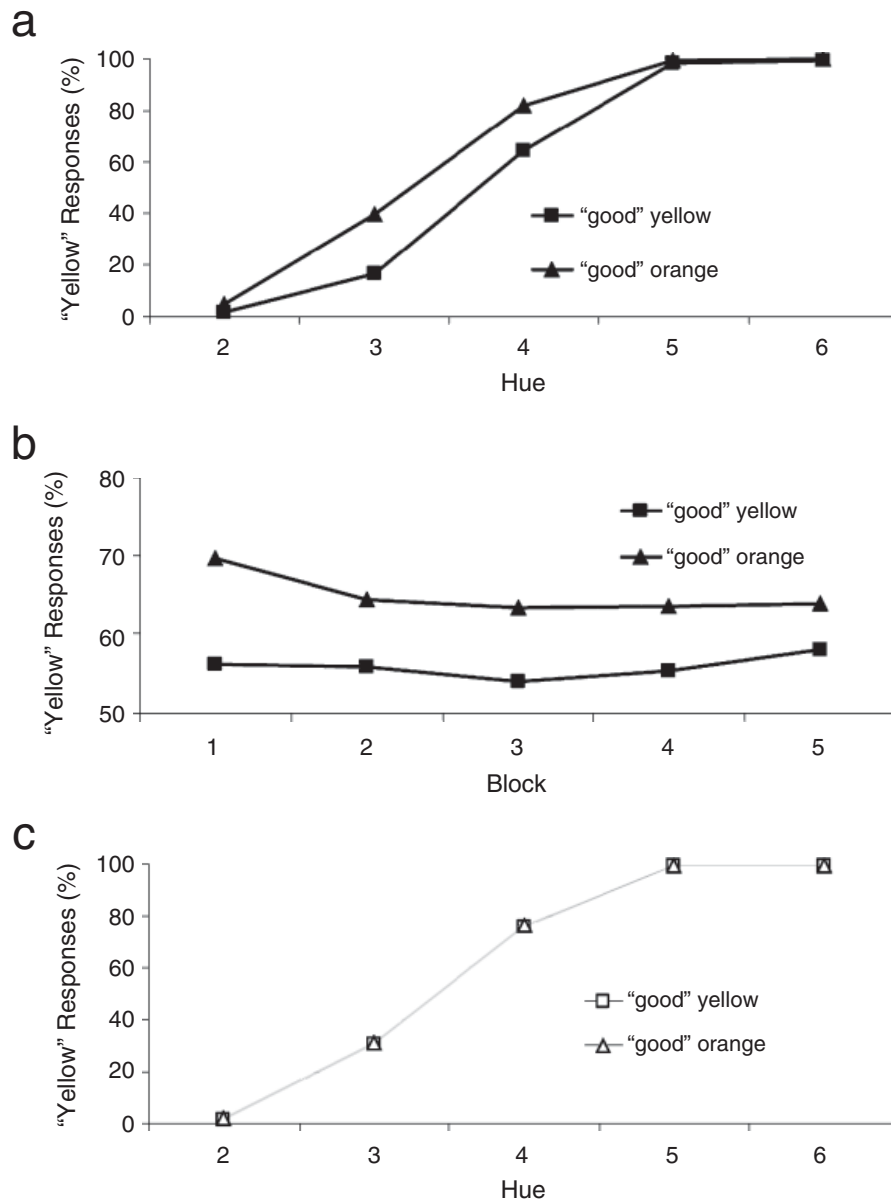
a

b

c

Fig. 3. Results from the color-categorization phase in Experiments 1 (objects with a prototypical color) and 2 (objects without a prototypical color). The graphs show the percentage of ''yellow'' responses as a function of (a) hue and (b) test block in Experiment 1 and as a function of (c) hue in Experiment 2.

of ecological variance, and adjusting a decision criterion would not help to reduce ecological variance. Note that such a functional argument holds only for a delayed feedback effect, as in the current case, because on-line feedback is not beneficial for optimal perception (Norris et al., 2000, pp. 301–306). On an empirical level, it is difficult to explain the gradual decline of the recalibration effect over the testing blocks on the basis of this alternative account. A criterion-shift account predicts that the effect should be stable over the course of the experiment, and leaves unexplained why participants reshifted their criterion during the color-categorization phase. In contrast, a top-down

recalibration would have been weakened gradually as category boundaries were gradually updated during exposure to better examples (i.e., the endpoints of the test continuum) of the color categories in the color-categorization phase.

EXPERIMENT 2

In Experiment 1, participants were exposed to different limited ranges of hues, either a range from good yellow to the ambiguous hue or a range from the ambiguous hue to good orange. Therefore, the effect of exposure could have been due to an adaptation

(or range) effect, in which the ambiguous hue was perceived as different from the good yellow or orange. Such pure adaptation effects in the absence of a top-down influence have been observed in both speech and color perception (Bornstein & Korda, 1985). In our second experiment, we tested this possibility by exposing participants to the same hues, but on objects without a prototypical color.

Method

Participants
Fifty-two members of the Max Planck Institute's subject pool participated in the experiment for payment (€4). All were native speakers of Dutch and had normal color vision. None of them had participated in Experiment 1.

Stimuli and Procedure
The procedure included the same three phases as in Experiment 1, the only change being in the exposure phase. The ranges of hues from the previous experiment were tested, but they were presented on objects without a prototypical color (see the right-hand side of Fig. 2). For instance, the banana was replaced by a crayon, which has a similar shape but no prototypical color. The ratio of colored surface to achromatic surface in each picture was similar to the ratio for the corresponding prototypically colored object in Experiment 1. Again, half of the participants viewed objects in good yellow or an ambiguous hue, and the other half viewed objects in good orange or the same ambiguous hue. If the effect of exposure in the first experiment was due to a range effect, a similar effect would be expected in this experiment.

Results and Discussion
As Figure 3c shows, there was no difference between the group with good-yellow exposure (i.e., the orange-bias group) and the group with good-orange exposure ($b_{\text{yellow bias}} = 0.05, p_{\text{rep}} = .535$, with the orange-bias group mapped on the intercept). Therefore, the results of Experiment 1 cannot be attributed to an adaptation effect. Exposure to good yellow and ambiguous orange on objects without an intrinsic color does not lead to a recalibration in color categorization. Hence, the effect found in Experiment 1 must have been due to conceptual knowledge, and cannot be explained by adaptation effects.

GENERAL DISCUSSION

Our experiments demonstrate that observers use knowledge of an object's color to recalibrate their color categories. This finding has two important theoretical implications. First, it undermines the assumption that language perception is fundamentally different from perception in other domains. Two earlier experiments—investigating auditory and visual word recognition—showed similar top-down effects for phoneme and letter categories (Norris, Butterfield, McQueen, & Cutler, 2006;

Norris, McQueen, & Cutler, 2003). Because recalibration of categories also occurs in color perception, it is not unique to language, but rather is a general perceptual strategy that language processing takes advantage of (Dawkins, 1989, chap. 11; Deacon, 1997, chap. 4).

Second, our results show that perception does make use of top-down processing. This top-down processing is, however, delayed. Immediate top-down influences would override the perceptual evidence, and with such feedback, observers might hallucinate and perceive unripe and overripe bananas as yellow. Delayed feedback for learning prevents such illusions, but still utilizes prior probabilities provided by world knowledge to achieve perceptual constancy. We suggest that this mechanism has evolved to strike an optimal balance between top-down and bottom-up information in perception.

REFERENCES

Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America*, *3*, 1743–1751.

Baayen, H.R., Davidson, D.J., & Bates, D.M. (in press). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*.

Bornstein, M.H., & Korda, N.O. (1985). Identification and adaptation of hue: Parallels in the operation of mechanisms that underlie categorical perception in vision and in audition. *Psychological Research*, *47*, 1–17.

Brainard, D.H. (2004). Color constancy. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 948–961). Cambridge, MA: MIT Press.

Dawkins, R. (1989). *The selfish gene* (new ed.). Oxford, England: Oxford University Press.

Deacon, T.W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: Norton.

Golz, J., & MacLeod, D.I.A. (2002). Influence of scene statistics on colour constancy. *Nature*, *415*, 637–640.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K.R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*, 1367–1368.

Kraft, J.M., & Brainard, D.H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences, USA*, *96*, 307–312.

Levin, D.T., & Banaji, M.R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, *135*, 501–512.

Massaro, D. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Norris, D., Butterfield, S., McQueen, J.M., & Cutler, A. (2006). Lexically guided retuning of letter perception. *Quarterly Journal of Experimental Psychology*, *59*, 1505–1515.

Norris, D., McQueen, J.M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–324.

Norris, D., McQueen, J.M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.

Oliva, A., & Schyns, P.G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41*, 176–210.

Pitt, M.A., & McQueen, J.M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*, 347–370.

Samuel, A.G., & Pitt, M.A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language, 48*, 416–434.

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

Troost, J.M., & de Weert, C.M.M. (1991). Naming vs. matching in color constancy. *Perception & Psychophysics, 50*, 591–602.