Object knowledge modulates colour appearance

Christoph Witzel¶
Department of Psychology, University of Giessen, Otto-Behaghel-Straße 10F, 35394 Giessen, Germany;
e-mail: Christoph.Witzel@psychol.uni-giessen.de

Hanna Valkova
Department of Psychology, University of Giessen, Otto-Behaghel-Straße 10F, 35394 Giessen, Germany;
e-mail: Hanna.Valkova@web.de

Thorsten Hansen
Department of Psychology, University of Giessen, Otto-Behaghel-Straße 10F, 35394 Giessen, Germany;
e-mail: Thorsten.Hansen@psychol.uni-giessen.de

Karl R Gegenfurtner
Department of Psychology, University of Giessen, Otto-Behaghel-Straße 10F, 35394 Giessen, Germany;
e-mail: gegenfurtner@uni-giessen.de

Abstract. We investigated the memory colour effect for colour diagnostic artificial objects. Since knowledge about these objects and their colours has been learned in everyday life, these stimuli allow the investigation of the influence of acquired object knowledge on colour appearance. These investigations are relevant for questions about how object and colour information in high-level vision interact as well as for research about the influence of learning and experience on perception in general. In order to identify suitable artificial objects, we developed a reaction time paradigm that measures (subjective) colour diagnosticity. In the main experiment, participants adjusted sixteen such objects to their typical colour as well as to grey. If the achromatic object appears in its typical colour, then participants should adjust it to the opponent colour in order to subjectively perceive it as grey. We found that knowledge about the typical colour influences the colour appearance of artificial objects. This effect was particularly strong along the daylight axis.

1 Introduction

What colour is a smurf? If you are very familiar with this imaginary creature, a particular blue should come to your mind, and you should be able to answer immediately that a smurf is blue. Now, does this internalised knowledge about the smurf's typical colour by itself result in your perceiving a smurf as being blue? Even when it is grey?

It has been shown, for fruit images, that knowledge about their typical colour influences their colour appearance (Hansen et al 2006; Olkkonen et al 2008). If one knows, for example, that a banana is yellow, it appears to be yellow even when it is completely achromatic. This phenomenon is evidence for a memory colour effect. A memory colour is the typical colour of an object that we have memorised due to our past experience with the respective object (Bartleson 1960, page 73; Hering 1920 [1878], pages 6–7). The memory colour effect refers to the idea that memory colours modulate the colour appearance of the respective objects' actual colours. The memory colour effect requires that the object is strongly associated with a typical colour and that the beholder is highly familiar with the object. How strongly a particular object refers to a typical colour through memory colours is called colour diagnosticity (Biederman and Ju 1988, page 41; Tanaka and Presnell 1999, page 1141). For example, a banana may be considered as highly colour diagnostic, since it refers directly to yellow. In contrast, a sock is not colour diagnostic but colour neutral, since socks occur in many different colours. According to the original idea of memory colours, knowledge

¶ Corresponding author.

about the typical colour of an object should be acquired through frequent visual experience with the respective object (Bartleson 1960, page 73; Hering 1920 [1878], page 7). To test this idea, we investigated the memory colour effect for artificial objects. Artificial objects such as the smurf are man made. Consequently, in contrast to natural objects, their colours are not determined by nature but by humans. Everyday-life experiences with these objects are bound to the cultural and historical context in which they occur. A mailbox is a good example. While it is typically yellow in Germany, it is red in the United Kingdom, blue in the United States, and green in China. A memory colour effect for these objects would indicate that the colour appearance of objects is influenced through knowledge that has been learned during past experiences.

This research question is relevant for at least two fields of research. Firstly, it contributes to research about the influence of past experiences and learning on perception in general. For example, Schyns and colleagues have used novel stimuli (ie, stimuli that do not exist in everyday life) to show that people describe these stimuli differently depending on their learning experience (Schyns and Rodet 1997; Schyns et al 1998). Moreover, Gosselin and Schyns (2003) have shown that the assumption that a particular object exists in random noise images led to the actual perception of these objects in the images. Such research shows that human apprehension of reality is highly constructive and creative in nature. These assumptions have been at the core of metatheories about the human mind in empiricism (Hume 1894 [1748]), constructivism (Watzlawick 1984; Wittgenstein 1953), and situated cognition (King 2000; O'Connor and Glenberg 2003; St Julien 1997). Since colour is such an elementary visual attribute, it may be considered as a prime example for investigating the influence of past experience on cognition. And, indeed, the first studies on the memory colour effect were framed by this perspective (Adams 1923; Baker and Mackintosh 1955; Bruner et al 1951; Duncker 1939; Fisher et al 1956; Harper 1953; Helmholtz 1867; Hering 1920 [1878]). Secondly, the present investigations contribute to a major question in the field of colour research. From the perspective of colour research the question of whether we can perceive objects and their colour independently of each other is tied to the debate about functional segregation in high-level vision (Gegenfurtner 2003; Gegenfurtner and Kiper 2003). Functional segregation refers to the idea that colour is perceived separately from other elementary visual attributes, such as shape, texture, or depth. On the neuropsychological level this would imply that information about colour is processed by cortical cells that are functionally separable from those that process other visual attributes (Wandell 1995, page 334). The question of whether the later cortical stages that are at the basis of our final visual impression of the environment are functionally segregated is in the focus of today's colour research (Bloj et al 1999; Ling and Hurlbert 2004; Miceli et al 2001; Naor-Raz et al 2003, page 677; Werner 2007).

In the case of colour the beholder's final visual impression is called colour appearance. There are at least four phenomena of colour appearance where memory colours play a major role. Firstly, memory colours help object and scene recognition (Tanaka et al 2001). When the depicted objects are highly colour diagnostic, the presence of the object colours boosts significantly scene recognition (Gegenfurtner and Rieger 2000; Oliva and Schyns 2000; Wichmann et al 2002) as well as object recognition (Humphrey et al 1994; Nagai and Yokosawa 2003; Naor-Raz et al 2003; Nicholson and Humphrey 2003; Tanaka and Presnell 1999; Therriault et al 2009; but see Wurm et al 1993). Secondly, memory colours support colour constancy (Emmerson and Ross 1987; Hurlbert and Ling 2005; Ling and Hurlbert 2006; for a discussion see also Olkkonen et al 2008). Thirdly, memory colours are used in colour memory (Ratner and McCarthy 1990; Siple and Springer 1983; Van Gulick and Tarr 2010). When people have to memorise the colour of an object, their memory is led by the object's

typical colour. Finally, memory colours are involved in colour naming. It has been shown that people name an ambiguous colour differently depending on how the ambiguous colour is paired with colour diagnostic objects (Mitterer and de Ruiter 2008). This has also been shown for culturally specific associations between objects and colour names (Mitterer et al 2009). These findings indicate that memory colours work as a perceptual anchor that provides the reference for visual estimations and judgements. In the light of these newer findings, the old unresolved question of whether the role of memory colours is restricted to judgement biases or involves an actual perceptual retuning of colour appearance (Harper 1953) gains new relevance. To answer this question, it is crucial to investigate the direct involvement of memory colours in colour appearance.

Indeed, several studies have shown an effect of memory colours on measures of colour appearance. Early studies indicated that people overestimate the saturation of an object's diagnostic colours (Bartleson 1960; Duncker 1939; Harper 1953; Herring and Bryden 1970; Newhall et al 1957; Siple and Springer 1983; White and Montgomery 1976). However, some other studies found inconsistent results (Bolles et al 1959; Bruner et al 1951; Leibovich and Paolera 1970; Pérez-Carpinell et al 1998), particularly for artificial objects such as a schematic heart (Fisher et al 1956). Using different techniques, recent studies have confirmed that people overestimate the amount of the typical hue in colour diagnostic fruit stimuli (Hansen and Gegenfurtner 2006; Hurlbert and Ling 2005; less clear in Yendrikhovskij et al 1999, page 401, figure 5). While these findings may still be attributed to a judgemental bias, Olkkonen and colleagues have finally shown that even grey fruits induced the perception of their typical colour (Hansen et al 2006; Olkkonen et al 2008). Through a colour adjustment procedure, they let participants adjust eight fruits so that they appear to be completely achromatic, that is grey. They found that participants shifted the grey adjustments to the colour that is opposite to the original colour of the respective fruit. This indicates that people counteracted the impression of the typical colour when the object was actually grey. Hence, only by adjusting the object in the opponent colour did it subjectively appear grey to them. Olkkonen et al (2008) and Hansen and Gegenfurtner (2006) could even explain the inconsistency of memory colour effects in earlier studies that used outline shapes (Bolles et al 1959; Bruner et al 1951; Fisher et al 1956). They showed that the memory colour effect declines with the loss of perceptual information, which is high for photos and very low for outline shapes.

Fruits, however, are a special kind of object; namely, they are natural objects. In general, human colour vision is tuned to the natural environment throughout its phylogenesis (Nathans 1999; Surridge et al 2003). And ripe fruits in particular are assumed to play an important role in evolutionary adaptation (Regan et al 2001; Sumner and Mollon 2000). As with other mechanisms of human colour vision, the particularities of fruit objects might play a role in the memory colour effect because the human visual system is tuned to them. Fruits have specific natural textures, shapes, and colour distributions. Moreover, the colour distributions of the fruits used in the former experiments cover only a part of the colour space. Since there are few saturated blue, purple, and pink fruits in our geographic region that are at the same time highly colour diagnostic, the fruit stimuli were restricted to green, yellow, orange, and red hues. The aforementioned studies on the role of perceptual information in the memory colour effect (Hansen and Gegenfurtner 2006; Olkkonen et al 2008) used stimuli that lacked natural texture and colour distributions. The observation that memory colour effects decline for such objects may be understood as supporting the idea that these object features play a particular role in the memory colour effect (Olkkonen et al 2008). With regard to the actual determinants of the memory colour effect, another question arises. The observed decline of the memory colour effect for stimuli with reduced perceptual information might be due only to a decreased recognisability of the images. Alternatively, the

memory colour effect might also be tied to perceptual features such as three dimensionality, natural texture, or complex colour distributions. Furthermore, Joseph and Proffitt (1996) have shown a strong interaction between object recognition and stored knowledge about the typical object colour. This interaction occurred not only for images of objects but also for the names of objects. With regard to the memory colour effect, we may therefore wonder whether memory colour effects occur only for concrete objects or whether knowledge about the association between a colour and a more abstract concept may also elicit memory colour effects.

In the present study we wanted to make sure that the memory colour effect is really due to learning through experience, and not to the particularities of natural objects or colour distributions. For this reason we used artificial objects that are sampled throughout the whole colour space. Artificial objects are man made, and their colour is determined by humans too. As a result, their memory colours are bound to the cultural context in which the objects occur. This implies that the association between these objects and their colours must be learned in everyday life through experience with the objects. To investigate the perceptual determinants of the memory colour effect, we selected artificial objects that varied in the complexity of their perceptual features as well as in the abstractness of their colour diagnostic characteristics. In our main experiment we tested whether the memory colour effect also occurs for these artificial objects. If the memory colour effect is solely due to the learned object–colour association, it should appear independently of the particularities of their perceptual features. However, objects that are bound to a cultural context are particularly prone to variability in colour diagnosticity. At the same time, high colour diagnosticity is a prerequisite for the memory colour effect. In order to guarantee high colour diagnosticity for our stimuli, we conducted a preliminary reaction time experiment to identify stimuli with high diagnosticity.

2 Experiment 1: identification of colour diagnostic stimuli

Colour diagnosticity consists of two aspects. The objective aspect of colour diagnosticity is the typicality or characteristicness of the object colour. This means that in the outside world the object occurs unequivocally with only minor variations in a typical colour. The subjective aspect of colour diagnosticity is that the beholder must know about the typicality of the object colour. For example, when playing pool, the colour of the ball with number one is always yellow. However, people who do not play pool very frequently may not be able to recall the colour of the ball with number one. Moreover, for an interaction between memory colour and colour appearance the association between the object and its typical colour must be so strong that the object triggers the colour automatically. Finally, in order to produce a memory colour effect, the object should be highly recognisable on the image used as the stimulus. It must be recognisable, not only in its typical colour but also when the colour changes during the colour adjustment procedure. The stimuli to be selected for the colour adjustment procedure should maximise all these characteristics.

For this purpose, we developed a reaction-time paradigm. Tanaka and Presnell (1999) have measured colour diagnosticity by the consistency with which their participants name the typical colour of an object, and the consistency with which people include the typical colour among the first three characteristic attributes of an object. However, these measures are purely conceptual, not perceptual. They measure only semantic associations between objects and features. They are primarily based on deliberate reflections, not automatic reactions. Moreover, these measures do not guarantee that the images of the objects are well recognisable since there is no image presentation in this procedure. Naor-Raz et al (2003) have applied a Stroop interference task to colour diagnostic objects (cf Stroop 1935). They
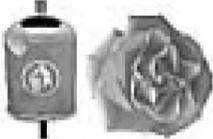
presented objects in colours that were congruent or incongruent with their typical colours. When people had to name the presented colour of the object, Naor-Raz et al could show an interference of the memory colour by reaction times. These reaction times measure automatic object–colour associations, since Stroop interference is an automatic effect. However, this measure might not be sensitive enough to differentially evaluate different degrees of colour diagnosticity. Since a correct answer to the task can be given without recognising the object, this measure will also not discard unrecognisable images by low accuracy rates.

Instead, we used speed and accuracy to measure how well achromatic images are associated with the typical colour of the depicted object. We converted a large pool of candidate images into greyscale images. In each trial we presented one of these achromatic images and let participants indicate as fast as possible the usual colour of the depicted objects. In order to control for subjective colour diagnosticity, these measures were collected for the same participants that took part in the colour adjustment experiment in a later session. In regard to the dependent variables, low accuracy rates indicate that images are either not recognisable or not colour diagnostic at all. Reaction times were used to measure automaticity. Comparatively low average reaction times are assumed to separate automatic object–colour associations from reflective, deliberative ones. At the same time, subjective colour diagnosticity of candidate stimuli should be consistent over time and across participants in order to produce a statistically reliable memory colour effect. Hence, candidate stimuli should also yield a comparatively low variability of reaction times, as measured through standard errors of mean across blocks and participants.

However, it is an open question as to whether low accuracy rates do identify inappropriate stimuli, and whether reaction times can differentiate between different levels of automatisation of the colour–object association. Moreover, it is also an open question as to whether there are supplementary influences besides the memory colour that significantly affect the accuracy rates and the reaction times. In particular, even though we controlled for the average luminance of the stimuli, there might still be some information about the original colour in the luminance distribution and in luminance contrast. For example, in objects that originally have a bright colour, such as yellow or pink, there is a wide transition range between shades and the directly illuminated colour areas, but a short transition between such areas and white gloss; for objects that had originally a dark colour, such as blue, the inverse is true.

In order to verify whether our measures really represent memory colours, we included eight kinds of control stimuli. For an overview of these kinds of control objects and the corresponding predictions, see table 1. Firstly, we included a block with coloured disks, whose actual colour had to be indicated. The reaction times for these responses should provide a lower boundary to the reaction times for the achromatic images of colour diagnostic objects. Secondly, we included objects that we considered a priori as unrecognisable and which should yield low accuracy rates and high reaction times. For example, the grey silhouette of an orange is basically a round disk, and it is difficult to recognise it as an orange. Thirdly, there was a colour-neutral object, a striped sock, which we expected to elicit random colour assignments and high reaction times. Fourthly, there was a woollen pompom that served as a novel object that does not occur in everyday life. This novel object was shown in its original colour once at the beginning of the experiment. Since people had to recall its colour explicitly, colour assignments should result in high reaction times with high accuracy rates. A small subgroup of the participants was familiarised in everyday life with this object. This group should dispose of higher automatisation and, hence, show lower reaction times. Fifthly, there were ambiguous objects, which exist in at least two colours in everyday life.

Table 1. Predictions and results for control stimuli. The leftmost column describes the category of control stimuli and gives an example for this category. The other columns show predictions and results for accuracy rates and reactions times. Results shown in green conform to predictions, those in red contradict predictions, and those in black cannot be differentiated because of ceiling effects; *photos* = natural photos; *outline* = outline shapes, *denat* = 'denaturalised' (without natural texture), ie, white painted. * than one row above. ** in the experiment the sock was upright (turned 90 deg. anticlockwise). *** No sensible answer was possible, hence responses should be around 50%.

Type		Accuracy rates predicted	Accuracy rates results	Reaction times predicted
Coloured disks uniform noise		highest	95-100%	lowest
Unrecognisable orange	Unrecognisable	low	23-90%	high
	outline	higher*	95%	lower*
	photos	higher*	98%	lower*
Colour-neutral *sock***		random answers***	75-100%	high
Novel object pompom	Non-learners (27)	high	91%	high
	Learners (4)	high	100%	lower*
Two colours–Two responses		low	52-88%	high
Two colours–One response		high	84-99%	high
Olkkanen et al (2008)	photos	high	98%	low
	denat	lower*	99%	higher*
	outline	lower*	93%	higher*
Duncker (1939)	leaf (outline)	low	94%	high
	donkey (outline)	low	100%	high

For example, we included a pepper, which exists in red, yellow, and green. These objects should yield low accuracy rates because of their ambiguity. Sixthly, there were ambiguous objects, whose colour assignment could be determined through explicit inference because the alternative colour was not available as a response option. For example, there was a grey chess piece but no response option for 'black', only for 'white'. These objects should yield high accuracy rates and also high reaction times due to the time that is necessary for the explicit inference. Together, the stimuli that involve explicit inference or recall should provide an upper boundary for those reaction times that are good candidates to reveal automatic colour diagnosticity. If our paradigm measures the degree of colour diagnosticity well, the candidate stimuli should spread between the lower and upper borders. Seventhly, we also added the stimuli from Olkkonen et al (2008) as well as the outline shapes from Duncker (1939) to the pool of candidate stimuli. The correlation between the reaction times and the memory colour effect of the stimuli from Olkkonen et al (2008) may inform us about the relationship between the two measurements and validate this procedure. Finally, we also added images of objects that have typically an achromatic colour in order to identify an adequate control stimulus for the colour adjustment method (see below).

To choose the objects for the memory colour experiment, we applied the following criteria to the stimulus selection. Stimuli that are not recognisable or not subjectively colour diagnostic are not appropriate for the memory colour experiment. Hence, we discarded all stimuli that had accuracy rates lower than 95%. Furthermore, the stimuli should have the lowest average reaction times and low standard errors of mean. At the same time their reaction times should be low enough to separate these stimuli well from stimuli with high reaction times. Finally, the memory colours of the stimuli should cover different parts of the colour space, especially those parts that were not sampled in the study of Olkkonen et al (2008).

2.1 Method

There were several methodological and pragmatic constraints to the overall design of the experimental paradigm. Firstly, we had to restrict the number of repetitions for each stimulus, since a high repetition rate would induce a stimulus-specific automatisation of responses. In order to obtain enough data, we compensated by collecting data from a large group of participants. Moreover, due to another experiment (not reported here), we combined the pool of images of artificial objects with a pool of natural objects. In order to indicate the typical colour of a grey object, participants had to assign a Basic Colour Term to the object. Basic Colour Terms are particularly useful for this task, since they refer to colour categories that are most intuitive (in terms of particularly low reaction times) and that are most consistent between individuals (Berlin and Kay 1969; Boynton and Olson 1990; Guest and Van Laar 2000; Sturges and Whitfield 1997). In order to cover the whole colour space, we used all eight chromatic Basic Colour Terms. Additionally, we included the achromatic categories 'grey' and 'white' for the candidate control stimuli. Finally, we designed the response mode by grouping the colour categories into pairs. The reason for using just two colour terms in a block was to prevent spurious reaction times through key search and increased false negative errors through key confusions. The categories were paired based on three criteria. Firstly, the two categories should not be adjacent in colour space in order to avoid hesitations and inter-individual differences in the categorisation of ambiguous object colours. Secondly, in order to obtain equal prior probabilities for correct answers, the amount of available stimuli should be approximately the same in the respective two categories. Thirdly, the colours of the two categories should have approximately the same lightness. This should avoid the possibility of luminance information influencing the responses. In a compromise to meet these criteria, we paired blue with brown, red with yellow, green with

orange, violet with grey, and pink with white. Please see table 2 for an overview of how colour categories were paired.

Table 2. Stimuli of reaction time experiment. The rows show the different category pairings that defined the two response options (order of columns is arbitrary). At the top of each cell the colour name is shown with the overall number of stimuli in parentheses. The different kinds of objects are listed separately (natural, artificial, or novel objects), followed by the number of the respective objects in this category. In parentheses behind this number, objects are further specified as follows. In the sample of artificial objects: *obj* = object (represents itself), *sgn* = sign (nonlinguistic symbol), *lg* = logo (written label); in the sample of natural objects: *ph* = natural photos, *den* = 'denaturalised' (without natural texture), *out* = outline shapes, *cart* = cartoons (clip-art drawings).

Category 1			Category 2		
BLUE (26)			BROWN (25)		
Artificial:	11	(2*obj*, 4 *sgn*, 5 *lg*)	Artificial:	6	(6 *obj*)
Natural:	15	(4 *ph*, 3 *den*, 4 *out*, 4 *cart*)	Natural:	19	(7 *ph*, 3 *den*, 3 *out*, 6 *cart*)
YELLOW (26)			RED (26)		
Artificial:	10	(6 *obj*, 2 *sgn*, 2 *lg*)	Artificial:	11	(5 *obj*, 5 *sgn*, 1 *lg*)
Natural:	15	(4 *ph*, 4 *den*, 3 *out*, 4 *cart*)	Natural:	15	(4 *ph*, 3 *den*, 3 *out*, 4 *cart*)
Novel:	1	(*Pompon*)			
GREEN (26)			ORANGE (22)		
Artificial:	9	(4 *obj*, 3 *sgn*, 2 *lg*)	Artificial:	9	(7 *obj*, 2 *lg*)
Natural:	17	(5 *ph*, 3 *den*, 5 *out*, 4 *cart*)	Natural:	13	(4 *ph*, 3 *den*, 3 *out*, 3 *cart*)
VIOLET (17)			GREY (14)		
Artificial:	1	(*Milko*)	Artificial:	2	(2 *obj*)
Natural:	12	(3 *ph*, 3 *den*, 3 *out*, 5 *cart*)	Natural:	12	(3 *ph*, 3 *den*, 3 *out*, 3 *cart*)
PINK (16)			WHITE (19)		
Artificial:	4	(2 *obj*, 2 *lg*)	Artificial:	2	(*obj*)
Natural:	12	(3 *ph*, 3 *den*, 3 *out*, 3 *cart*)	Natural:	17	(6 *ph*, 3 *den*, 3 *out*, 5 *cart*)
YELLOW EXAMPLES (5)			RED EXAMPLES (4)		
Artificial:	2	(1 *ph*, 1 *out*)	Artificial:	2	(1 *ph*, 1 *out*)
Natural:	2	(1 *ph*, 1 *out*)	Natural:	2	(1 *cart*, 1 *out*)
Novel:	1	(*pompon*)			

2.1.1 *Participants.* Thirty-one participants (twenty-five women and six men; average age = 25 years) were recruited by announcement and participated for financial remuneration. Four of the participants were familiarised with the novel object. In their case the novel object was placed on the desktop in their office more than two months before the experiment. All participants had normal colour vision as tested with the Ishihara plates (Ishihara 2004). All experiments in this study were carried out in accordance with the relevant institutional and national regulations and legislation and with the World Medical Association Helsinki Declaration as revised in October 2008 (http://www.wma.net/en/30publications/10policies/b3/).

2.1.2 *Apparatus.* The monitor used to display stimuli was an Iiyama MA203DT monitor driven by a NVIDIA graphics card with a colour resolution of 8 bits per channel. Spatial resolution was set to 1152 x 864 pixels and the refresh rate to 75 Hz. For calibration the spectra of the monitor primaries were measured with a Photo Research PR650 spectroradiometer. For gamma correction, primary intensities were measured with a UDT Instruments model 370 optometer with a model 265 photometric filter. These measurements were used to make look-up tables to correct for nonlinearities. Experiments were written in MatLab (MathWorks Inc 2007) with the Psychophysics toolbox extensions (Brainard 1997; Pelli 1997). The seed for the randomisation procedures in the experiment was set in relation to computer time. An

Table 3. Stimuli of reaction time experiment. The rows show the different category pairings that defined the two response options (order of columns is arbitrary). At the top of each cell the colour name is shown with the overall number of stimuli in parentheses. The different kinds of objects are listed separately (natural, artificial, or novel objects), followed by the number of the respective objects in this category. In parentheses behind this number, objects are further specified as follows. In the sample of artificial objects: *obj* = object (represents itself), *sgn* = sign (nonlinguistic symbol), *lg* = logo (written label); in the sample of natural objects: *ph* = natural photos, *den* = 'denaturalised' (without natural texture), *out* = outline shapes, *cart* = cartoons (clip-art drawings).

Category 1			Category 2		
BLUE (26)			BROWN (25)		
Artificial:	11	(2*obj*, 4 *sgn*, 5 *lg*)	Artificial:	6	(6 *obj*)
Natural:	15	(4 *ph*, 3 *den*, 4 *out*, 4 *cart*)	Natural:	19	(7 *ph*, 3 *den*, 3 *out*, 6 *cart*)
YELLOW (26)			RED (26)		
Artificial:	10	(6 *obj*, 2 *sgn*, 2 *lg*)	Artificial:	11	(5 *obj*, 5 *sgn*, 1 *lg*)
Natural:	15	(4 *ph*, 4 *den*, 3 *out*, 4 *cart*)	Natural:	15	(4 *ph*, 3 *den*, 3 *out*, 4 *cart*)
Novel:	1	(*Pompon*)			
GREEN (26)			ORANGE (22)		
Artificial:	9	(4 *obj*, 3 *sgn*, 2 *lg*)	Artificial:	9	(7 *obj*, 2 *lg*)
Natural:	17	(5 *ph*, 3 *den*, 5 *out*, 4 *cart*)	Natural:	13	(4 *ph*, 3 *den*, 3 *out*, 3 *cart*)
VIOLET (17)			GREY (14)		
Artificial:	1	(*Milko*)	Artificial:	2	(2 *obj*)
Natural:	12	(3 *ph*, 3 *den*, 3 *out*, 5 *cart*)	Natural:	12	(3 *ph*, 3 *den*, 3 *out*, 3 *cart*)
PINK (16)			WHITE (19)		
Artificial:	4	(2 *obj*, 2 *lg*)	Artificial:	2	(*obj*)
Natural:	12	(3 *ph*, 3 *den*, 3 *out*, 3 *cart*)	Natural:	17	(6 *ph*, 3 *den*, 3 *out*, 5 *cart*)
YELLOW EXAMPLES (5)			RED EXAMPLES (4)		
Artificial:	2	(1 *ph*, 1 *out*)	Artificial:	2	(1 *ph*, 1 *out*)
Natural:	2	(1 *ph*, 1 *out*)	Natural:	2	(1 *cart*, 1 *out*)
Novel:	1	(*pompon*)			

ActiveWire-compatible input device was used with an ActiveWire driver to record responses (ActiveWire Inc 2003). This was done to avoid noise in reactions times, such as the one produced by the keyboard buffer. Variability of pure time measurements was below 1 ms. During the experiment the distance to the monitor was controlled through a chin rest that was 65.5 cm away from the monitor.

2.1.3 *Stimuli.* For an overview of the composition of the stimulus set, please see table 2. We assembled a set of overall 218 images of objects, either from the Internet or by taking photos ourselves. Among these there were only sixty-five artificial objects. Two others were the colour-neutral object and the novel object. The colour-neutral object was a sock with orange and red stripes. The novel object was a yellow handmade, woollen pompon.

The other stimuli were 151 images of natural objects. This ensemble of natural images consisted of forty-six photos of original objects, thirty-two photos of objects with an artificial texture (ie, white-painted objects or reproductions in porcelain or plastic), forty-one clipart cartoons and drawings, as well as thirty-two outline shapes of some of the photos. Among these were the different versions of the eight fruits and vegetables (here summarised as 'fruits') used by Olkkonen et al (2008). These were courgette, lettuce, grapes, banana, lemon, orange, carrot, and strawberry. While there were eight natural photos and eight outline shapes of these photos, there were only seven photos of white-painted fruits because the lettuce could not be painted in white. There were also the outline shapes of the rose leaf and the donkey used by Duncker (1939).

Moreover, the following stimuli were control stimuli. Four outline shapes were barely recognisable because they were more or less round disks. These were the silhouettes of a blueberry, of the Earth, of an orange, and of a red cabbage. Three stimuli were ambiguous. Firstly, there was a standard dustbin for bus stops in Germany. Even though the dustbin in our photo was originally green, in Germany this kind of dustbin occurs in green as well as in orange. Secondly, even though our pepper was originally red, peppers also exist in yellow and green. Finally, we included the yellow Ferrari logo, since most people associate the red car with Ferrari. Six stimuli were also ambiguous, but the correct colour assignment could be inferred from the available response options. This was the case for the photo of a blue emergency light such as the ones on police cars (which also exists in yellow), green traffic lights (that are not easily distinguishable from red ones, when shown as an achromatic image), the red side of a ping-pong racket (that could as well be the black side), the grey photos of the orange and the green grapes from Olkkonen et al (that might as well depict a grapefruit or blue grapes, respectively), and, finally, a white chess piece adjusted to middle grey (that might also have been black originally).

In accordance with the criterion of approximately equal numbers within each category pair, there were twenty-six blue and twenty-five brown stimuli, twenty-six yellow and twenty-six red, twenty-six green and twenty-two orange, seventeen violet and fourteen grey, and finally sixteen pink and nineteen white stimuli. In addition to these 218 stimuli, there were five yellow and four red images for the practice trials. They consisted of photos (four), cartoons (one) and outline shapes (four) of natural (four), artificial (four), and novel (pompom) objects.

If necessary, objects were manually segmented from the background using Adobe Photoshop (Adobe Systems Inc 2005). We set the background of these objects to a homogeneous achromatic background of half the maximum monitor luminance (ie, 28 cd m^{-2}). In order to make these images achromatic, we converted them from RGB to Derrington-Krauskopf-Lennie (DKL) colour space (Derrington et al 1984; Krauskopf et al 1982; MacLeod and Boynton 1979). This space consists of one achromatic luminance and two chromatic axes. Hence, we can manipulate colour and luminance independently. We set the chromatic values of the images to zero to obtain an achromatic image. In order to prevent information about hue through luminance, we shifted the luminance distribution along the luminance axis so that the average luminance of the greyscale image was isoluminant with the background. However, some of the images would not be (or not be completely) visible if they were isoluminant with the background. The reason for this was either that these images consisted of a uniform area such as the outline shapes or that there were areas at the edge of the object that had the luminance of the average image. In these cases we shifted the luminance so that the average luminance of the images was slightly higher than the background (34 cd m^{-2}).

The images have been resized so that the ordinal relationships of the original objects were maintained. At the same time, size differences have been compressed so that small objects could still be identified easily and large objects were still parafoveal. The smallest object (photo of a tooth) subtended 1.6 deg x 2.2 deg visual angle (1.8 cm x 2.5 cm); the largest (a photo of the sky), 12.1 deg x 10.4 deg (14 cm x 12 cm). Resizing was done by a nearest neighbour method (ie, without interpolation) before colour conversion in order to avoid distortion of the colour distributions.

For the measurement of baseline reaction times we used uniform and luminance-noise disks. The colour of the uniform disks was defined by (approximately) the prototypes of the respective ten Basic Colour Terms (red, green, purple, and so on). The noise disks were produced by converting the uniform disks with the respective colour into DKL colour space. Then, the luminance dimension was set to brown noise. This implies that pixels were set to random luminance intensities with a spatial frequency of $1/f^2$. Brown noise corresponds

approximately to the spatial frequency of the luminance in the fruit images (Hansen et al 2008, page 4). As a result these noise disks had an organic appearance. The size of the disks was 1.9 deg (2.2 cm) in diameter.

2.1.4 *Procedure.* In the main task the achromatic images were presented on the screen and participants had to indicate the typical colour of the respective object. Figure S1 in the supplementary material illustrates the course of one trial of the main task. Each trial began with a black fixation point on the grey background for 1000 ms. Then one of the images was presented until a response was given. For this purpose participants had two keys available. These keys corresponded to the two Basic Colour Terms that were coupled as described above. Key assignment to the two colours of a colour category pair was randomised across participants but stayed constant across the blocks for each participant. Before each block with presentations of the grey objects there was a block with repeated presentations of the coloured disks. In this way participants got used to the assignment of the categories to the two keys so as to minimise spurious response patterns for the grey objects. After pressing a response key, the corresponding colour name was displayed for 500 ms to reinforce the coupling between colour names and response keys. Every naming block began with a countdown of three seconds in order to get participants ready at the two response keys before the first image was presented.

Before starting the experimental programme, the experimenter gave an oral overview of the experiment. Then the experiment started with standardised written instructions on the screen. The time for reading through the instructions also guaranteed that people adapted to the grey background of the experimental setting. As the very first task participants were shown the novel stimulus (the woollen pompom) in its original colour. They were instructed to memorise its colour for the main task. In order to continue, they had to assign one of the eight chromatic Basic Colour Terms to this object. Then, a short practice block began. Since the practice objects were originally red and yellow (see stimuli section or table 2), the practice block began with the presentations of yellow and red uniform and noise disks. In the practice block each of the four disks was presented just once, which resulted in an overall of only four disk presentations. After a pause the eight practice objects and the novel object were presented in grey in random order. In the main part there was such a block for each of the five category pairs. The order of these blocks was randomly permuted. Overall, there were three series of all five blocks. In this main part the four coloured disks were repeated five times each, resulting overall in twenty trials per block. After the presentation of the last stimulus of the fifteenth block the colour-neutral stimulus, the sock, was presented without transition, as if it was part of this last block. Possible answers to this stimulus depended on which category pair was tested in the last block. The complete experiment lasted about 60 min.

In view of the low number of presentations of each image (three), it was important to prevent spurious responses due to fatigue or boredom. To enhance motivation, we included feedback after each block and a hall of fame at the end of the whole experiment. For the block-wise feedback as well as for the hall of fame, a score was calculated. This score combined reaction times and accuracy rates. Feedback was presented verbally and consisted of German translations of 'fantastic' (on average about 100% correct, < 800 ms), 'excellent' (about 100% correct, < 1000 ms), 'very good' (about 95%, < 1500 ms), 'good' (about 90%, 2 s), or 'okay' (about 75%, < 10 s). If a participant performed worse than okay, he or she was asked to contact the experimenter. This, however, never happened. In the hall of fame, the ten best scores were recorded. Participants were informed about the feedback and the hall of fame in the introductory instructions. There, it was emphasised that high scores could be achieved only with maximum accuracy.

2.2 Results

For the single stimuli we calculated average reaction times and accuracy rates over the thirty-one participants, over the three blocks per participant, and in the case of the disks, over the five repetitions per block. For the coloured disks we simply averaged these measures over the disk types (noise and colour) to obtain a central tendency for all disks. In the case of the objects, however, the central tendency of these measures depends strongly on the image sampling (eg, the proportion of unrecognisable images). In order to compare the reaction times and accuracy rates for single objects with an appropriate central tendency of a group of objects, we will therefore report the median of the single measures since the median automatically excludes outlier stimuli (eg, unrecognised images). Moreover, the standard errors for each object were calculated over all data points (thirty-one participants x three blocks), and not over the means per participant. This was done because both interindividual and intraindividual variations were important for our stimulus choice, and because the averages per participant are strongly affected by the single values due to the low amount of only three repetitions (ie, blocks). For all the measures of reaction times, only correct answers were considered. Apart from that, no outliers were discarded from the analysis based on the size of reaction times. The significance level for all statistical tests was set to alpha = 0.05.

2.2.1 *Control stimuli.* Overall, reaction times and accuracy rates for the control stimuli followed the expected pattern. An overview of these results together with the predictions may be found in table 1. For the coloured disks the overall average reaction time was 399 ms, with a minimum of 307 ms for the pink uniform disks and a maximum of 531 ms for the orange uniform disks. The mean accuracy rate was 98%, with a minimum of 95% for the pink noise disk and a maximum of 100% for yellow noise disk.

The accuracy rate and the average reaction times of particularly unrecognisable outline shapes were 23% and 1277 ms for the blueberry, 80% and 1326 ms for the Earth, 90% and 1037 ms for the orange, and 78% and 1227 ms for the red cabbage, respectively. For comparison, the median of the accuracy rates and the median of the average reaction times for all thirty-two outline shapes were 95% and 721 ms, respectively. The forty-seven photos yielded medians of 98% and 613 ms for the accuracy rates and the reaction times, respectively.

For the colour-neutral stimulus (sock) average reaction time was 1256 ms with a minimum of 586 ms and a maximum of 4113 ms. Depending on the response pair of the last block, different colour categories could be assigned to the sock. Category assignments were four to blue by the four people with a blue–brown final block, nine to red in the eleven samples with a red–yellow final block, seven to orange in the seven green–orange cases, three to violet for the four violet–grey, and four to pink in the five pink–white cases.

The novel stimulus (pompom) yielded an accuracy rate of 91% and an average reaction time of 1089 ms for the twenty-seven participants, who saw it only once at the beginning of the experiment. The four people who were familiarised with this object answered with an average accuracy of 100% and an average reaction time of 873 ms. This difference between the two groups of participants was only marginally significant in a one-tailed t-test, $t(84) = 1.4$, $p = 0.079$.

Among the ambiguous images, the (green) bus-stop bin resulted in an accuracy rate of 58% and an average reaction time of 963 ms, the (red) pepper in 88% and 956 ms, and the Ferrari logo in 52% and 912 ms, respectively. The average accuracy rates and reaction times for stimuli with resolvable ambiguity were 92% and 772 ms for the blue light, 84% and 1142 ms for traffic lights, 97% and 903 ms for the ping-pong racket, 99% and 686 ms for the check figure, 98% and 600 ms for the photo of the orange, and finally 97% and 676 ms for the photo of the grapes, respectively.

Finally, the silhouettes of Duncker (1939) resulted in an average accuracy rate and reaction time of 100% and 591 ms for the donkey and 94% and 1001 ms for the rose leaf. The medians of the accuracy rates and average reaction times for the eight photos of Olkkonen et al (2008) were 98% and 612 ms. For the corresponding eight silhouettes these were 93% and 780 ms, respectively. For the seven photos of the white-painted fruits these medians were 99% and 656 ms. We compared reaction times between the three formats (photo, white painted, outline shape) through t-tests. For this purpose, reaction times were averaged across participants, and t-tests were paired by objects and one tailed. The lettuce did not exist in the white-painted format. Thus, comparisons that involved the photos of white-painted fruits were restricted to seven objects, excluding the lettuce. The reaction-time differences between the seven natural and the seven white-painted fruits were on average 44 ms and significantly larger than zero, $t(6) = 2.3$, $p = 0.03$. The average difference between white-painted fruits and outline shapes was 104 ms and marginally significant, $t(6) = 1.8$, $p = 0.06$. The average difference between the eight natural photos and the eight outline shapes was 156 ms and highly significant, $t(7) = 3.1$, $p < 0.01$. Olkkonen et al used only fifteen of these stimuli in the second part of their first experiment. For this sample the five original fruits yielded 600 ms, the five white-painted fruits 660 ms, and the five outline shapes 719 ms. The correlation between the reaction times and the memory colour effects of these fifteen stimuli was $r = -0.52$ and with $p = 0.047$ significant (see figure 1). Memory colour effects are measured through a memory colour index (MCI), which is explained in Olkkonen et al (2008, page 6) as well as in the second part of this paper on the colour adjustment experiment.
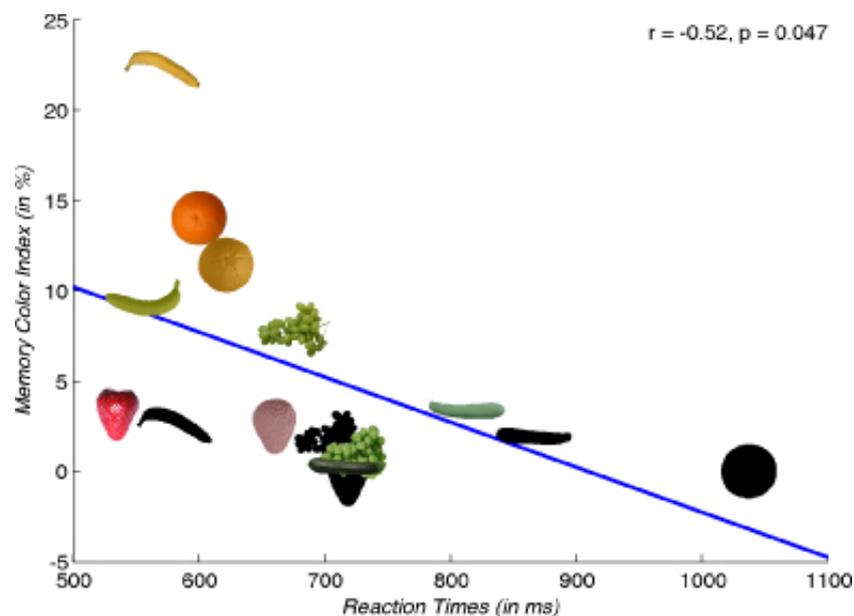
Figure 1. Correlation between reaction times and memory colour effects for the stimuli used in the second part of the first experiment in Olkkonen et al (2008). The x-axis shows the reaction times measured as indices for colour diagnosticity and recognisability. The y-axis shows the memory colour effect as measured through the memory colour index.

2.2.2 *Candidate stimuli.* The central tendency of the reaction times depends on the category pairing—for example, the saliency of the categories. For this reason we will report the median of the set of stimuli within a category pair as a reference for the reaction times of the single objects. We will highlight the three 'best' stimuli. These are those stimuli that yielded the lowest reaction times among the candidate artificial stimuli of each category. Moreover, we verified whether the reaction times discriminate well between the candidate objects. For

this purpose we compared the last of the three best stimuli (in terms of reaction times) of each category with the worst among the stimuli of the whole category pair. The 'worst' is the stimulus, which yielded highest reaction times among those that have not been excluded on the basis of the 95% accuracy rate criterion. For this comparison we calculated a one-tailed t-test.

For the overall fifty-one stimuli of the blue–brown pair the median reaction time was 614 ms (see figure 2a). The images of a blue light (92%), of Alf (94%), of a pharmacy vial (80%), and of a brown pair of sunglasses (59%) led to less than 95% correct responses. For the blue category a Nivea cream tin (574 ms), a blue traffic sign (591 ms), and the smurf (633 ms) yielded the three lowest average reaction times. They also resulted in the smallest standard errors (14 ms, 18 ms, and 20 ms, respectively). For brown the chair (559 ms, 22 ms), the closet (583 ms, 22 ms), and the violin (608 ms, 23 ms) led to lowest average reaction times and standard errors. The worst stimulus with an accuracy rate greater than 95% was the (blue) O2 logo. Its average reaction time (873 ms) was significantly different from the one of the smurf, $t(177) = 4.0$, $p < 0.01$, as well as from the one of the closet, $t(177) = 4.9$, $p < 0.01$.

For the overall fifty-two stimuli of the yellow–red pair the median reaction time was 590 ms (see figure 2b). The McDonalds logo (86%), the priority street sign (94%), the radioactivity sign (91%), the Ferrari logo (52%), the Reclam booklet (89%), and the German train company logo (92%) resulted in accuracy rates below 95%. For yellow the German mailbox (572 ms), the UHU glue tube (598 ms), and the tennis ball (599 ms) elicited the lowest average reaction times for artificial objects. These stimuli had the third (20 ms), second (18 ms), and seventh (23 ms) highest standard errors among the yellow artificial objects. For red, reaction times for the Coke logo (575 ms), a schematic heart (576 ms), and the photo of a fire extinguisher (590 ms) were lowest. Standard errors were fifth (16 ms), first (14 ms), and third highest (14 ms), respectively, among the red artificial objects. The (red) ping-pong racket (s. stimuli) yielded the worst average reaction time (903 ms) of the yellow–red pair. It differed significantly from the reaction times for the tennis ball, $t(181) = 5.2$, $p < 0.01$, as well as for the fire extinguisher, $t(180) = 5.8$, $p < 0.01$.

The median reaction time for the forty-eight stimuli of the green–orange pair was 710 ms (see figure 2c). The ewer (94%), the exit sign (89%), the wine bottle (84%), the bus-stop bin (58%), the green traffic lights (84%), Ernie from *Sesame Street* (92%), and the Elmex toothpaste (76%) elicited less than 95% correct answers. For green the ping-pong table (715 ms), the German police car door (748 ms), and the recycling symbol (781 ms) yielded the lowest average reaction times. They also resulted in the three lowest standard errors (23 ms, 28 ms, and 26 ms, respectively). For orange this was the case for Die Maus (The Mouse, a German TV figure, 608 ms), the basket ball (654 ms), and the warning vest (676 ms). Standard errors were also the three smallest (15 ms, 17 ms, and 13 ms, respectively). The worst of the artificial objects, with an accuracy rate above 95%, was the (green) European Bio symbol (828 ms). Among the best green stimuli, only the reaction times for the ping-pong table differed significantly from the one of the Bio symbol, $t(182) = 2.2$, $p = 0.02$. For orange, however, the difference between the warning vest and the Bio symbol was significant, $t(179) = 2.9$, $p < 0.01$.

The median reaction time for the total of thirty-one stimuli of the violet–grey pair was 660 ms (see figure 2d). The Milka logo (92%) and the grey dustbin (92%) yielded accuracy rates of less than 95%. For violet the Milka logo was the only artificial object and led to an average reaction time of 740 ms. For grey only the lantern remained with a reaction time of 725 ms. For the thirty-five stimuli of the pink–white pair the median reaction time was 628 ms. Among the six artificial objects, only the €10 bill (90%) had to be excluded by the 95% criterion. The remaining three pink artificial objects were the Pink Panther, the Barbie puppet logo, and the German Telekom logo. Their mean reaction times and standard errors were

596 ms, 773 ms, and 792 ms, as well as 19 ms, 48 ms, and 22 ms, respectively. The difference between the mean reaction times of the Pink Panther and the Barbie logo was significant, $t(181) = 2.9$, $p < 0.01$. Among the stimuli of the achromatic categories (grey and white) the golf ball led to the smallest average reaction time (609 ms) and to the smallest standard error (19 ms).

2.3 Discussion

As the first step, we wanted to verify whether this experimental paradigm is useful to measure subjective colour diagnosticity of objects and the recognisability of their images. If the paradigm works, we intended to identify those images that have the highest potential to elicit a memory colour effect according to these measures. This should allow the selection of the most appropriate stimuli for the colour adjustment experiment.

2.3.1 *Evaluation of paradigm.* We wanted to verify three assumptions about this paradigm. Firstly, accuracy rates were supposed to identify those stimuli that are unrecognisable or that lack any colour diagnosticity. Secondly, we wanted to guarantee that the information of the memorised typical colour was the main factor that determines our dependent variables. In particular we wanted to exclude that information of the luminance distribution influences our measurements. And, finally, reaction times should be used to measure the automaticity of the colour–object association. For this purpose they should show sensitivity to the degree of automaticity of the answers. Table 1 facilitates the comparison between predictions and results.

In general, accuracy rates were lowered for those stimuli we judged as difficult to recognise or ambiguous in their typical colour (mostly < 90%). In contrast, ambiguous objects whose colour could be inferred because of the available answer options led to higher accuracy rates (mostly > 95%). However, not all of these stimuli yielded accuracy rates of approximately 50% (eg, between 25% and 75%), as was the case for the bus-stop bin and the Ferrari logo. For example, the silhouette of the orange led to responses slightly above 90%, despite its very unspecific form. Moreover, the outline shapes in general did result in a quite high accuracy rate (95%) in comparison with the photos (98%). This probably had to do with the fact that the silhouettes of objects corresponded in size and shape to the photos of the same objects. Finally, the error rates for categorising the coloured disks (between 0% and 5%) indicate that there might be up to 5% of spurious errors. Taken together, we may conclude that the limit of 5% errors is very sensible to exclude completely inappropriate stimuli without excluding appropriate images due to spurious errors.

In order to prevent the influence of luminance information on the responses, the areas of the typical colours of the stimuli were isoluminant with the background. Nevertheless, we observed that accuracy rates and reaction times were influenced by information beyond that of recognisability and colour diagnosticity. Some ambiguous stimuli such as the pepper (88%) yielded accuracy rates that were much higher than chance (50%). Moreover, some of the stimuli that were expected to necessitate explicit inference led to comparatively low reaction times, such as the photos of the orange (600 ms), of the green grapes (676 ms), and of the white chess piece (686 ms). In some cases this might be due to the fact that some of the original exemplars of the ambiguous stimuli were more typical than the alternative variant, such as the red versus the yellow pepper, or the orange versus the grapefruit. However, the responses to the colour-neutral sock also seemed not to be equally distributed over the category pairs: the sock was assigned more often to its original colours than to the alternatives if the original colours were among the category pairs (82% red instead of yellow and 100% orange instead of green). This might imply that the luminance distribution provides some relevant colour information. However, it might also be the case that participants simply knew this kind of sock or associated very intense colours like red and orange with the striped
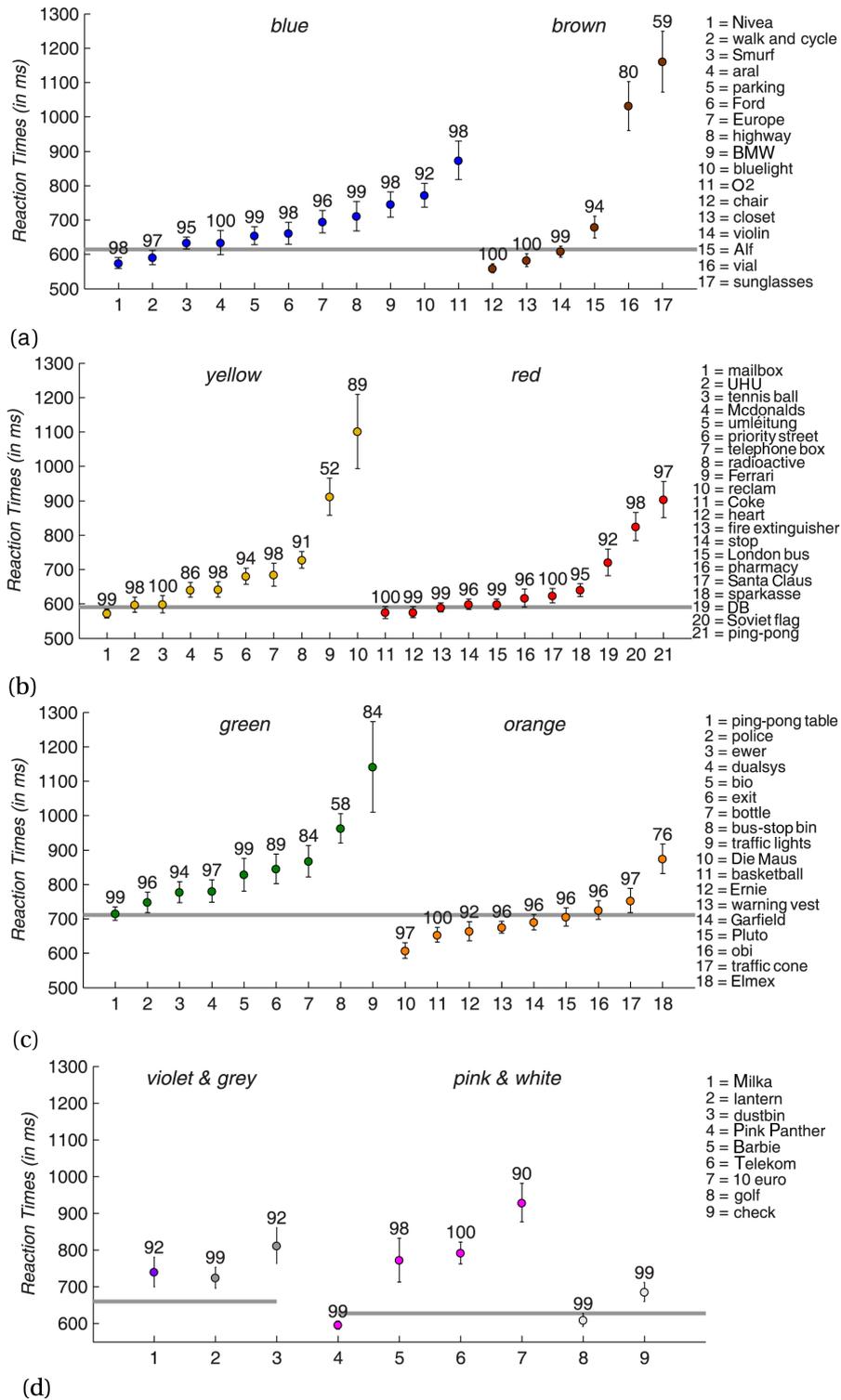
Figure 2. Reaction times and accuracy rates for candidate objects. The different parts of the figure correspond to the category pairings (cf table 2). The *x*-axis lists the artificial objects of each category in the order of their average reactions times. The *y*-axis represents reaction times in ms for correct responses. Coloured disks show reaction times averaged over thirty-one participants; the respective error bars depict standard errors of mean. The grey line corresponds to the median reaction time of the whole category pair, that is, including the natural objects. Category pairs are: (a) blue–brown, (b) yellow–red, (c) green–orange, and (d) violet–grey and pink–white.

pattern of the sock. In sum, we cannot completely exclude that the luminance distributions provide information about the original colours.

Reaction times differentiated well between different levels of familiarity. Through the maximum reaction times for coloured disks (531 ms) we determined a lower boundary for reaction times at about 550 ms. In contrast, participants reacted only after long thinking times (> 1 s) to the colour-neutral sock, indicating that its appearance perplexed them. Additionally, reaction times for those stimuli we had previously judged as unrecognisable also exceeded 1 s. Therefore, the region around 1000 ms may be a reasonable upper boundary to evaluate the degree of the participant's familiarity with the object–colour association. Within this range from 550 ms to 1000 ms we could dissociate automatic from explicitly deliberated responses—with some limitations. Stimuli to which participants could respond correctly through explicit inference and recall should result in comparatively high reaction times with moderate error rates. This was the case for ambiguous stimuli such as the blue lights (772 ms and 92%) and the ping-pong bat (903 ms and 97%), as well as for the woollen novel object (1.089 ms and 91%). Moreover, the reaction times to the novel object showed the right tendency to differentiate between the untrained group and the group that was familiarised with the real object (873 ms). However, the difference between the two groups was only marginally significant ($p = 0.079$). But not all predictions could be confirmed. Some of the ambiguous stimuli yielded reaction times that were close to or even lower than the median such as the photo of the orange (600 ms) or of the grapes (676 ms). Moreover, the median reaction times could not provide a reliable border to separate automaticity from explicit inference since the medians varied quite considerably across colour category pairs. The minimum was 594 ms for the yellow–red pair and the maximum 717 ms for the green–orange pair. This difference could not be explained by different numbers of stimuli since these pairs contained an approximately equal number of stimuli ($n = 52$ and 48). Nevertheless, we still found within each category pair significant differences between fast and slow stimuli. 'Fast stimuli' can be identified as those for which the difference to the stimulus with the largest reaction time was significant. This indicates that a meaningful division between stimuli with low and high reaction times is possible. Among these fast stimuli the lowest reaction times varied between 559 ms (brown chair) and 715 ms (green ping-pong table). The fact that the fastest reaction times are close to the lower reaction time border (550 ms) suggests that the knowledge about the respective object–colour association is automatised.

Furthermore, we found for the fifteen stimuli of Olkkonen et al (2008, second part of first experiment) that the reaction times correlated with the memory colour effect. This shows that the characteristics of the stimuli measured by this paradigm play a role in the memory colour effect. However, the total variance explained by this correlation is only 27%. Moreover, a main part of this correlation seems to be due to the fact that the outline shapes were less recognisable. Hence, they yielded not only higher error rates, but also higher reaction times as well as lower memory colour effects (see figure 1). In fact, the order of reaction times resembled the inverse order of the memory colour effects for outline shapes, photos of white-painted fruits, and photos of original fruits: the five outline shapes had highest reaction times (719 ms) and lowest memory colour effects (1.2%), the five photos of natural fruits lowest reaction times (600 ms) and highest memory colour effects (9.5%), and the others were in between (660 ms and 5.6%). Furthermore, Duncker's (1939) silhouette of a rose leaf seemed also to be poorly recognisable. Its accuracy rate (94%) and reaction time (1001 ms) were beyond the borders for accuracy rate and reaction time defined above (< 95%, > 1000 ms). Therefore, the relationship between reaction times in our experiment and memory colour effects support the idea of Olkkonen et al (2008) that the unreliable results of

earlier studies on the memory colour effect were due to the low spontaneous recognisability of the outline shapes.

In sum, we may conclude that our paradigm provides a sensible measure of colour diagnosticity. This measure allows the identification of those stimuli that hold a potential to elicit memory colour effects. More precisely, the paradigm successfully separates stimuli with a particularly high degree of automatised object–colour association from those with only uncertain or deliberative object–colour association. And at this level of resolution the reaction times may predict memory colour effects: objects with particularly low reaction times, such as the photo of the banana, yielded high memory colour effects; objects with extremely high reaction times, such as the outline shape of the orange, led to low memory colour effects. Therefore, the paradigm may separate stimuli with a particularly high potential for memory colour effects from those with a particularly low potential for memory colour effects. It is no surprise, however, that the relationship between reaction times and memory colour effects does not resolve into finer nuances. The precise variation of both reaction times and memory colour effects may be influenced by factors other than colour diagnosticity. On the one hand, factors such as the objects' absolute size, the saliency of their shapes, as well as luminance contrast may influence the size of reaction times. On the other hand, factors such as overall saturation, the saliency of colour as an object feature, and—as we will see later—the hue itself may have an impact on the size of memory colour effects.

Note that these and other particularities of the images in our stimulus pool also limit a general interpretation of the reaction times. For example, among the artificial objects we found particularly low reaction times for the chair (559 ms), the German mailbox (572 ms), the Nivea tin (574 ms), and the Coke logo (575 ms). Among the natural objects there were still lower reaction times, in particular for the clipart cartoon of a banana (520 ms) and a strawberry (520 ms). One might wonder whether the association between object and colour is particularly strong for these objects. Furthermore, the median reaction time across objects was comparatively high for the green, violet, and pink category, namely 735 ms, 722 ms, and 709 ms, respectively. In contrast, it was not only the median reaction times of the achromatic categories grey (564 ms) and white (584 ms) that were particularly low; those of the yellow, red, and brown categories were also comparatively low with 585 ms, 599 ms, and 608 ms, respectively. Here, one might wonder whether these striking differences were due to the nature of the involved colour categories. However, our findings cannot yet be generalised neither for the depicted objects nor for the colour categories. To answer the two questions above a procedure that objectively controls for comparable object images and stimulus sets in all categories is required.

2.3.2 *Stimulus selection.* According to the average reaction times, green stimuli were less appropriate to verify memory colours. In contrast to other colour category pairs, reaction times for the best green stimuli were higher than 700 ms. And only one of them, the ping-pong table, led to significantly lower reaction times than the image with highest reaction times. For the orange stimuli reaction times were also quite high (> 650 ms), except for the best stimulus (608 ms). Moreover, there were already several green and orange stimuli in the study by Olkkonen et al (2008). The inverse was true for blue and red stimuli. For this reason we decided to include only one stimulus of the green and orange categories and—in turn—three of the blue and red ones. For pink there was also only one stimulus that yielded a reaction time of under 700 ms, and this was significantly lower than the average reaction time for the next higher stimulus. The violet stimulus would usually be excluded. But, since Olkkonen et al did not have any violet stimulus and since it was the only one we had for violet, we included it nevertheless. From the group of brown and yellow stimuli we took

two. Finally, we chose the most colour diagnostic among the grey and white stimuli and the colour-neutral stimulus as control stimuli.

In this way the following sixteen stimuli were chosen for the colour adjustment experiment:

- the Nivea tin, the blue traffic sign, and the smurf for the blue category (3),
- the chair and the closet for the brown category (2),
- the German mailbox and the UHU glue stick for the yellow category (2),
- the Coke logo, the schematic heart, and the fire extinguisher for the red category (3),
- the German TV figure Die Maus for the orange category (1),
- the ping-pong table for the green category (1),
- the Pink Panther for the pink category (1),
- the Milka chocolate bar for the violet category (1),
- the golf ball as the achromatic control stimulus (1), and
- the sock as the colour-neutral control stimulus (1).

With the exception of the green and violet categories, these stimuli yielded reaction times between 550 ms and 650 ms. In this way they were close to the lower (550 ms) and far from the upper boundary (1000 ms) for reaction times as defined above.

3 Experiment 2: investigation of memory colour effects

This experiment investigated whether knowledge about object colour modulates the colour appearance of the highly colour diagnostic artificial objects determined in the previous experiment. For this purpose we used the same methods as Olkkonen et al (2008). In order to measure colour appearance we let our participants adjust the colour of the objects. We tested the same hypothesis for these manmade objects as Olkkonen et al (2008) did for the fruits: if memory colour influences the appearance of the actual colour, the objects should appear slightly in their typical colour when they are achromatic, ie grey. Therefore, when adjusting the object to grey, the participants should shift the colour toward the colour opponent to the object colour. In this way they would counteract their subjective impression of apparent object colour at the grey point. So, in order to measure the main dependent variable we let participants adjust the colour of each object to achromatic grey. To determine the memory colour that people have in mind for each object, we also let people adjust the colours of the objects to their typical colours. Moreover, since there might be systematic distortions of the colour appearance of the objective grey, we also measured observers' subjective grey point. Therefore, we let observers adjust disks with either a uniform colour or a brown noise $(1/f^2)$ luminance distribution. These subjective grey points represent the actual colour that the respective observer perceives as achromatic when there is no interference of memory colours. For this reason the subjective grey point of each individual observer separately was used to evaluate memory colour effects.

As announced in the introduction, the artificial objects used in this experiment varied in the complexity of their perceptual features as well as in the abstractness of their colour diagnostic characteristics (see figure 3). On the one hand, the perceptual complexity varied on three levels. The most complex stimuli were photos of three-dimensional objects. These stimuli had a surface texture, and provided complex colour distributions, ie colour distributions that contained a multitude of different hue and saturation nuances. The stimuli in this class were the red fire extinguisher, the brown chair and closet, the green ping-pong table, the yellow mailbox and UHU glue stick, as well as the blue Nivea tin. Then, there were also two-dimensional stimuli with complex colour distributions, namely the violet Milka chocolate bar picture and the red Coca Cola logo. Finally, two-dimensional images that were composed of only uniform colour surfaces were considered as perceptually least complex.

These included the Pink Panther, the orange Die Maus, the red heart, as well as the blue smurf and traffic sign (the latter was not shown as a realistic photo, but as an icon). On the other hand, the degree of abstractness also varied on three levels. Firstly, there were concrete objects that directly represented themselves. Among these were the red fire extinguisher and heart, the brown chair and closet, the green ping-pong table, the Pink Panther, the orange Die Maus, as well as the blue smurf. In contrast, there were also objects whose characteristic features were symbols or writings, such as a traffic sign or a brand logo. These features must be interpreted symbolically and refer to abstract ideas such as a traffic rule or a company identity, respectively. Among the symbols, secondly, there was the blue traffic sign and the yellow mailbox. Thirdly, logos with written names must even be interpreted linguistically before they may refer to a name, which by itself refers to a company. To this last group belonged the yellow UHU glue stick, the blue Nivea tin, the violet Milka chocolate bar, and the red Coca Cola logo. This selection of stimuli allows us to test whether memory colour effects may also be elicited by two-dimensional objects with simple colour distributions. And we can verify whether these effects also occur for stimuli whose object identity is defined through abstract symbolic associations.

In addition, we included the aforementioned two control objects to verify whether there were systematic distortions of colour adjustments due to the presentation of objects in general. Firstly, we added the golf ball as an object that is naturally achromatic. This stimulus was used to verify whether there are systematic deviations of colour adjustments independently of memory colours for objects with natural colour and luminance distributions. Secondly, we also used the sock with orange–red stripes. A sock is not colour diagnostic since it may exist in different colours. We showed this sock in its original colour only once at the beginning of the whole experiment. It might be that there are influences of seeing the objects in the original colour before achromatic adjustment. If this was the case the sock should reveal memory colour effects even though its colour had not been internalised through experience.
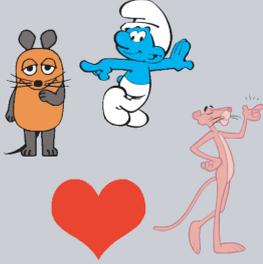
A memory colour effect is shown if the achromatic adjustments are shifted away from the subjective grey point towards the colour opponent to the object's typical colour. If the memory colour effect is only due to the learned object–colour association, it should appear for all fourteen objects in the same way, independently of their perceptual complexity and the abstractness of their characteristics. Finally, if all assumptions about the memory colour effect apply, both of the control objects (golf ball and sock) should not be shifted systematically away from the subjective grey point.

3.1 Method
Overall, the method and setup of this experiment were the same as in Olkkonen et al (2008). Details may be found there. Here, we focus on some slight modifications to this method, which were applied to improve methodological stringency. In addition, there was one overall difference to the design of Olkkonen et al (2008) in that we had fewer repetitions per object. This was due to the fact that we had more objects (sixteen instead of eight), and that we did not want to extend the experiment to more than one hour. To compensate for the smaller amount of data per participant, we recruited more participants.

3.1.1 *Participants.* In this experiment, twenty-five of the thirty-one participants of the first experiment participated for financial remuneration. This sample was composed of twenty-three women and two men; average age was 26 years. All participants were naive as to the purpose of the experiment.

3.1.2 *Apparatus.* This experiment took place in the same experimental chamber with controlled illumination as used by Olkkonen et al (2008). This setup differs from the one used for the reaction time experiment (see above). The setup had been recalibrated for

	3-dimensional	*2-dimensional*	*2D with uniform colour patches*
Object			
Symbol			
Logo			

Figure 3. Features of the fourteen colour diagnostic stimuli used in the colour adjustment experiment. Columns show how the stimuli vary in the complexity of their perceptual features. Stimuli on the right depict three-dimensional objects with a texture and a complex colour distribution. Stimuli on the left are two dimensional and consist merely of uniformly coloured areas. The two stimuli in the centre column are two dimensional, but have a texture and a complex colour distribution. Rows represent the degree of abstractness of the identificatory feature. Stimuli in the top row are objects. This implies that they represent themselves. The main features of the stimuli in the centre and in the bottom row are symbols or writings, respectively. They have to be interpreted symbolically. Moreover, they refer to abstract ideas such as a traffic rule in the case of the traffic sign or a company identity in the case of the brand logos.

this experiment. The resulting Judd-revised CIE chromaticity coordinates for the monitor primaries were now: $R = (0.615\ 0.348\ 20.6)$, $G = (0.283\ 0.606\ 59.6)$, and $B = (0.156\ 0.083\ 8.3)$. The look-up tables for gamma correction were also updated. Lamps were recalibrated by letting four observers adjust the lightness and hue of the lamps to be equal to the one of the grey background of the monitor. The measured Judd-corrected chromaticity coordinates for the lamps were $x = 0.302$, $y = 0.358$, and $Y = 42.2$. The monitor was placed at the end of a 60 cm-long tunnel, which opened out into the experimental chamber. This tunnel prevented reflections of the illumination on the monitor. The chamber was 125 cm long. Hence, the distance between chin rest and monitor was 185 cm. For randomisation the seed was set in relation to computer time.

3.1.3 *Stimuli.* Besides the stimuli selected by the preliminary experiment, we added three stimuli for practice trials: a yellow rubber duck, a lifesaver with red stripes, and a blue *Aral* logo. The first two were part of the practice trials and the last was part of the main stimuli of the preliminary reaction time experiment. The uniform and the noise disk used to measure the subjective grey point were constructed in the same way as in the preliminary experiment. The background luminance was set to half the maximum luminance of the monitor. The Judd-revised chromaticity coordinates of the background were $x = 0.311$, $y = 0.344$, $Y = 44.2$.

As described in Olkkonen et al (2008), the original RGB images were converted to the colour-opponent DKL space (cf experiment 1). This allowed the standardisation of luminance and implementation of colour adjustments during the experiment. Also, the colour adjustment was made only for the chromatic dimensions while holding luminance constant. Some of the objects, however, might contain more than one colour. For example, the blue smurf has a white hat and white trousers. With the fruits in Olkkonen et al (2008) this problem appeared only for the strawberry. In this case the authors could remove the green part of the strawberry without making it unrecognisable. A corresponding treatment, however, would completely destroy the identity of some of the artificial objects. Therefore, we converted the colour distribution of such parts to achromatic grey. During the colour adjustment of the characteristic colour distribution, these parts did not change. This procedure was used for the following seven objects: the smurf, the traffic sign, the Nivea tin, the ping-pong table, Die Maus, the fire extinguisher, and the Milka chocolate bar. Figure S1 in the supplementary material provides an overview of the areas that were kept achromatic during the colour adjustments as well as an overview of the appearance of the respective objects at the average typical and achromatic adjustments. Only for Die Maus, the fire extinguisher, and the Milka chocolate bar did these manipulations concern chromatic areas at all. The colourful area that was held constant in the Milka chocolate bar was a tiny cow bell. For the fire extinguisher it concerned a small yellow button that could have been grey as well, ie that was not colour diagnostic. For Die Maus this manipulation turned the brown arms, legs, and ears into grey, which also did not strongly change the appearance of this object. For the other objects the concerned areas were originally achromatic. For example, this was the case for the text and handle of the fire extinguisher, the Nivea logo, and the smurf's clothes. The only reason for controlling these was that in the digital images of the objects the respective areas may not be completely achromatic. Instead, they may have slight shades of colours due to the overall colour of, or due to shadings within, the whole image. During the colour adjustment this might have led to artefacts when participants oversaturate the stimulus. In sum, we may assume that keeping these areas achromatic during the adjustment did not strongly affect the appearance of these seven objects.

This treatment, however, leads to a dilemma. Olkkonen et al (2008) set the mean luminance of each complete image to the luminance of the background. If we do this for example with the smurf, the blue areas would be shifted towards a luminance that is darker than the original one. This is due to the large white parts of the smurf. This treatment would provide a completely distorted luminance cue to the beholder. Moreover, it would have shifted some of the colours outside the monitor gamut by increasing their luminance. Hence we decided to control only the luminance of the area with the typical colour, which is adjusted during the experiment. If we set the mean luminance of this area to the luminance of the background, however, this would blur many of the objects with colour distributions (eg Nivea tin, ping-pong table). It would also render uniform colour areas completely equal to the background. In this way the colour adjustment task would become a discrimination task for images such as the smurf, the traffic sign, or the heart. For these reasons we decided to keep the polarity (+/−) of the difference between the luminance of the background and the original average luminance of the typical colour. As a result, a colour diagnostic area that was originally lighter than the background would also be lighter in the experiment, and vice versa. We just limited the difference to a standardised luminance shift of 8.8 cd m^{-2} upwards or downwards from the background luminance. The average luminance of the following stimuli was set to the luminance of 35.4 cd m^{-2} (background luminance −8.8 cd m^{-2}): the blue traffic sign, the blue Nivea tin, the violet Milka chocolate bar, the red Coca Cola logo, the red heart, the red fire extinguisher, the green ping-pong table, and the red–orange sock. The average

luminance of the typical colour area of the blue smurf, the Pink Panther, and the orange Die Maus was set to 53.1 cd m⁻² (background luminance +8.8 cd m⁻²). The average luminance of the typically coloured areas of the yellow UHU glue stick, the yellow German mailbox, the brown chair, the brown closet, and the white golf ball could be set to the luminance of the background. For these images this did not lead to the aforementioned problems because they contained strong luminance contrasts due to shadows. Finally, the two kinds of disks were set to +8.8 cd m⁻² above the background luminance.

Like in the first experiment, the size of the images was adjusted to keep the relative order of sizes of the original objects, but size differences were squeezed. The sizes of the images varied on this setup between 1.24 deg x 1.24 deg visual angle (4 cm diameter) for the golf ball and 4.48 deg x 2.48 deg visual angle (14.5 cm x 8 cm) for the closet. The disks were set to 1.70 deg (5.5 cm) in diameter. Finally, during the whole experiment the lamps in the chamber were set to approximately the same lightness and hue as the grey monitor background as described in the 'apparatus' section.

3.1.4 *Procedure.* In one trial of the main task the image of an object appeared in a random colour on the screen. Participants were instructed to adjust its colour either to the object's typical colour or to grey, depending on the experimental condition. In contrast to Olkkonen et al (2008) this random colour was drawn from a circle in the colour space. The radius of this circle corresponds to the saturation, the azimuth to the hue. Colours were sampled so that they had the same hue distance to the adjacent colours. Hence, the azimuth distance between adjacent hues was 360° divided by the number of stimuli. This number was sixteen for the typical and eighteen for the achromatic adjustments, which included the two disks. This was done in order to prevent colours from being drawn from particular parts of the colour space by chance. In order to randomise the colours, this circle was turned by a random azimuth for each participant, and colours were assigned randomly to each object. As a result, the colour space was sampled homogeneously for each participant and colours were still random. This hue circle was located at approximately half of the maximum saturation. In this way, participants were forced to make adjustments in every trial of both conditions because the start colours were too unsaturated for the typical and too colourful for the achromatic colours.

In order to adjust the colours of the objects, participants used the number pad of the keyboard. Here, the keys '2', '4', '6', and '8' corresponded to the poles of the chromatic DKL axes. The poles of the axes were assigned to these keys so that the keys reflect the opponent logic of the colours: '4' corresponded to the greenish and '6' to the opponent reddish pole; '2' could be pressed to make the image more bluish and '8' to shift the hue towards the opponent yellowish colour. As the adjustment approached the grey point, ie the origin of the colour space, the colour distribution was compressed, as described in detail by Olkkonen et al (2008). In this way, images were in fact completely achromatic at the objective grey point. Participants could switch between a coarse (key '9') and a fine (key '3') adjustment step. At the beginning of each trial the coarse step adjustment was turned on. After finishing the adjustment of a particular object, participants had to press the space bar to confirm the adjusted colour and to proceed with the presentation of the next stimulus. However, in order to guarantee that participants were fine-tuning their adjustments before continuing with the next stimulus, participants could confirm and end the respective adjustment only if they were in fine-tuning mode. If a participant got stuck with his or her adjustment, he or she could press the key '1' to reset the object to the start colour. Finally, owing to the limitations of the monitor gamut objects cannot become infinitely saturated. If a participant passed with more than 20% of the pixels over the gamut, an announcement appeared. The participant had to shift the colour distribution back into the gamut before continuing any adjustment.

Therefore, the adjustments were blocked for all colour directions except the direction that was opponent to the direction in which the gamut had been transgressed.

After an oral introduction to the experiment, instructions were presented on the screen so that participants adapted to the illumination. Before the main task we presented the colour-neutral object (the sock), in its original colours (red with orange stripes), and asked participants to memorise this colour in order to adjust it afterwards. They had to press space to end the presentation and to continue with the instructions. Then the experiment proceeded with the part with either the typical or the achromatic adjustments. The order of these conditions was randomised. Each of these parts began with practice trials. After a pause, the main blocks began. In one block all stimuli were adjusted once in random order. There were three blocks per condition. In the achromatic condition the main as well as the practice trials also included the adjustment of the uniform and the noise disk. The complete experiment lasted on average 56 min.

3.2 Results

For the colour adjustments of the objects, some single measurements appeared to be inconsistent with the aim of the task. For example, in the achromatic conditions some stimuli have been adjusted to their typical colour, and vice versa. We discarded outliers of this kind through the following criterion. For each object we lumped all measurements together, ie all repetitions across all participants. Then, we calculated the average and the standard deviation of these adjustments along either the x-axis or the y-axis. For each object we discarded adjustments that deviated from the average along the x-axis or along the y-axis (or both) by two standard deviations. For comparisons of averages, t-tests will be used. If not mentioned otherwise, they will test across participants. Note that in some such t-tests the degrees of freedom may be only twenty three instead of twenty four for twenty-five participants. This was the case when for one of the participants all three adjustments of a particular object were outliers in regard to the criterion above. However, neither the inclusion of such outliers nor the exclusion of the participant with the highest amount of outliers changed the pattern of the results in any way. Alpha was set to 0.05.

3.2.1 *Typical adjustments.* The average typical adjustments are shown in figure 4a. By the criterion of two standard deviations, 8% of the typical adjustments were excluded. The interindividual standard errors averaged over the colour diagnostic objects were 0.035 and 0.036 for the x-axis and y-axis (of the DKL space), respectively. In comparison, the standard errors of the typical adjustments of the golf ball were 0.018 and 0.028, and of the sock 0.031 and 0.035, along the x-axis and y-axis, respectively.

3.2.2 *Achromatic adjustments.* The average achromatic adjustments are shown in figure 4b. Of the achromatic adjustments, we discarded 8% as outliers. Interindividual standard errors averaged over colour diagnostic objects were 0.009 and 0.014 for the x-axis and y-axis, respectively. Achromatic adjustments were aligned towards a particular direction so that x and y values were significantly positively correlated, $r = 0.80$, $p < 0.01$, $n = 18$. On the basis on the covariance, a principal component with a slope of 72 deg. could explain 96% of the variance of these adjustments.

We determined the subjective grey point of each participant by averaging the adjustments for the uniform and noise disks. When averaged over all participants, we observed that the subjective grey points tended to be shifted towards blue, $x = -0.015$, $y = -0.020$, azimuth = 232 deg. This shift was significant on the x-axis, $t(24) = -3.0$, $p < 0.01$, and the y-axis, $t(24) = -2.4$, $p = 0.03$. The achromatic adjustments of the grey-shaded golf ball and the colour-neutral sock were shifted in approximately the same direction as the subjective grey point (azimuths of 214 deg. and 225 deg., respectively). The average adjustments of the white golf ball were

−0.019 and −0.012 with interindividual standard errors of 0.012 and 0.021 on the x-axis and y-axis, respectively. The difference of these adjustments to the subjective grey point was not significant in a paired t-test, neither for the x-axis, $t(23) = 0.25$, $p = 0.81$, nor for the y-axis, $t(23) = 0.41$, $p = 0.69$. The average adjustment of the colour-neutral sock was −0.018 and −0.018 with standard errors of 0.010 and 0.019 on the x-axis and y-axis, respectively. And these adjustments were not significantly different from the subjective grey point, $t(23) = 0.065$, $p = 0.95$ on the x-axis and $t(23) = −0.22$, $p = 0.83$ on the y-axis.

When removing the control stimuli (golf ball, sock, and disks), the correlation of x and y values ($r = 0.76$) as well as the direction of the principal component stays almost identical (azimuth = 72 deg., variance explained = 96%). We calculated a paired t-test over the average adjustments per participants in order to compare whether the adjustments deviated significantly from the subjective grey point on the x-axis or the y-axis. With *Bonferroni* adjustment (alpha = 0.025), the achromatic adjustments of the heart, the ping-pong table, the smurf, the blue traffic sign, and the Nivea tin differed significantly on both axes from the adjustments of the subjective grey. For the Coca Cola, the chair, the mailbox, the UHU stick, and the Milka bar this was true only for the y-axis. For the Pink Panther the difference along the x-axis was almost significant ($p = 0.026$). The adjustments of the fire extinguisher, Die Maus, and the closet did not differ significantly along any axis (all $p > 0.05$). The shifts along the axes, however, do not reflect whether they occurred in the opponent direction of the typical colour of the object. To get a first impression, the comparison of the achromatic adjustments in figure 4b with the typical adjustments in figure 4a allow for a rough graphical evaluation of this question.

3.2.3 *Memory colour index.* In order to evaluate how much the achromatic adjustments were shifted to the opposite colour of the typical adjustments, we calculated the MCI proposed by Hansen et al (2006, page 1367; see also Olkkonen et al 2008, page 6). For the MCI the achromatic adjustments are projected on the axis of the typical adjustments that leads through the subjective grey point. The distance of this projection from the subjective grey point measures how strong the shift along this axis was. For the MCI this measure is divided by the length, ie the saturation, of the typical adjustment. In this way, the MCI represents the ratio of achromatic shift relative to the colourfulness of the typical colour. The sign (+/−) of the MCI reflects the direction in which the adjustment is shifted away from the subjective grey point. A positive MCI indicates an achromatic adjustment opposite to the typical adjustment. A negative MCI implies, contrary to the memory colour effect, that there is a shift of the achromatic adjustments towards the same direction as the typical adjustments. The MCI has been calculated separately for each participant using their subjective grey point. The average MCIs for each colour diagnostic object are shown in figure 5.

The MCI for ten of the fourteen colour diagnostic objects was positive. This implies that the achromatic adjustments of these objects were shifted towards the colour opposite to their typical colour. The average MCI over all fourteen colour diagnostic objects was 3.31% and significant across the objects, $t(13) = 2.97$, $p < 0.01$. Right-tailed t-tests across the averages of the participants revealed that for seven of these fourteen objects this shift was significantly higher than zero. These objects were the following, listed in descending order of their MCI: the Nivea tin with an average MCI of 10.3%, $t(24) = 5.9$, $p < 0.01$, UHU with an MCI of 8.5%, $t(23)=3.9$, $p<0.01$, the blue traffic sign with an MCI of 8.4%, $t(23) = 5.8$, $p < 0.01$, the Milka bar with an MCI of 6.0%, $t(23) = 4.6$, $p < 0.01$, the mailbox with an MCI of 4.8%, $t(24) = 3.4$, $p < 0.01$, the smurf with an MCI of 4.7, $t(24) = 4.1$, $p < 0.01$, and the Pink Panther with an MCI of 3.0%, $t(23) = 2.7$, $p < 0.01$. In contrast, for the heart a left-tailed t-test revealed that the MCI of −4.4% was significantly smaller than zero, $t(24) = −4.2$, $p < 0.01$. Finally, the MCI for the
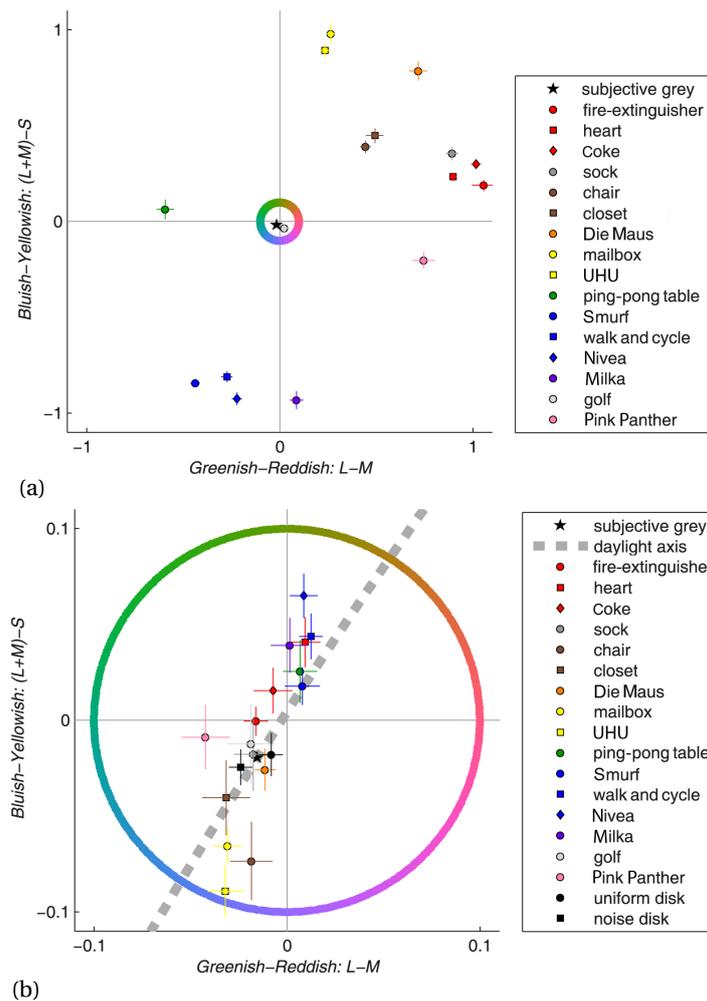
Figure 4. Adjustments for each object in DKL colour space. The DKL space is scaled so that along the axes an absolute value of 1 corresponds to the monitor gamut. The coloured ring represents hue variation along the azimuth in DKL colour space. It is shown with a radius of value 0.1 in both graphics. In this way, it also indicates the difference in scale between the two graphics. The coloured symbols show the adjustments, averaged over repetitions and participants. The lines crossing the disks are the standard errors of mean for the *x*-axis and *y*-axis across participants. The black pentagram represents the average subjective grey-point. (a) Typical adjustments for each object. (b) Achromatic adjustments for each object. The DKL space is zoomed in to values between −0.1 and 0.1. The dashed grey line is the daylight axis, shifted so that the correlated colour temperature of 6500 (standard illuminant *d65*) falls into the subjective grey point. Figure (a) shows that participants adjusted the correct typical colours quite reliably. In figure (b) the shifts of the achromatic adjustments are stronger along the daylight axis. A comparison of the corresponding symbols in figures (b) and (a) shows that most achromatic adjustments were shifted towards the opposite direction of the typical adjustments as assumed for the memory colour effect.

colour-neutral sock was 0.2% and was far from a significant difference from zero, $t(23) = 0.1$, $p = 0.90$.

3.3 Discussion

We investigated whether achromatic adjustments of the artificial objects' colours were shifted towards the opposite direction of the typical adjustments. This, we assumed, should be due to the influence of memory colours on colour appearance. We found such a shift for most of the objects. As measured through the MCI, this shift could reach up to over 10% of the typical colour such as it was the case for the Nivea tin. The amount of this shift indicates how
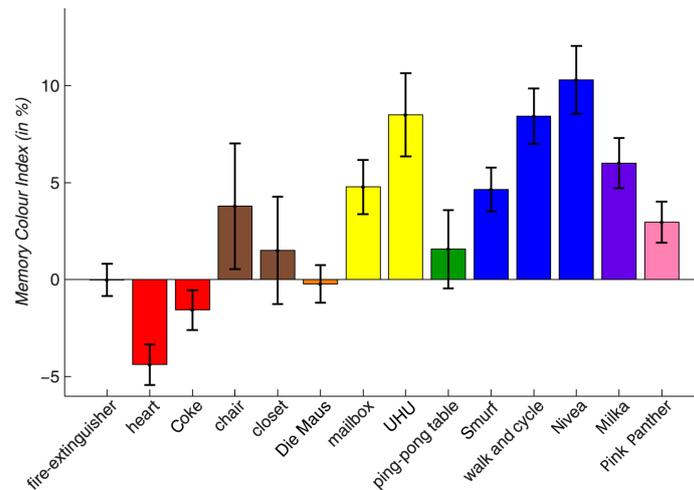
Figure 5. Memory colour indices (MCIs) for each object. Bars show the MCI in percent averaged over participants. Error bars represent standard errors of mean. Bars are ordered according to the azimuth of the objects' typical colours. Double asterisks (**) indicate that the average MCI of the respective object is significantly different from zero in a paired one-sided t-test across participants and with an alpha of 0.01. Congruent with the memory colour effect, for ten out of fourteen objects the MCI is higher than zero, and for seven of these objects this difference was significant. For the heart the MCI is significantly lower than zero. However, across all objects the MCI is still significantly above zero, thus confirming the pattern of the memory colour effect.

much of the typical colour people still see when the object is actually grey, since it is this apparent colour that they counteract by shifting the object's colour towards the opponent colour. At the same time, the golf ball should not show any systematic shift of the achromatic adjustments because its typical colour is achromatic. The sock was expected not to show a systematic shift, either, because it is a colour-neutral object. The participants saw it only once in its original colour at the beginning of the experiment. This was enough for them to set its typical colour correctly, but it should not be sufficient to internalise its object–colour association. And, indeed, there was no systematic shift away from the subjective grey point for either of these two control objects. So, the systematic shifts are indeed bound to people's knowledge about the object–colour association, and may be interpreted as memory colour effects.

Nonetheless, we also observed three particularities that need further discussion. Firstly, the subjective grey point, as measured through colour-neutral disks, was significantly shifted towards blue. Secondly, we found an alignment of achromatic colour adjustments along an axis that connects blue and yellow. Finally, we observed that the memory colour effect did not appear to have the same magnitude for all objects. In particular, red and orange stimuli did not yield positive MCIs. In the next three sections we will discuss possible origins of these phenomena and their implication for the interpretation of positive MCIs as memory colour effects. Together they will answer the questions which features of the stimuli are responsible for positive MCIs, and whether the MCI really reflects memory colour effects.

3.3.1 *Grey-point shifts.* We observed that the subjective grey point, as measured through the achromatic adjustments of the uniform and noise disks, was slightly shifted away from the grey point of the monitor background towards a bluish hue. The same was true for the achromatic adjustments of the two control stimuli of our experiment, the golf ball and the colour-neutral sock. Such a shift might be due to a mismatch between the monitor background and the lamps. However, the preceding studies on the memory colour effect revealed the same phenomenon (Hansen et al 2006, figure 2b; Olkkonen et al 2008, figures

3a and 3b). The monitor and background had been readjusted by four new observers for the present study. For this reason it is very unlikely that the same distortion reoccurred by accident for both, our calibration and the one of the preceding experiments. Additionally, other colour adjustment studies also found such a systematic shift of achromatic colours towards blue, when adjustments were made on a monitor (Granzier et al 2009; Oicherman et al 2009). These studies systematically investigated shifts of colour adjustments in a particular direction and used very different setups.

Owing to the colour adjustment technique used in this experiment, objects with a colour distribution, such as the UHU glue stick, exhibit a colour distribution at the subjective grey point. This colour distribution spans from the objective grey point [ie the origin (0, 0) of DKL colour space] to the subjective grey point. Only when these objects are adjusted to the objective grey-point is their colour distribution compressed so that there are only differences in lightness, but none in hue and saturation (for details see figure S3 in the supplementary material). As a result, objects with colour distributions have different nuances in hue and saturation when adjusted to the subjective grey point. However, for all of these objects the colour distributions at the subjective grey point are very narrow (see in particular figure S3c-d). For colour diagnostic objects, which yield a memory colour effect, the distribution at the average achromatic adjustment is much larger (see in particular figure S3e-f). This, however, is not the case for the colour-neutral object (the sock), which also has a complex colour distribution (see figure S4 in the supplementary material). Overall, the grey point shift cannot account for the shifts of the achromatic adjustments of the colour diagnostic objects for at least two reasons. Firstly, in our study the shifts for the colour diagnostic objects also occur in the direction that is opposite of the grey point shift. Indeed, the blue and violet stimuli (Nivea, traffic sign, smurf, Milka) that were adjusted in the direction opposite to the grey point shift yielded the strongest shifts away from the grey point, and hence highest MCIs. Furthermore, the shifts of the colour diagnostic objects were much stronger than the shift of the subjective grey point. One idea might be that this is the case because the grey point shifts are stronger for complex objects in contrast to the simple disks. Then this should also occur for the adjustments of the two control objects, the sock and the golf ball. But it did not: the achromatic adjustments of the control objects were not shifted beyond the subjective grey point, but coincided almost exactly with it. Another observation that contradicts this idea is that objects without a complex colour distribution, such as the smurf or the traffic sign, did also yield memory colour effects.

In sum, there seems to be a systematic shift of the subjective grey point for colour adjustments on the monitor. However, this phenomenon does not undermine the interpretation of our results as memory colour effects. Additionally, we compensated for this systematic grey point shift by using the subjective grey point when evaluating memory colour effects.

3.3.2 *Correlation along the blue–yellow axis.* We found a correlation between the x and y values of the achromatic adjustments of all stimuli. This shows that the achromatic adjustments in general were aligned along an axis that passes from the bluish to the yellowish part of the colour space (cf figure 4b). This phenomenon was also present in previous studies on the memory colour effect in that the x- and y-coordinates of all achromatic adjustments (including the two kinds of disks) also correlated positively. In Hansen et al (2006) this correlation was high, but due to the small number of stimuli only marginally significant, $r = 0.633$, $p = 0.067$, $n = 9$. The second part of the first experiment of Olkkonen et al (2008) contained the highest number of different stimuli (ie five photos of fruits, five photos of white-painted fruits, five outline shapes, and two kinds of disks). Here, this correlation is clearly confirmed, $r = 0.86$, $p < 0.01$, $n = 17$. In both studies the total variance of the x- and y-axes may be represented by a principal component to 88% and 96%, respectively. The

slope of this principal component is the same in both studies (68 deg.) and very close to the one we found for the achromatic adjustments in our study (72 deg.). In fact, a closer look at the respective figures (Hansen et al 2006, figure 2b; Olkkonen et al 2008, figures 3a and 3b) reveals that the achromatic adjustments were not exactly oriented in the direction of the colour that is opponent to the typical colour. Instead, they were all slightly but consistently tilted away from the exact opponent colour towards the bluish direction. Furthermore, this enhanced variation of the achromatic adjustments along the blue–yellow axis is consistent with the MCIs. In our study MCIs correlated positively with the distance of the achromatic adjustments to the subjective grey-point along this axis, $r = 0.55$, $p = 0.04$, $n = 14$. While this correlation was not significant for the small sample of stimuli in Hansen et al (2006), $r = 0.41$, $p = 0.36$, $n = 7$, it clearly was present in Olkkonen et al (2008, experiment 1 part 2), $r = 0.68$, $p < 0.01$, $n = 15$. The question arises of why achromatic adjustments were shifted more strongly along the blue–yellow axis.

A candidate correlate of the blue–yellow axis may be the variation of daylight hue. Natural daylight varies along an axis from blue to yellow (Mollon 2006; Wyszecki and Stiles 2000, page 7). This is due to the combination of different amounts of direct yellow sunlight and bluish diffusion through Rayleigh scattering. Consequently, the absolute hue of an object may vary along this blue–yellow axis across the day, whereas it stays relatively stable on the orthogonal greenish-to-reddish axis. This implies that in their natural environment people are exposed to uncertain information about the object's hues along the daylight axis. As a result they might also be less certain when estimating an object's colour along this axis. This idea is supported by research on the subjective white point. When participants were to set white points, these estimations vary more strongly along the blue–yellow axis (Beer et al 2006; Halen et al 2010). In order to verify whether this is also true for the achromatic settings in our experiment, we analysed intraindividual and interindividual variations of the achromatic settings for the control stimuli (see figure 6). These were the uniform disk, the noise disk, the golf ball, and the sock. Interindividual variation refers to the variation of achromatic adjustments for each of the four objects averaged for each participant (four objects x twenty-five participants minus outliers). In contrast, intraindividual variation consists of the variation of achromatic adjustments around the average for each individual and each object (four objects x twenty-five participants x three repetitions minus outliers). For this intraindividual variation, x and y-coordinates correlated with $r = 0.57$ ($p < 0.01$, $n = 279$). This common variation may be represented by a principal component with a slope of 64° and an explained variance of 83%. For the interindividual variation, the correlation between x and y values was $r = 0.72$ ($p < 0.01$, $n = 98$), and the principal component was tilted by an angle of 65 deg. (explained variance = 90%). This implies that people are less certain about the estimation of grey along an axis close to 65 deg. The daylight axis may be approximated by the *correlated colour temperature* (CCT) (Wyszecki and Stiles 2000, page 7). The slope of the CCT in the segment of the DKL space that contains the typical stimulus colours (x and y varying from -1 to 1) is 57 deg. So, the main variations of the estimations of subjective grey (65 deg.) align quite well with the variation of daylight (57 deg.). This indicates that people are uncertain in their estimation of achromatic colours along the axis on which the hue of an object varies under natural illuminations.

The main variation of the achromatic adjustments for all sixteen objects had a slope of 72 deg. The deviation of 16 deg. of this variation from the daylight axis may be due to the fact that this variation is also oriented by memory colour effects. Because of this, its slope should depend on the stimulus sampling in that the typical colours of the objects in the sample determine the orientations of the respective memory colour effects. This may deflect the achromatic adjustments from the exact orientation of the daylight axis. Nevertheless, we still find a considerable alignment with the blue–yellow axis (cf figure 4b). So, the actual slope of

this axis might be a combination of both memory colour effects and uncertainty about the colour estimation along the daylight axis. In this way, uncertainty along the daylight axis may well be the explanation for the alignment of achromatic adjustments with an axis that goes from blue to yellow. Of course, it might also be the case that luminance information interacts with daylight variation since the amount of blue and yellow hue in daylight is related to the height of the sun during the day. An object that appears comparatively light might be set to a more yellowish grey since higher lightness levels are associated with more yellow through more direct sunlight.
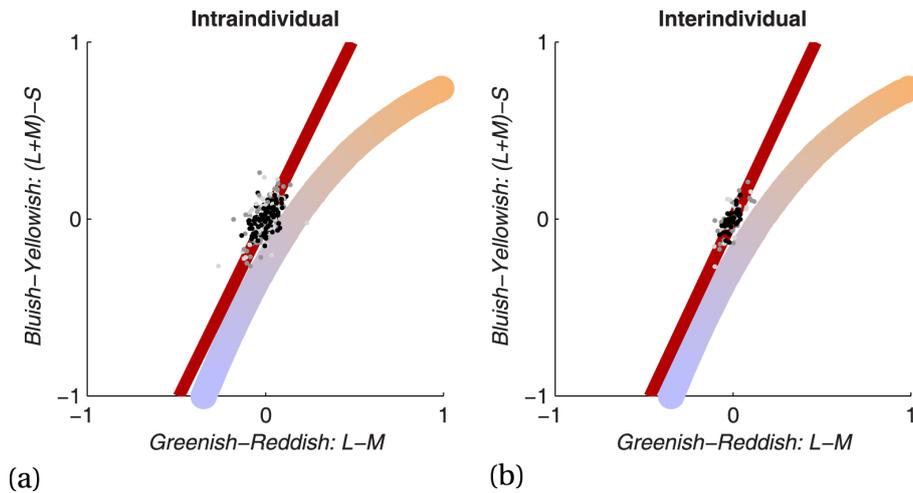
Figure 6. Variability of achromatic adjustments for control stimuli. The dots represent the single adjustments of four control stimuli: The two kinds of disks (black dots), the golf ball (light grey), and the colour-neutral sock (dark grey). The red curve is the principal component that represents the main common variation of *x* and *y* values. In order to illustrate daylight variation, the correlated colour temperature is shown as a curve that is coloured according to the actual hues that correspond to the coordinates. (a) Intraindividual variability: the coloured dots depict the single adjustments for all participants after subtraction of the respective participant's mean. (b) Interindividual variability: the coloured dots show the average adjustments for each object and each participant.

3.3.3 *Determinants of memory colour effects.* There were strong differences of memory colour effects across our colour diagnostic objects. These variations in MCIs cannot be explained through the level of colour diagnosticity as reflected in the reaction times of our preliminary experiment. Reaction times and MCIs were not correlated for our stimuli, $r = 0.13$, $p = 0.66$, $n = 14$.

Furthermore, these differences across objects may not be attributed to surface complexity or the level of abstractness of the colour diagnostic features (cf figure 5 with figure 3). Indeed, we observed that memory colour effects appeared independently of these two dimensions. On the one hand, among those images eliciting a memory colour effect there were not only three-dimensional objects like the Nivea tin, the mailbox, and the UHU glue stick, but also two-dimensional objects such as the Milka chocolate bar. Even objects that consisted just of uniform colour patches yielded memory colour effects, such as the smurf or the Pink Panther. At the same time, the images that did not elicit memory colour effects also included three-dimensional instances, such as the fire extinguisher. On the other hand, not only concrete objects such as the smurf or the Pink Panther led to memory colour effects. Memory colour effects also occurred for more abstract object–colour associations. They were elicited by symbols such as the blue traffic sign or the Post symbol on the German mailbox. Moreover, they also appeared for logos, whose main characteristic is a written name that refers to a company. This is the case for the Milka and the Nivea logo.

Instead, the main determinant of the variability of the MCIs is the typical colour of the objects (figure 5). In fact, it seems that the MCIs are highest for objects with typical colours that are close to the yellow–blue axis such as yellow, blue, and violet. And they are low for objects with colours that lie orthogonal to this axis such as red, orange, green, and pink. To apprehend the alignment of the objects' typical colours along the blue–yellow axis, we calculated the absolute angular distance of the adjustments of the typical colours from the angle of the principal component of the achromatic adjustments. These distances predicted well the MCI of each colour diagnostic object, $r = -0.67$, $p < 0.01$, $n = 14$. This shows that memory colour effects are enhanced along this axis. The boosted variation of the achromatic adjustments along the blue–yellow axis is a main determinant of the variability of the MCIs.

Still, not all variations of the MCIs may be explained through the enhancement of shifts along the blue–yellow axis. We also found that the x-coordinates of the typical adjustments (not the achromatic ones, as in the previous section) do well predict the MCIs, $r = -0.69$, $p < 0.01$, $n = 14$. This is due to the fact that some colours with low x values, such as blue and purple, led to particularly high MCIs, whereas colours with high x values, such as orange and red, yielded extremely low MCIs. For example, memory colour effects for red objects such as the Coca Cola logo or the fire extinguisher were close to zero. For the heart, we even found a significantly negative MCI. These results are in line with previous findings. Most of the classical studies that failed to replicate consistent memory colour effects used red objects (Bolles et al 1959; Bruner et al 1951, experiment 1; Fisher et al 1956). In particular, the heart led to results that were ambiguous (Harper 1953, page 88; cf heart and '*Y*' in table 1) or even reverse to those predicted through the memory colour effect (Fisher et al 1956, page 550). Furthermore, Olkkonen et al (2008) did not find consistent memory colour effects for the red strawberry either. At the same time, the uncertainty along the daylight axis may not completely account for the achromatic adjustments of the red objects. If there is high uncertainty in the direction of the typical colour of an object, there should be a shift that is congruent with the memory colour effect. If there is little uncertainty in the direction of the typical colour, the adjustments should vary along the daylight axis, but the average of this variation should not be systematically shifted. In contrast the heart and the Coca Cola logo were shifted significantly away from the subjective grey point (see results for achromatic adjustments) in a direction that is not congruent with the memory colour effect (cf figure 5). As can be seen in figure 4b, the red objects are shifted in the yellow direction and for the heart the magnitude of this shift is even equivalent to the shifts that were due to memory colour effects, such as the ones for the smurf, the Milka bar, and the blue traffic sign. Together, these findings show that red colours may elicit special effects on colour appearance that interact with or supersede memory colour effects.

In sum, there are several factors other than colour diagnosticity that modulate the strength of the memory colour effect. This may explain why, for our artificial objects, there is no direct correlation between memory colour effects (experiment 2) and colour diagnosticity as measured by reaction times (experiment 1). Certainly, recognisability and subjective colour diagnosticity of the objects are necessary conditions to produce memory colour effects. On the one hand, the impact of recognisability is shown by the correlation between reaction times and memory colour effects for the fruit stimuli of Olkkonen et al (2008). The photos of the painted fruits and the outline shapes have the same colour diagnosticity as the photos of the original fruits since they depict the same objects. At the same time, they expose fewer characteristic features than the original fruits in that they miss the typical texture and, in the case of the outline shapes, 3-dimensional shape. And, indeed, all outline shapes yielded low MCIs and, apart from the banana, comparatively high reaction times (cf figure 1; see also discussion of experiment 1). On the other hand, the necessity of colour diagnosticy for

memory–colour effects is shown by the fact that the sock, being an object with no subjective colour diagnosticity, did not yield a memory colour effect.

Apart from recognisabiltiy and colour diagnosticity as necessary preconditions, we found that the variation along the daylight axis as well as the special effects for red objects modulates the size of the memory colour effects. This may already be observed for the fruit stimuli of Olkkonen et al (2008). For example, the photo of the original banana, with a colour close to the daylight axis, yielded particularly high memory colour effects while the photo of the red strawberry produced particularly low memory colour effects. This was the case even though the reaction times for the photos of both fruits were comparable (cf figure 1). As a consequence, the correlation between reaction times and memory colour indices cannot completely predict the memory colour effects (only by 27% of the variance, see results of part 1). For our second experiment we chose objects that were all perfectly recognisable and maximally colour diagnostic according to accuracy rates and reaction times in the first experiment. The maximisation of colour diagnosticiy through the stimulus selection drastically reduced the variability of reaction times on a coarse scale. And it is at this scale that our reaction time measurements reflect colour diagnosticity (see evaluation of paradigm in the discussion of experiment 1). The residual variability of reaction times across these preselected objects may not only be caused by differences in colour diagnosticity. Instead, it may also be due to the strength of the contrast between the two colour categories in a response pair, the size of the stimulus samples per category, as well as variations in apparent lightness (see introduction to methods and evaluation of paradigm of experiment 1). Although we cannot totally exclude variations of colour diagnosticity across this stimulus sample; it is clear that compared with the experiment of Olkkonen et al such variations are comparatively low due to the preselection. The absence of a meaningful variation in colour diagnosticity across the selected artificial objects may explain the absence of a correlation between reaction times and memory colour indices. For our data the typical colours of the objects were much stronger predictors of the memory colour effects than the reaction times.

4 Conclusion

We investigated whether there is a memory colour effect for artificial, man-made objects. Through a reaction time experiment we guaranteed that the images of these man-made objects are well recognisable and that the objects are highly colour diagnostic for all participants. We may draw four kinds of conclusions.

Firstly, we found clear memory colour effects for artificial objects. Since these objects are tied to a particular cultural context, their association with a typical colour must have been learned in everyday life. Therefore, we conclude that acquired knowledge about objects modulates their colour appearance. These findings provide further evidence that object recognition and colour appearance interact in high-level vision. Moreover, they show that these interactions are mediated through past experience. In this way, they also support the idea that learning influences perception. Future studies about the memory colour effect should investigate the exact way in which learning of object colours affects their colour appearance. Therefore, these future studies should directly control the learning process itself.

Secondly, memory colour effects were particularly high for colours that were close to the daylight axis. The further colours were away from the daylight axis, the lower was the memory colour effect. At the same time, our findings indicate that people are less certain in the colour estimation along the daylight axis. This may be explained by the fact that the actual colour of a particular object varies along this axis under different natural daylight illuminations. We conclude that memory colours compensate the uncertainty about the absolute colours of

objects. In order to encounter this uncertainty, people resort more to their memory colours for colour estimations along the daylight axis than is the case for colours orthogonal to this axis. This supports once again the idea that colour appearance in particular and vision in general is strongly adapted to ecological constraints. It may be fruitful to investigate systematically the precise effects of uncertainty and of variations along the daylight axis on memory colours and colour memory. Moreover, future investigations should also address the precise implications of luminance information in colour adjustments and its connection to memory colour effects. We also found some particularities for red objects that could not be explained by the interaction between uncertainty and memory colour effects. At the same time, particular effects for red have also been shown for colour preferences (Bornstein et al 1976; Franklin et al 2009; Maier et al 2009) and for human behaviour in general (Elliot and Niesta 2008; Elliot et al 2007; 2009; Hagemann et al 2008; Hill and Barton 2005). Therefore, we wonder whether our results are linked to the fact that saturated red colours have special communicative functions in the natural environment, such as indicating alert, nutritious fruits, and sexual dispositions (for a discussion see, eg, Fernandez and Morris 2007, page 10).

Thirdly, we could observe that the memory colour effect for our stimuli appeared independently of their perceptual complexity and of the abstractness of their colour diagnostic characteristics. On the one hand, our results exemplify that the memory colour effect also exists for two-dimensional objects with uniform colours. This implies that the effect does not presuppose three dimensionality, texture, or complex colour distributions. Moreover, we may further specify the findings of previous studies (Hansen and Gegenfurtner 2006; Olkkonen et al 2008), which show that classical experiments yield particularly unstable memory colour effects because they used outline shapes as stimuli. Since the memory colour effect does not depend on the perceptual complexity of the stimuli, the weak memory colour effect for outline shapes seems to be due not to the poverty of their surface structure, but to the fact that they are less recognisable. This is further supported by the finding that the outline shapes yielded higher reaction times than the photos in our first experiment. Taken together, our findings suggest that the memory colour effect appears most strongly for stimuli that correspond to the visual experiences with which people were originally familiarised in their everyday life. On the other hand, we found that even symbols and writings that refer to abstract ideas may elicit a memory colour effect. Together with earlier studies (eg Joseph and Proffitt 1996; Naor-Raz et al 2003) these results point towards the important role of conceptual knowledge about object–colour associations in object recognition (for a discussion see, eg, Tanaka et al 2001). In how far the memory colour effect is modulated by perceptual features or whether it may be induced by pure conceptual knowledge is an open question for future research.

Finally, the evaluation of our reaction time paradigm provides a methodological contribution. The reaction times for the colour categorisation of achromatic images measured well their subjective, automatic colour diagnosticity. Our results show that this measure may identify candidate stimuli that have the potential to yield memory colour effects. However, the precise relationship between colour diagnosticity and memory colour effects could not be investigated. In the current implementation other determinants apart from colour diagnosticity may have modulated the size of the reaction times and the strength of the memory colour effects. Our findings suggest that at least the recognisability and the typical colours (ie, the effect of the daylight axis and red) of the objects must be controlled in order to measure the strength of this relationship. Moreover, our measurements cannot yet provide general insights about object-specific colour diagnosticity or the relationship between colour categories and object colours. In order to investigate effects of particular objects or categories,

it might be fruitful to apply our method to stimulus sets that are comparable across categories and that contain standardised depictions of the respective objects.

References

ActiveWire Inc, 2003 *ActiveWire (Version 1.0.14)* (Palo Alto, CA: ActiveWire Inc) ◄

Adams GK, 1923 "An experimental study of memory color and related phenomena" *The American Journal of Psychology* **34** 359–407 ◄

Adobe Systems Inc, 2005 *Adobe Photoshop CS2 (Version 9.0.1)* (San Jose, CA: Adobe) ◄

Baker KE, Mackintosh I, 1955 "The influence of past associations upon attributive color judgments" *Journal of Experimental Psychology* **49** 281–286 doi:10.1037/h0044052 ◄

Bartleson CJ, 1960 "Memory colors of familiar objects" *Journal of the Optical Society of America* **50** 73–77 doi:10.1364/JOSA.50.000073 ◄

Beer RD, Dinca A, C , 2006 "Ideal white can be yellowish or bluish, but not reddish or greenish" *Journal of Vision* **6** 417–417 doi:10.1167/6.6.417 ◄

Berlin B, Kay P, 1969 *"Basic color terms: their universality and evolution"* (Berkeley, CA: University of California Press) ◄

Biederman I, Ju G, 1988 "Surface versus edge-based determinants of visual recognition" *Cognitive Psychology* **20** 38–64 doi:10.1016/0010-0285(88)90024-2 ◄

Bloj M G, Kersten D, Hurlbert A C, 1999 "Perception of three-dimensional shape influences colour perception through mutual illumination" *Nature* **402** 877–879 ◄

Bolles R C, Hulicka I M, Hanly B, 1959 "Colour judgment as a function of stimulus conditions and memory colour" *Canadian Journal of Psychology/Revue canadienne de psychologie* **13** 175–185 doi:10.1037/h0083774 ◄

Bornstein M H, Kessen W, Weiskopf S, 1976 "The categories of hue in infancy" *Science* **191** 201–202 doi:10.1126/science.1246610 ◄

Boynton R M, Olson C X, 1990 "Salience of chromatic basic color terms confirmed by three measures" *Vision Research* **30** 1311–1317 doi:10.1016/0042-6989(90)90005-6 ◄

Brainard D H, 1997 "The Psychophysics Toolbox" *Spatial Vision* **10** 433–436 doi:10.1163/156856897X00357 ◄

Bruner J S, Postman L, Rodrigues J, 1951 "Expectation and the perception of color" *The American Journal of Psychology* **64** 216–227 doi:10.2307/1418668 ◄

Derrington A M, Krauskopf J, Lennie P, 1984 "Chromatic mechanisms in the lateral geniculate nucleus of macaque" *Journal of Physiology* **357** 241–265 ◄

Duncker K, 1939 "The influence of past experience upon perceptual properties" *The American Journal of Psychology* **52** 255–265 doi:10.2307/1416111 ◄

Elliot A J, Maier M A, Binser M J, Friedman R, Pekrun R, 2009 "The effect of red on avoidance behavior in achievement contexts" *Personality and Social Psychology Bulletin* **35** 365–375 doi:10.1177/0146167208328330 ◄

Elliot A J, Maier M A, Moller A C, Friedman R, Meinhardt J, 2007 "Color and psychological functioning: the effect of red on performance attainment" *Journal of Experimental Psychology: General* **136** 154–168 doi:10.1037/0096-3445.136.1.154 ◄

Elliot A J, Niesta D, 2008 "Romantic red: red enhances men's attraction to women" *Journal of Personality and Social Psychology* **95** 1150–1164 doi:10.1037/0022-3514.95.5.1150 ◄

Emmerson P G, Ross H E, 1987 "Variation in colour constancy with visual information in the underwater environment" *Acta Psychologica* **65** 101–113 doi:10.1016/0001-6918(87)90021-7 ◄

Fernandez A A, Morris M R, 2007 "Sexual selection and trichromatic color vision in primates: statistical support for the preexisting-bias hypothesis" *The American Naturalist* **170** 10–20 doi:10.1086/518566 ◄

Fisher S C, Hull C, Holtz P, 1956 "Past experience and perception: memory color" *The American Journal of Psychology* **69** 546–560 doi:10.2307/1419079 ◄

Franklin A, Bevis L, Ling Y, Hurlbert A, 2009 "Biological components of colour preference in infancy" *Developmental Science* **13** 346–354 doi:10.1111/j.1467-7687.2009.00884.x ◄

Gegenfurtner K R, 2003 "Cortical mechanisms of colour vision" *Nature Reviews Neuroscience* **4** 563–572 doi:10.1038/nrn1138 ◄

Gegenfurtner K R, Kiper D C, 2003 "Color vision" *Annual Review of Neuroscience* **26** 181–206 doi:10.1146/annurev.neuro.26.041002.131116 ◄

Gegenfurtner K R, Rieger J, 2000 "Sensory and cognitive contributions of color to the recognition of natural scenes" *Current Biology* **10** 805–805 doi:10.1016/S0960-9822(00)00563-7 ◄

Gosselin F, Schyns P G, 2003 "Superstitious perceptions reveal properties of internal representations" *Psychological Science* **14** 505–509 doi:10.1111/1467-9280.03452 ◄

Granzier J J, Brenner E, Smeets J B, 2009 "Do people match surface reflectance fundamentally differently than they match emitted light?" *Vision Research* **49** 702–707 doi:10.1016/j.visres.2009.01.004 ◄

Guest S, Van Laar D, 2000 "The structure of colour naming space" *Vision Research* **40** 723–734 doi:10.1016/S0042-6989(99)00221-7 ◄

Hagemann N, Strauss B, Leissing J, 2008 "When the referee sees red" *Psychological Science* **19** 769–771 doi:10.1111/j.1467-9280.2008.02155.x ◄

Halen K, Juricevic I, McDermott K, Webster M A, 2010 "Variations in achromatic settings across the visual field" *Journal of Vision* **10** 394–394 doi:10.1167/10.7.394 ◄

Hansen T, Gegenfurtner K R, 2006 "Color scaling of disks and natural objects at different luminance levels" *Visual Neuroscience* **23** 603–610 doi:10.1017/S0952523806233121 ◄

Hansen T, Giesel M, Gegenfurtner K R, 2008 "Chromatic discrimination of natural objects" *Journal of Vision* **8** 1–19 ◄

Hansen T, Olkkonen M, Walter S, Gegenfurtner K R, 2006 "Memory modulates color appearance" *Nature Neuroscience* **9** 1367–1368 doi:10.1038/nn1794 ◄

Harper R S, 1953 "The perceptual modification of colored figures" *The American Journal of Psychology* **66** 86–89 doi:10.2307/1417972 ◄

Helmholtz H L F v, 1867 *Handbuch der physiologischen Optik* 1st edition (Leipzig, Germany: Leopold Voss) ◄

Hering E, 1920 [1878] *Grundzüge der Lehre vom Lichtsinn* (Berlin, Germany: Springer) ◄

Herring B S, Bryden M P, 1970 "Memory color effects as a function of viewing time" *Canadian Journal of Psychology/Revue canadienne de psychologie* **24** 127–132 doi:10.1037/h0082840 ◄

Hill R A, Barton R A, 2005 "Psychology: red enhances human performance in contests" *Nature* **435** 293–293 doi:10.1038/435293a ◄

Hume D, 1894 [1748] *An Enquiry concerning Human Understanding* (Oxford, UK: Clarendon Press) ◄

Humphrey G K, Goodale M A, Jakobson L S, Servos P, 1994 "The role of surface information in object recognition: studies of a visual form agnosic and normal subjects" *Perception* **23** 1457–1481 doi:10.1068/p231457 ◄

Hurlbert A C, Ling Y, 2005 "If it's a banana, it must be yellow: the role of memory colors in color constancy" *Journal of Vision* **5** 787–787 doi:10.1167/5.8.787 ◄

Ishihara S, 2004 *Ishihara's Tests for Colour Deficiency* (Tokyo, Japan: Kanehara Trading Inc.) ◄

Joseph J E, Proffitt D R, 1996 "Semantic versus perceptual influences of color in object recognition" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22** 407–429 doi:10.1037/0278-7393.22.2.407 ◄

King A, 2000 "Situated Cognition" in *Encyclopaedia of Psychology* volume 7, Ed. A E Kazdin pp 289–291 (Washington, DC: Oxford University Press) doi:10.1037/10522-127 ◄

Krauskopf J, Williams D R, Heeley D W, 1982 "Cardinal directions of color space" *Vision Research* **22** 1123–1131 doi:10.1016/0042-6989(82)90077-3 ◄

Leibovich B, Paolera M B, 1970 "Research on color perception and after-image in relation to significant images" *Annales Médico-Psychologiques* **128** 735–742 ◄

Ling Y, Hurlbert A C, 2004 "Color and size interactions in a real 3D object similarity task" *Journal of Vision* **4** 721–734 doi:10.1167/4.9.5 ◄

Ling Y, Hurlbert A C, 2006 "Colour-memory-dependent colour constancy: 2D vs 3D real surfaces" paper presented at the Third European Conference on Colour in Graphics, Imaging, and Vision ◄

MacLeod D I A, Boynton R M, 1979 "Chromaticity diagram showing cone excitation by stimuli of equal luminance" *Journal of the Optical Society of America* **69** 1183–1186 doi:10.1364/JOSA.69.001183 ◄

Maier M A, Barchfeld P, Elliot A J, Pekrun R, 2009 "Context specificity of implicit preferences: the case of human preference for red" *Emotion* **9** 734–738 doi:10.1037/a0016818 ◄

MathWorks, 2007 *Matlab: The Language of Technical Computing Version R2007a* (MathWorks Inc: El Segundo, CA) ◄

Miceli G, Fouch E, Capasso R, Shelton J R, Tomaiuolo F, Caramazza A, 2001 "The dissociation of color from form and function knowledge" *Nature Neuroscience* **4** 662–667 doi:10.1038/88497 ◄

Mitterer H, de Ruiter J P, 2008 "Recalibrating color categories using world knowledge" *Psychological Science* **19** 629–634 doi:10.1111/j.1467-9280.2008.02133.x ◄

Mitterer H, Horschig J M, Musseler J, Majid A, 2009 "The influence of memory on perception: it's not what things look like, it's what you call them" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **35** 1557–1562 doi:10.1037/a0017019 ◄

Mollon J D, 2006 "Monge: The Verriest lecture, Lyon, July 2005" *Visual Neuroscience* **23** 297–309 doi:10.1017/S0952523806233479 ◄

Nagai J-i, Yokosawa K, 2003 "What regulates the surface color effect in object recognition: color diagnosticity or category?" Technical Report on Attention and Cognition 28 ◄

Naor-Raz G, Tarr M J, Kersten D, 2003 "Is color an intrinsic property of object representation?" *Perception* **32** 667–680 doi:10.1068/p5050 ◄

Nathans J, 1999 "The evolution and physiology of human color vision: insights from molecular genetic studies of visual pigments" *Neuron* **24** 299–312 doi:10.1016/S0896-6273(00)80845-4 ◄

Newhall S M, Burnham R W, Clark J R, 1957 "Comparison of successive with simultaneous color matching" *Journal of the Optical Society of America* **47** 43–54 doi:10.1364/JOSA.47.000043 ◄

Nicholson K G, Humphrey G K, 2003 "The effect of colour congruency on shape discriminations of novel objects" *Perception* **32** 339–353 ◄

O'Connor K, Glenberg A M, 2003 "Situated Cognition" in *Encyclopedia of Cognitive Science volume 4* Ed. L Nadel pp 19–25 (London: Nature Publishing Group) ◄

Oicherman B, Luo M R, Rigg B, Robertson A R, 2009 "Adaptation and colour matching of display and surface colours" *Color Research & Application* **34** 182–193 doi:10.1002/col.20492 ◄

Oliva A, Schyns P G, 2000 "Diagnostic colors mediate scene recognition" *Cognitive Psychology* **41** 176–210 doi:10.1006/cogp.1999.0728 ◄

Olkkonen M, Hansen T, Gegenfurtner K R, 2008 "Color appearance of familiar objects: effects of object shape, texture, and illumination changes" *Journal of Vision* **8** 1–16 doi:10.1167/8.5.13 ◄

Pelli D G, 1997 "The VideoToolbox software for visual psychophysics: transforming numbers into movies" *Spatial Vision* **10** 437–442 doi:10.1163/156856897X00366 ◄

Pérez-Carpinell J, Fez M D d, Baldoví R, Soriano J C, 1998 "Familiar objects and memory color" *Color Research & Application* **23** 416–427 doi:10.1002/(SICI)1520-6378(199812)23:6<416::AID-COL10>3.0.CO;2-N ◄

Ratner C, McCarthy J, 1990 "Ecologically relevant stimuli and color memory" *The Journal of General Psychology* **117** 369–377 doi:10.1080/00221309.1990.9921143 ◄

Regan B C, Julliot C, Simmen B, Vienot F, Charles-Dominique P, Mollon J D, 2001 "Fruits, foliage and the evolution of primate colour vision" *Philosophical Transactions of the Royal Society B: Biological Sciences* **356** 229–229 doi:10.1098/rstb.2000.0773 ◄

Schyns P G, Goldstone R L, Thibaut J-P, 1998 "The development of features in object concepts" *Behavioral and Brain Sciences* **21** 1–17 ◄

Schyns P G, Rodet L, 1997 "Categorization creates functional features" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **23** 681–696 doi:10.1037/0278-7393.23.3.681 ◄

Siple P, Springer R M, 1983 "Memory and preference for the colors of objects" *Perception & Psychophysics* **34** 363–370 doi:10.3758/BF03203049 ◄

St Julien J, 1997 "Explaining learning: the research trajectory of situated cognition and the implications of connectionism" in *Situated Cognition: Social, Semiotic, and Psychological Perspectives* Eds D Kirshner, J A Whitson pp 261–280 (Mahwah, NJ: Lawrence Erlbaum Associates) ◄

Stroop J R, 1935 "Studies of interference in serial verbal reactions" *Journal of Experimental Psychology* **18** 643–662 doi:10.1037/h0054651 ◄

Sturges J, Whitfield T W A, 1997 "Salient features of Munsell colour space as a function of monolexemic naming and response latencies" *Vision Research* **37** 307–313 doi:10.1016/S0042-6989(96)00170-8 ◄

Sumner P, Mollon J D, 2000 "Chromaticity as a signal of ripeness in fruits taken by primates" *Journal of Experimental Biology* **203** 1987–2000 ◄

Surridge A K, Osorio D, Mundy N I, 2003 "Evolution and selection of trichromatic vision in primates" *Trends in Ecology & Evolution* **18** 198–205 doi:10.1016/S0169-5347(03)00012-0 ◄

Tanaka J W, Presnell L M, 1999 "Color diagnosticity in object recognition" *Perception & Psychophysics* **61** 1140–1153 doi:10.3758/BF03207619 ◄

Tanaka J W, Weiskopf D, Williams P, 2001 "The role of color in high-level vision" *Trends in Cognitive Sciences* **5** 211–215 doi:10.1016/S1364-6613(00)01626-0 ◄

Therriault D J, Yaxley R H, Zwaan R A, 2009 "The role of color diagnosticity in object recognition and representation" *Cognitive Processing* **10** 335–342 doi:10.1007/s10339-009-0260-4 ◄

Van Gulick A E, Tarr M J, 2010 "Is object color memory categorical?" *Journal of Vision* **10** 407–407 doi:10.1167/10.7.407 ◄

Wandell B A, 1995 *"Foundations of vision"* (Sinaur Associates, Inc: Sunderland, MA) ◄

Watzlawick P (Ed.), 1984 *The Invented Reality: How Do We Know What We Believe We Know? (Contributions to Constructivism)* (New York, NY: Norton) ◄

Werner A, 2007 "Color constancy improves, when an object moves: high-level motion influences color perception" *Journal of Vision* **7** 1–14 doi:10.1167/7.14.19 ◄

White C W, Montgomery D A, 1976 "Memory colors in afterimages: a bicentennial demonstration" *Perception & Psychophysics* **19** 371–374 doi:10.3758/BF03204246 ◄

Wichmann F A, Sharpe L T, Gegenfurtner K R, 2002 "The contributions of color to recognition memory for natural scenes" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **28** 509–509 doi:10.1037/0278-7393.28.3.509 ◄

Wittgenstein L, 1953 *Philosophical Investigations* 3rd edition, translated by G E M Anscombe (New York, NY: MacMillan) ◄

Wurm L H, Legge G E, Isenberg L M, Luebker A, 1993 "Color improves object recognition in normal and low vision" *Journal of Experimental Psychology: Human Perception and Performance* **19** 899–911 doi:10.1037/0096-1523.19.4.899 ◄

Wyszecki G, Stiles W S, 2000 *Color Science: Concepts and Methods, Quantitative Data and Formulae* 2nd edition (New York, NY: John Wiley & Sons) ◄

Yendrikhovskij S N, Blommaert F J J, Ridder H d, 1999 "Representation of memory prototype for an object color" *Color Research & Application* **24** 393–410 doi:10.1002/(SICI)1520-6378(199912)24:6<393::AID-COL3>3.0.CO;2-Z ◄

Christoph Witzel received one degree in psychology and another one in political science and cultural anthropology at the University of Heidelberg in Germany. He is now doing his PhD with Karl Gegenfurtner at Gießen University. For more information visit http://www.allpsych.uni-giessen.de/chris/.

Hanna Valkova first obtained a degree in educational science at the A.S. Pushkin Brest State University in Brest, Belarus. She supplemented this degree by another one in psychology at Gießen University. Now, she is completing a training to become a Behavioural psychotherapist at the AVT in Cologne (Germany).

Thorsten Hansen received his diploma in computer science (hons.) in 1997 from the University of Braunschweig and in 2002 his doctoral degree (Dr. rer. nat.) from the Department of Neural Information Processing at the University of Ulm. Currently he is a research scientist at the Department of Psychology at the University of Giessen.

Karl Gegenfurtner studied psychology in Regensburg (Germany) and then did a PhD in Experimental Psychology at New York University. After spending time as a PostDoc in New York and Tübingen, he became Professor of Psychology in Magdeburg. Since 2001 he has been at Giessen University (http://www.allpsych.uni-giessen.de/karl/).
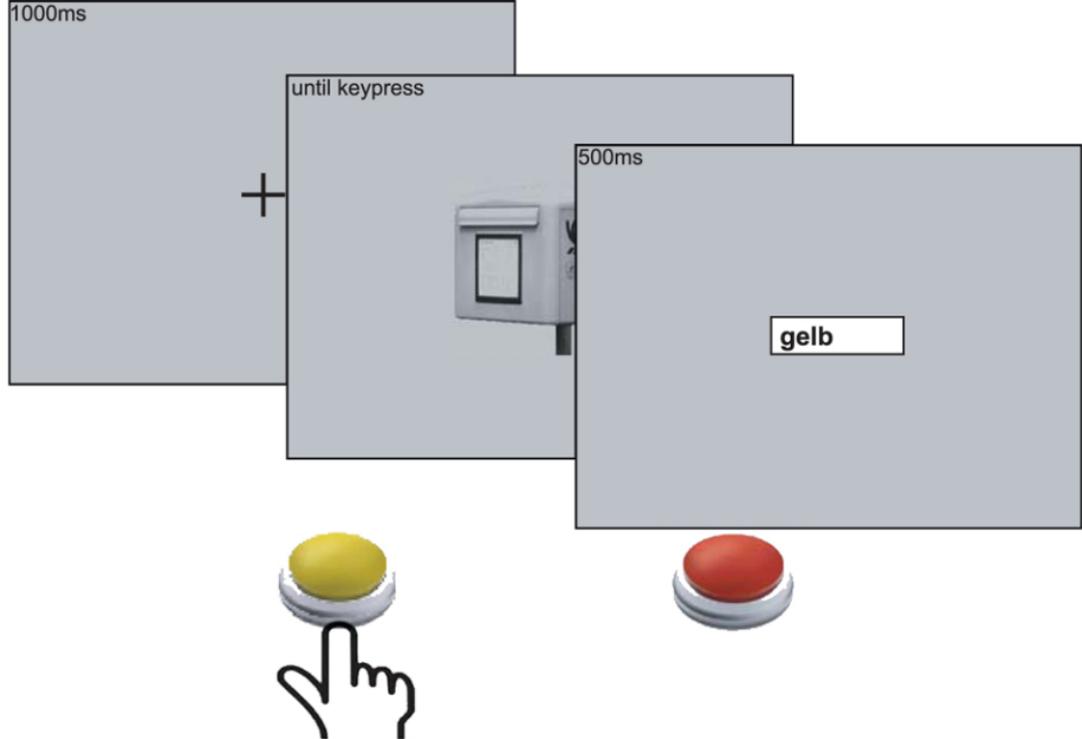
Figure S1. Procedure of the reaction time experiment. The course of one trial is shown as proceeding from the back to the front of the graphic. First a fixation point appeared for 1000ms. Then an achromatic image was shown until the participant responded; here the German mailbox is shown as an example. The participant had to decide which of two colour categories corresponds to the typical colour of the depicted object. At the end of the trial, the answered colour name was shown for 500ms. The stronger the association between object and colour, the faster people can indicate the typical colour of the object. Before these main trials, participants completed the same task with coloured disks instead of greyscale images of objects.
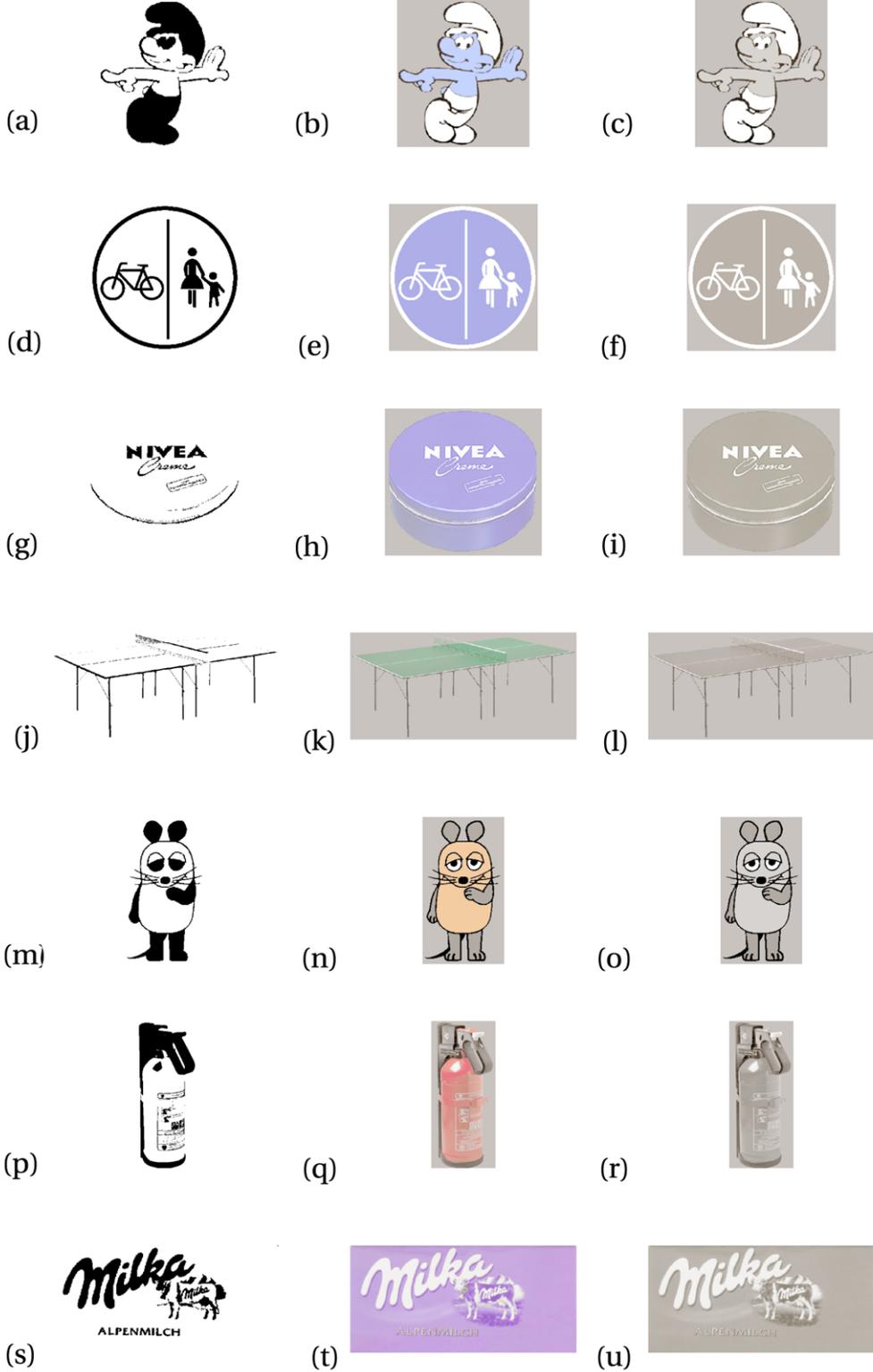
Figure S2. Areas kept achromatic during colour adjustment. The rows show the different objects with the respective areas: a.-c.) smurf, d.-f.) traffic sign, g.-i.) Nivea tin, j.-l.) ping-pong table, m.-o.) Die Maus, p.-r.) fire extinguisher, and s.-u.) Milka chocolate bar. The first column shows in black which areas of the objects were kept constantly achromatic during colour adjustment. The second and third column show how these objects looked at the average adjustment of the typical and achromatic colour, respectively.
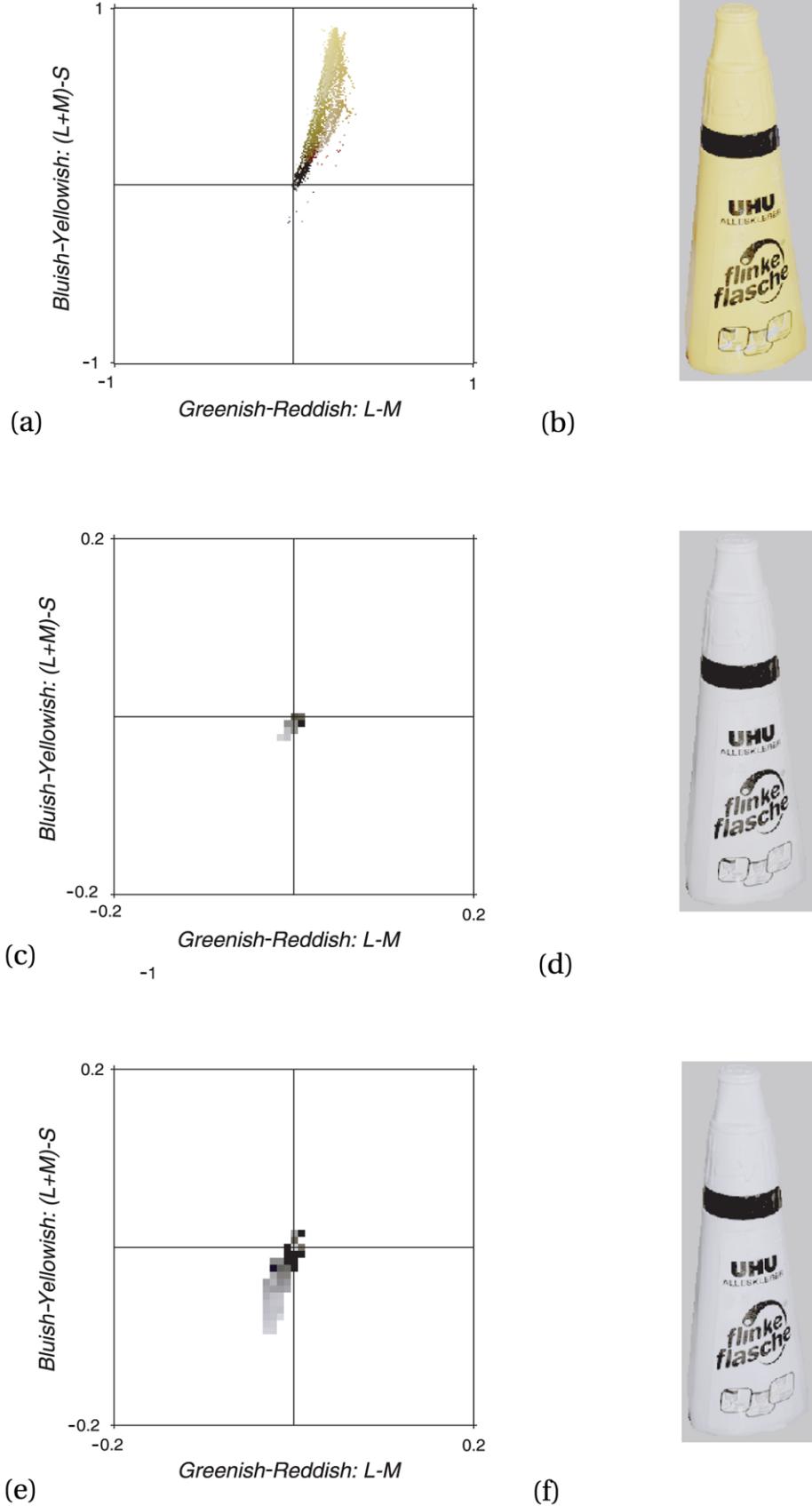
Figure S3. Colour distributions of a colour diagnostic object (UHU glue stick). The left column (a, c, e) shows the colour distributions in DKL colour space for different colour adjustments; the right column (b, d, f) shows the object with the respective colour distribution. DKL-Colour coordinates of these distributions are projected onto the plane spanned by the L-M and S-(L+M)-axes. The single points in the distributions are plotted in the respective colours. The adjustments shown here are averaged across blocks and participants. a.-b.) Typical adjustment of this object. c.-d.) The object at the subjective white-point. e.-f.) Achromatic adjustment of this object. Note that for the distributions at the subjective white-point (c.) and for the achromatic adjustment (d.) the graphics are zoomed to the fivefold size of the one for the distribution of the typical adjustment. At the subjective white-point the colour distribution (c-d.) is very narrow compared to the one of the achromatic setting (e-f.).
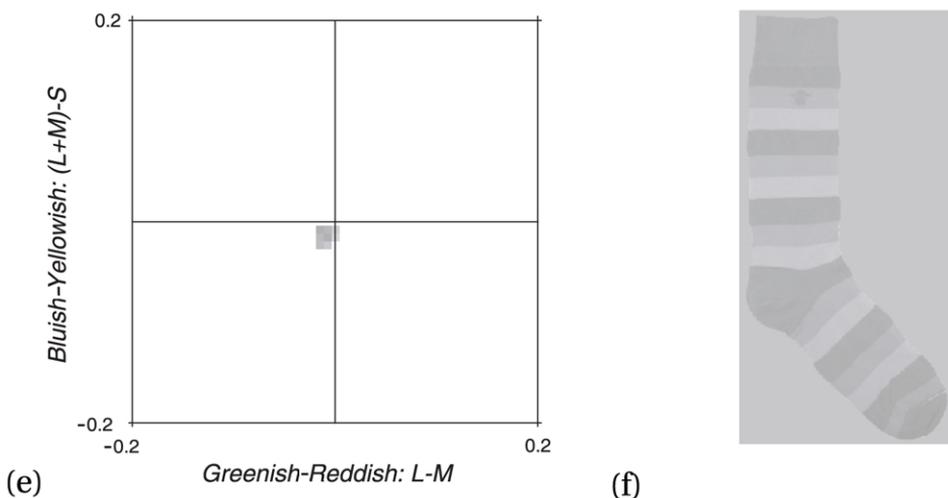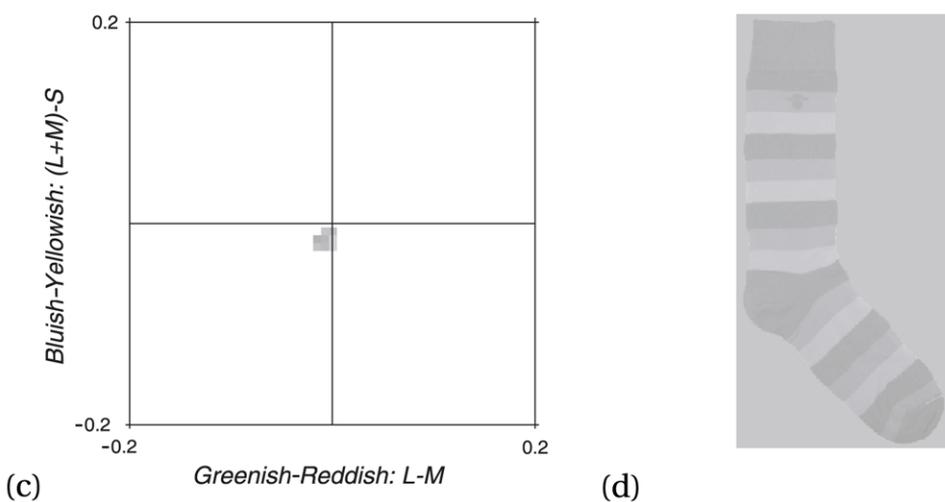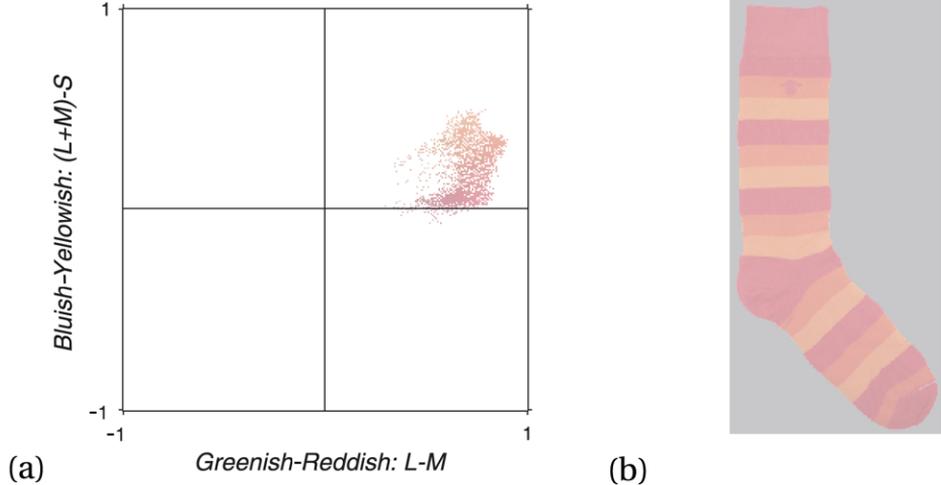
Figure. S4. Colour distributions of the colour-neutral object (sock). Details as in fig. S3. Like for the colour diagnostic object in figure S3, the colour distribution at the subjective white-point (c-d.) is very narrow. In contrast to the colour diagnostic object, here the colour distribution of the achromatic setting (e-f.) is almost completely the same as for the subjective grey point.