universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Color me your color: A cross-cultural investigation into implicit associations between color and valence“

verfasst von / submitted by

Claudia Kawai, M.A.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktorin der Naturwissenschaften (Dr. rer. nat.)

Wien, 2021 / Vienna, 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 796 605 298
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Psychologie
Betreut von / Supervisor:	Prof. Dr. Ulrich Ansorge

List of Papers

This thesis is based on the three following articles, which are referred to in the text by their Roman numerals. The contribution of Claudia Kawai is printed below each paper.

I **Kawai**, **C**., Lukács, G., & Ansorge, U. (2020). Polarities influence implicit associations between colour and emotion. *Acta Psychologica, 209*, 103143. https://doi.org/10.1016/j.actpsy.2020.103143

Contributions: Conceptualization and study design in collaboration with the supervisor UA and co-author GL. Software in collaboration with GL. Solely responsible for data collection and statistical analysis. Preregistration and manuscript in collaboration with UA and GL.

II **Kawai**, **C**., Lukács, G., & Ansorge, U. (2021). A new type of pictorial database: The Bicolor Affective Silhouettes and Shapes (BASS). *Behavior Research Methods*. 1-18. https://doi.org/10.3758/s13428-021-01569-7

Contributions: Conceptualization in collaboration with the supervisor UA and co-author GL. Design, software, preregistration, supervision of data collection and statistical analysis in collaboration with GL and UA. Solely responsible for image collection. Manuscript in collaboration with GL and UA.

III **Kawai**, **C**., Zhang, Y., Lukács, G., Chu, W., Zheng, C., Gao, C., Gozli, D., Wang, Y., & Ansorge, U. (2021). *The good, the bad, and the red: Implicit color-valence associations across cultures* [Manuscript submitted for publication]. Department of Cognition, Emotion, and Methods in Psychology, University of Vienna.

Contributions: Conceptualization and study design in collaboration with the supervisor UA and co-author WC. Software in collaboration with GL. Stimulus translation and selection in collaboration with WC and YZ. Collaborative data collection in Xi'an with YZ, collaborative data collection in Macau with CZ. Solely responsible for data collection in Austria and for statistical data analysis. Preregistration in collaboration with UA and GL. Manuscript in collaboration with UA, proofread by GL, DG and YW.

Table of Contents

1. Introduction

Imagine a machine with two colored but unlabeled buttons: One button is red and the other is white. The machine is currently powered down. To power it up, which button would you press? Would you first press the red button or the white one? Now, imagine the buttons were red and green. Which button would you choose here? The red button is present in both scenarios—would you select it (or not press it) in both cases, or does the color of the second button influence your choice?[1]

The environment we live in is one of the central influences shaping us into the beings we are (e.g., Gibson, 1966; Greeno 1994). No matter where we are from, color carries meaning because we use it as a cue to convey messages (e.g., red stop lights, green 'ok' buttons). Repeatedly experiencing color in certain situations in our surroundings leads to the formation of color associations (e.g., Elliot et al., 2007; Palmer & Schloss, 2011).[2] This binding of situated perceptual cues with certain co-occurrences (behaviors, stimulus appraisals, outcomes or consequences) forms the basis for embodied views on cognition. They hold that encountering a perceptual cue, like a green light, for example, simulates instances of prior situations where this cue was experienced (e.g., Barsalou, 1999, 2008; Havas & Matheson, 2013). This mechanism can explain how cross-modal associations between concrete physical percepts and abstract concepts—'red is negative' or 'brightness is up'—can arise (see also *conceptual metaphor theory*, Lakoff & Johnson, 1980, 2008).

Biological origins are assumed to play a role in ascribing meaning to color (e.g., primates using red color displays to attract mates, see Pflüger et al., 2014; see also Elliot et al., 2010) as well as societal learning. Studies in which people were asked which emotions they

[1] No statistically significant results are available for this study, since the sample size presently stagnates at one. CK has a European background and answered 'red' in Scenario I, and 'green' in Scenario II.

[2] This depends on the individual's visual acuity, of course.

associate with certain colors revealed astounding similarities across the globe—but, at the same time, individual differences that were present in these explicit cross-cultural comparisons revealed language and culture as modulating factors (e.g., Adams & Osgood, 1973; Jonauskaite et al., 2020). This led researchers to the conclusion that both *nature and nurture* exert their influences when it comes to color association formation (for a discussion see Adams & Osgood, 1973; Elliot & Maier, 2012).

One instance of the culture-specificity of affective color meaning can be observed when comparing China to Western countries: While green seems generally linked to positive emotions across the board, white invokes significantly more negative associations in China than in the West (Jonauskaite et al., 2020, 2019). Red is linked to both positive (romance, love) and negative (anger, danger) concepts in the majority of tested countries, but in Chinese culture, it additionally carries the meaning of happiness, life, and prosperity (e.g., He, 2011; Jiang et al., 2014). Going back to the machine with the two colored buttons: Maybe it *does* matter where you come from after all. If people around the world were asked to power up this machine in the two different scenarios, how would they decide? Would we find differences between the cultures?

At the core of this thesis lies the cross-cultural comparison of color-valence associations in different color systems, using implicit measures (Study III). Under examination were two major culture groups: Western and Chinese.[3] The groundwork for this comparison was laid by two prior projects: In Study I, we investigated the influence of polarity on the relationship of

[3] To clarify: "Chinese", here, refers generally to the population and cultural characteristics of Mainland China, which I distinguish from special administrative regions like Macau. Jiang et al. (2014) demonstrated a difference in color-related cross-modal associations between Mainland and Hong Kong Chinese, and our Study III will show a similar divergence for Macanese participants. Note that the explicit color-emotion associations we cite and compare our findings to were either collected from Hong Kong Chinese (Adams & Osgood, 1973) or unspecified as to their region in Chinese territory (Jonauskaite et al., 2020, 2019). "Western culture", here, subsumes German-speaking Europeans from Studies I and III, as well as US Americans from Study II.

red and green color to valence; in Study II, we developed a stimulus database containing affective silhouette pictures. Due to their language-independence, it is possible to study color-valence associations across different cultures with the exact same materials.

The following section takes a closer look at the methodology of the implicit measures that we used throughout our studies. Therein, I address the key terms *polarity correspondence* (Proctor & Cho, 2006) and *color systems* (Sections 1.1.1 and 1.1.2, respectively). Subsequently, I summarize our empirical work consisting of three papers (Section 1.2). All articles, each with their individual in-depth discussions on the respective focal points, are included in full text in Section 2. This thesis concludes with an overarching discussion of the subject matter and the implications that arise from my investigation into different aspects of implicit color-valence associations (Section 3).

1.1 Measuring Color-Valence Associations Implicitly

Studies using explicit measures, like reporting, selecting, or rating emotion associations with colors (Clarke & Costall, 2008; Kaya & Epps, 2004; Valdez & Mehrabian, 1994) can uncover a wealth of associations, from common to individual. The same applies to cross-cultural comparisons in which universalities as well as idiosyncrasies of color meaning and affect are commonly reported (Adams & Osgood, 1973; Jonauskaite et al., 2020). It is worth pointing out that such measures depend on explicit memory and awareness of color-emotion relations. Implicitly learned associations which persist, but may not be disclosed in explicit reports, can be targeted using implicit measures (for a discussion of explicit measures see Ansorge, 2021).[4] Interestingly, despite being an easy-to-implement method that is fairly standardized but flexible at the same time, direct cross-cultural comparisons of implicit color-emotion associations are

[4] Among the most commonly used implicit methods for investigating cross-modal interactions are the Implicit Association Test (IAT, Greenwald et al., 1998) and perceptually primed semantic classifications. We chose to apply the latter for our investigations (see Section 1.1.1).

scarce (see Hong et al., 2020 for a recent example). To close this gap in research, we decided to use the same implicit method throughout our studies. It is reasonable to assume that, in general, implicit color-valence associations mirror findings from explicit associations (see above), and that the similarities and differences found across cultures show up in both measures in a complementary fashion.

1.1.1 Polarity Correspondence

The method we chose to measure implicit color-valence associations was speeded bimanual semantic classification. This is an often-applied protocol in which, for example, words carrying a certain perceptual feature (such as words appearing in black or white font color) are to be categorized according to a semantic dimension (such as their emotional valence) via button presses (see Meier et al., 2004). According to a priori specified hypotheses, congruent and incongruent pairings of perceptual and conceptual features can be defined. To infer that a concept is implicitly associated with a certain perceptual feature (e.g., *black is negative*), response times and error rates are expected to be lower for the congruent stimuli (here white-positive and black-negative) when compared to the incongruent stimuli (here white-negative and black-positive).

Lakens (2012) points out the importance of polarity correspondence effects which arise in this specific type of task. In short, the polarity correspondence principle (Proctor & Cho, 2006) holds that structural symmetry in the design (two valences, two response possibilities, two colors) leads to a mapping of these dimensions (conceptual, response, and perceptual, respectively). A positive and a negative category pole are assigned to both the binary response categories (here positive vs. negative valence) and the perceptual dimension (here white vs.

black).[5] Response facilitation occurs only where the conceptual category and the perceptual feature share the same pole (i.e., where both are + or both are −). When the poles do not correspond, a conflict occurs, resulting in slower and less accurate categorization responses. Typically, responses to +polar items display a processing advantage over −polar items (e.g., positive words are classified faster than negative words) and the summed polarity benefits across conditions manifest in an asymmetry in the size of the congruence effects (differences between incongruent and congruent responses), with larger differences for +polar than for −polar responses (e.g., small differences for negative words presented in white or black, larger differences for positive words). For a more detailed explanation on polarity benefits see Lakens (2012).

If no structural overlap in the experimental block is possible or suggested in the first place however (e.g., when stimuli of two valence categories but only one color is present), we would not expect polarity correspondence effects to emerge (see Study I).[6]

1.1.2 Color Systems

A novel element of scrutiny in a cross-cultural context and central to the research in this thesis are *color systems*.[7] The way humanity has come to use color in many diverse communi-

[5] Markedness is one suggested factor that determines which concepts are assigned to the +pole, see also the *salience asymmetry hypothesis* by Rothermund and Wentura (2004), and the *dimension specificity hypothesis* by Schietecat et al. (2018a), see Section 1.1.2.

[6] For clarification, it is still expected to see the characteristic processing advantage for +polar items, that is, *+pole response facilitation.* Only the correspondence effects should not arise since there is no symmetric structural overlap (e.g., one color but two valence categories to respond to).

[7] With the term *color systems* we refer to the use of more than one color, repeatedly appearing within a certain context. Typically arising from the specific use of these colors in relation to one another (e.g., communicative functions), specific message-signaling calcifies by association. Encountering a color within its color system (e.g., red within a red-green color system) may highlight the respective associations (e.g., red as negative signal to green). Importantly, these color-system related associations may not be salient in other contexts (e.g., isolation) or color systems (e.g., red-white).

cative functions throughout everyday life is in the form of opposing colors (on the role of contrariety in human cognition see, e.g., Bianchi & Savardi, 2008; Osgood & Richards, 1973). One instance of this was illustrated in the very beginning of this thesis with red. It can be part of different color systems which, depending on contexts of accompanying colors, can alter the meaning of their constituent colors: Red in the company of white may send a different message than red in company of green ('activate' vs. 'stop', respectively).

The *dimension-specificity hypothesis* (Schietecat et al., 2018a, 2018b) suggests an approach to predict how implicit cross-modal associations play out, combining the polarity correspondence principle and the semantic differential (Osgood et al., 1957). The latter builds on the premise that conceptual knowledge can be structured along the dimensions of evaluation (valence), activity (arousal), and potency (dominance), each with a +pole and a −pole as endpoints. The same bipolar structuring that applies to concepts can also be assumed for perceptual cues like position, brightness, or color. The better the fit between the color and emotion dimension, the stronger the size of the association effect. More concretely, red is used particularly often as a negative 'stop', 'loss', or 'danger' signal, in a context where green is present simultaneously to signal the inverse positive concepts, like 'go', 'gain', or 'safety'. When testing implicit valence associations of red in two different color systems using IATs, Schietecat et al. (2018b) showed that the red-green color system produced larger polarity attributions than the red-blue color system.[8] Since IAT results are usually given in aggregated form, the exact attributions of all color-valence combinations are not directly evident. Yet following the notion that

[8] For instance, Schietecat et al. (2018b) showed that a red-blue opposition highlights the salience of the *activity* dimension (think tap water), with red being associated with warmth (+active) and blue with coldness (−active). In contrast, a red-green opposition highlights the *evaluation* dimension (think traffic lights), with red being associated with negativity (−positive) and green with positivity (+positive). Congruence effects (responses to incongruent minus congruent pairings) are stronger for red-negativity/green-positivity than red-negativity/blue-positivity, since in the former pairing, the two colors (green and red) can be mapped to a +pole and a −pole of the *same* semantic dimension as the two salient emotional categories (positive valence and negative valence), while the latter pairing is conceptually farther apart on an evaluation dimension (resulting in smaller dimension-specific polarity attributions) and demonstrably closer on an activation dimension.

a red-green color system evokes an interpretation in line with its frequent oppositional color use, we can assume that red implicitly becomes *more* negatively associated in a context of opposing green. Conversely, red might be *less* negatively associated in a context of opposing blue, since no strong color system is in use that suggests strong polar opposition in terms of valence, and this might extend to other colors as well. We hypothesize that in a context of white, negative conceptualizations of the color red are less emphasized than in a context of green, at least for a Western sample.[9]

It is, however, not necessarily the case that the color systems tested above with a Western sample take effect the same way across different cultures. Just as color-emotion associations 'in isolation' can be affected by cultural idiosyncrasies (see Section 1), so may the strength or conventionality of certain color systems. In traditional Chinese culture, for example, there might be a prevalent red-positive white-negative opposition (e.g., *wedding*, 红事, lit. 'red' + 'matter' vs. *funeral*, 白事, lit. 'white' + 'matter' in Simplified Chinese, see He, 2011). Without cross-cultural testing, it is unclear which color-emotion (or in a broader sense color-valence) association is predominant in a particular context (see Study III).

1.2 Findings and Contributions

Having introduced the central notions this thesis builds on (implicit association measures, polarity correspondence, and color systems), I now summarize our research into context factors in implicit color-valence associations and their culture-dependency.

[9] Elliot and Maier (2012) touch upon the different behavior elicited by red (approach vs. avoidance) in their Color-in-Context Theory, but do not make opposing color a context factor in and of itself.

1.2.1 Implicit Color-Valence Associations in Polar and Non-Polar Contexts: Red vs. Green

Before contrasting implicit associations between cultures, we wanted to investigate the role of polarity (structural symmetry vs. asymmetry) in implicit color-valence associations. In Kawai et al. (2020), we demonstrated that strong red-negativity/green-positivity congruence effects emerge for Austrian participants, but only in a color opposition condition. Where emotional words are presented in red and green front color intermixed (structurally symmetrical), we observed faster and more accurate classifications for positive words when presented in green and for negative words when presented in red. In contrast, no such response facilitation for negative words in red was found in the monochromatic condition, where green cues were absent (structurally asymmetrical). This can formally be explained by stronger polarity correspondence contributions in the symmetrical mixed-color condition that cannot be present in monochromatic contexts (see Section 1.1.1).[10] Importantly, these results also suggest that red is not necessarily and automatically associated with negative emotion (cf. Moller et al., 2009) when *not* accompanied by green.

1.2.2 Implicit Color-Valence Associations in Pictorial Material: Affective Silhouettes

Study I relied solely on linguistic material, since comparisons were performed within a single language group (German speakers). To contrast color-valence associations between China and the West, however, we thought it important to include a test that could be conducted independent of language, where all participants were presented with the same stimulus material. Pictures are cues that reliably induce emotions (e.g., Lang et al., 1997), however no picture set was available that could both elicit a wide range of emotion responses and at the same time

[10] In the article, we also ruled out a conflict adaption account as an alternative explanation for the asymmetry of polarity correspondence effects (i.e., stronger congruence effects for +polar than –polar items), see Kawai et al. (2020) for details.

allow for easy color manipulation.[11] So we established a set of 583 affective black-on-white silhouette pictures and recruited a large representative US sample (n = 777) as well as a large diverse Mainland Chinese sample (n = 869) to norm them: Half of the participants rated the pictures for valence, the other half rated them for arousal. Study III presents our pictorial database: the Bicolor Affective Silhouettes and Shapes (BASS, https://gasparl.github.io/BASS).

The cross-cultural affective ratings themselves revealed interesting characteristics: Valence ratings were highly correlated between the US and the Chinese raters, but arousal ratings were less so (due to a u-shaped vs. more linear valence-arousal-relationship, respectively). Interestingly, these cross-cultural differences were smaller among the younger adults than the older. This potential age-mediated East-West difference in explicit emotion ratings needs to be explored further and we refer the interested reader to Kawai, Lukács, and Ansorge (2021) for a discussion.

To return to the thesis' central topic of implicit associations: The paper also reports an online verification experiment, which featured a subset of the BASS silhouettes, to test implicit color-valence associations in two different color systems: red versus green, and black versus white. We again used the bimanual valence classification protocol, where positive and negative silhouettes had to be categorized by Austrian/German students. According to expectations, interactions between valence and the perceptual features emerged for the silhouette stimuli, showing red-negative/green-positive associations in the color block, and black-negative/white-positive associations in the brightness block (see also e.g., Lakens et al., 2012; Meier et al., 2004; Specker & Leder, 2018).

These two initial investigations (Studies I and II) provided important groundwork for our eventual cross-cultural comparison. First, we established that, for Westerners, an implicit

[11] Color filters applied over photorealistic pictures from the International Affective Picture System (IAPS; Lang et al., 1997) or Open Affective Standardized Image Set (OASIS; Kurdi et al., 2017), for example, resulted in heavily impaired recognizability, whereas line-drawings (e.g., Snodgrass & Vanderwart, 1980) did not meet our criteria for highly affective content.

green-positivity/red-negativity relationship prevails in color opposition (but not necessarily in isolation), which comes out via linguistic as well as pictorial stimulus modalities. Second, the BASS provided us with a tool to select pictures that were rated similarly across the two culture groups, adding a high degree of control to the stimulus material.

1.2.3 Implicit Color-Valence Associations in Cross-Cultural Comparison: West vs. China

In international surveys, China does not stand out in its explicit associations to either red or green from Western countries (Adams & Osgood, 1973; Jonauskaite et al., 2020). On these grounds, similar results would also be expected for implicit measures in the red-green opposition between the two culture groups. For white, however, explicit associations reportedly diverge (China's nation-specific sadness association with the term *white*, Jonauskaite et al., 2020). This suggests, that in an implicit test of red versus white valence associations, cross-cultural differences are likely to emerge. He (2011) even argued that in "Chinese culture, white is contrary to red" (p. 161).

For our cross-cultural comparison, reported in Kawai, Yang, et al. (2021), we tested these two color systems (red-green and red-white) using words (Experiment 1) and silhouette pictures from the BASS (Experiment 2, online study). The latter compared Austrians/Germans to Mainland Chinese, while the former also included a third participant group from Macau, which can be said to combine influences from Chinese as well as Western culture (for different color associations within China see Jiang et al., 2014).

In both verbal and pictorial stimulus modes, our implicit measures showed significant cross-cultural differences where *explicit* measures may have predicted only little discrepancy, namely in the red-green color system. Participants from Austria (and Macau in Exp. 1) displayed stronger green-positivity/red-negativity associations than participants from Mainland China. Surprisingly, in the red-white color system condition, where explicit measures would have predicted a higher degree of cross-cultural discrepancy (i.e., explicit white-sadness associations specific to China), results were very similar across the tested cultures. For white as well as red colored trials, positive stimuli were responded to faster than negative ones. No red-negativity response facilitation occurred for the West, highlighting the fact that negative associations for red become salient when green is present as an antipole but not white (nor black [unpublished], data on https://osf.io/dfs9e/). For Chinese participants, neither color system led to red-negative response facilitation, and there was no selectively stronger implicit white-negative association in China compared to the West.

This demonstrates, for the first time, that color systems can exert culture-specific effects on valence associations (strong red-green opposition in West), calling for consideration—or better yet investigations—into this context factor in explicit measures as well.

2. Empirical Work

2.1 **Kawai, C.**, Lukács, G., & Ansorge, U. (2020). Polarities influence implicit associations between colour and emotion. *Acta Psychologica*, *209*, 103143. https://doi.org/10.1016/j.actpsy.2020.103143

Acta Psychologica 209 (2020) 103143

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Polarities influence implicit associations between colour and emotion

Claudia Kawai[a,*], Gáspár Lukács[a], Ulrich Ansorge[a,b]

[a] Faculty of Psychology, University of Vienna, Austria
[b] Vienna Cognitive Science Hub, University of Vienna, Austria

ARTICLE INFO

Keywords:
Polarity correspondence
Congruence
Colour
Emotion
Valence

ABSTRACT

Colours are linked to emotional concepts. Research on the effect of red in particular has been extensive, and evidence shows that positive as well as negative associations can be salient in different contexts. In this paper, we investigate the impact of the contextual factor of polarity. According to the polarity-correspondence principle, negative and positive category poles are assigned to the binary response categories (here positive vs. negative valence) and the perceptual dimension (green vs. red) in a discrimination task. Response facilitation occurs only where the conceptual category (valence) and the perceptual feature (colour) share the same pole (i.e., where both are plus or both are minus). We asked participants (n = 140) to classify the valence of green and red words within two types of blocks: (a) where all words were of the same colour (monochromatic conditions) providing no opposition in the perceptual dimension, and (b) where red and green words were randomly mixed (mixed-colour conditions). Our results show that red facilitates responses to negative words when the colour green is present (mixed-colour conditions) but not when it is absent (monochromatic conditions). This is in line with the polarity-correspondence principle, but colour-specific valence-affect associations contribute to the found effects.

1. Introduction

The effect (and affect) of colour has long been–and will likely remain to be–a topic of interest to the public eye, pop culture, and sciences. Accordingly, research has addressed various subject areas, such as colour associations (Clarke & Costall, 2008; Crozier, 1999; Kaya & Epps, 2004), affective judgments (Briki & Hue, 2016; Terwogt & Hoeksma, 1995), cross-cultural preferences (Adams & Osgood, 1973; Saito, 1996), and colours' effects on mood (Akers et al., 2012) and behaviour (Labrecque & Milne, 2012; Lichtenfeld, Elliot, Maier, & Pekrun, 2012; Wilson, 1966).

Central to the current paper are the dynamic associations linking emotional[1] valence and colour. Red and green are the colours we will focus on, as research shows that these two colours can vary maximally on the valence (evaluation) dimension (see Schietecat, Lakens, IJsselsteijn, & De Kort, 2018b). Conventionally, financial gains, upward trends, secure situations and agreeing actions are depicted in green colour, and conversely, losses, downward trends, potential danger and disagreeing actions in red colour. Despite this conventional use that connects green to positive contexts and red to its negative counterpart, research on the colour red painted a rather colourful, inconsistent picture.

1.1. The red-effect(s)

Given the appropriate context, a colour can elicit very different associations, potentially even of opposing valences (e.g., Schloss, Strauss, & Palmer, 2013; see Elliot & Maier, 2012, for a review of context-dependence of colour). When it comes to explicit associations for red, both negative (anger, danger, blood) and positive (love, passion, warmth) concepts[2] are salient (Kaya & Epps, 2004; Jonauskaite et al., 2019; Soriano & Valenzuela, 2009). Likewise, when it comes to

* Corresponding author at: Faculty of Psychology, University of Vienna, Liebiggasse 5, A-1010 Vienna, Austria.
E-mail address: claudia.kawai@univie.ac.at (C. Kawai).

[1] Emotions are states of the human mind that are typically characterized by a particular subjective feeling (e.g., anger, sadness, joy), accompanied by typical cognitive representations (e.g., beliefs about the state of affairs represented by the emotion; e.g., that somebody took my sun cream away from me on purpose), physiological activity (in the brain and in the peripheral nervous system; e.g., activation of the pituitary gland), and by action tendencies (e.g., aggressive behaviour in the case of anger). In contrast, the term *valence* only fetches that part of the feeling of an emotion having to do with its felt overall positivity versus negativity.

[2] For most of the current manuscript, we use the term *concept* to cover both, more abstract semantic representations and more concrete sensory or sensorimotor representations. Only occasionally, we refer to the latter as *features*. However, for our argument about the contribution of polarity correspondence to colour-valence congruence effects, the distinction between abstract semantic concepts and concrete sensory features is peripheral. Therefore, for the sake of convenience, on most occasions we jointly refer to both, colours and emotions, by the single term *concept*.

https://doi.org/10.1016/j.actpsy.2020.103143
Received 22 November 2019; Received in revised form 26 May 2020; Accepted 15 July 2020

implicit associations, there is evidence for an ambiguous role of red, indicating links to both positive and negative valence. On the one hand, red was found to be related to high social status and prestige (Wu, Lu, van Dijk, Li, & Schnall, 2018). On the other hand, red was shown to be associated with negativity in general (Kuhbandner & Pekrun, 2013; Schietecatet al., 2018b), and anger, danger, and failure-related emotion words in particular (Fetterman, Robinson, & Meier, 2012; Moller, Elliot, & Maier, 2009; Pravossoudovitch, Cury, Young, & Elliot, 2014).

Its highly polysemous nature is emphasized by all the different red-effects researchers have reported, some of which suggest a positive, appetitive influence on human behaviour, others an aversive one. Red increases the perceived attractiveness in the opposite sex (Elliot et al., 2010; Elliot & Niesta, 2008; Pazda, Elliot, & Greitemeyer, 2012; but see Francis, 2013; Lehmann & Calin-Jageman, 2017; Lehmann, Elliot, & Calin-Jageman, 2018; Peperkoorn, Roberts, & Pollet, 2016), but also the perceived negativity in affective pictures (Buechner & Maier, 2016; but see Smajic, Merritt, Banister, & Blinebry, 2014; Steele, 2014). Even influences on cognitive and physical performance have been reported, in the way that red enhances achievement in detail-oriented cognitive tasks (Mehta & Zhu, 2009), but undermines it in IQ tasks (Elliot, Maier, Moller, Friedman, & Meinhardt, 2007; Lichtenfeld, Maier, Elliot, & Pekrun, 2009) and motor performance (Briki, Rinaldi, Riera, Trong, & Hue, 2015, but see Akers et al., 2012).

Looking at the results, it seems evident that colour works in context, but which exact contextual factors determine if a given colour evokes an appetitive or aversive stimulus evaluation in the beholder is unknown to date. In general, a variety of factors can contribute to the salience or the weight of each particular conceptual dimension of a stimulus to processing (e.g., Ansorge, Kunde, & Kiefer, 2014; Elliot & Maier, 2012; Lebois, Wilson-Mendenhall, & Barsalou, 2015). One contextual influence responsible for whether or not a conceptual factor becomes salient is task relevance of a stimulus dimension (see Moors & De Houwer, 2006). Task relevance can be expressed explicitly (e.g., via task instructions), but might sometimes also be of a more subtle nature: to categorise stimuli by alternative responses may highlight features that discriminate between the responses (e.g., their spatial positions) that in turn gate influences of similar stimulus features (e.g., their locations) on performance (cf. Ansorge & Wühr, 2004). Below, we explicate one such subtle contextual factor related to task relevance that we hypothesise plays a decisive role in colour-valence interactions: polarity correspondence.

1.2. Polarity correspondence

Reviewing the seemingly contradictory results on the colour red, as stated above, one decisive influence of how a colour is perceived, or reacted to, is the presence of opposite pole stimuli. In fact, many studies showed that, in general, congruence or compatibility effects between conceptual and/or perceptual dimensions are present where stimuli from opposite poles on such conceptual *and* perceptual dimensions are present; however, such congruence effects are absent where only stimuli from *one* pole of either the conceptual or the perceptual dimension are present (cf. Ansorge & Wühr, 2004; Wühr, Biebl, & Ansorge, 2008). Importantly, studies that investigated the relationship between colour and valence (see above, e.g., Fetterman et al., 2012; Wu et al., 2018), often (albeit not always) did so by opposing one colour (e.g., red) to another chromatic or achromatic colour in the same task (e.g., blue or grey), thereby manipulating colour *within* blocks and participants. These studies typically applied binary-classification tasks, like the implicit association test (IAT; see Greenwald, McGhee, & Schwartz, 1998) or other semantic categorisation tasks, requiring bimanual classification responses ('Press key A or key B').[3] This is interesting because *polarity correspondence* as a contextual factor allows congruence effects between conceptual and perceptual dimensions such as emotion, colour, or position that critically depend on the presence of opposing poles in two dimensions (e.g., different valences *and* different colours) in such binary-classification experiments (Lakens, 2012; Lakens, Semin, & Foroni, 2012; Proctor & Cho, 2006).

The rationale behind the polarity correspondence principle is the following: Binary classification of target stimuli that vary on one conceptual dimension by alternative responses (e.g., pressing left key for stimuli above vs. pressing right key for stimuli below fixation) suggests to the participants to code target alternatives as plus (+) poles versus minus (−) poles.

Our first test of the polarity-correspondence principle made use of the dimensions' dichotomy as a necessary precondition for the colour-valence congruence effect: the binary task-relevant as well as an additional binary irrelevant dimension, in our case positive versus negative valence and red versus green colour, respectively, that seems to be required to allow for many compatibility effects in general (cf. Ansorge & Wühr, 2004) and for a polarity-based congruence effect in particular. In a series of studies, Lakens et al. (2012) showed how an effect of the white-good association only emerged when participants formed a black-bad association in opposition. In line with the necessity of variation on the second dimension involved in this congruence effect, in the absence of black cues, congruence effects for white-good stimuli were no longer observed (Lakens et al., 2012).

While this was a test of brightness-emotion associations, this logic has recently also been applied in an indirect way to the hue-emotion association case of red/green colour-valence congruence effects by Schietecat et al. (2018b). Between tasks, these authors varied if participants had to classify target words by their *valence* or by their *activity* value. In a version of the IAT, Schietecat et al. (2018b) found that the change of task reversed the association of the colour red with the respective pole of the task-relevant conceptual dimension: In a binary *valence*-classification task (positive vs. negative), red was congruent with the – pole (i.e., negative) targets, but in a binary *activity*-classification task (aggressive vs. calm), red was congruent with the + pole (i.e., active/aggressive). This reversal suggests that the binary classification in one conceptual dimension defined by task-relevance is decisive for the *direction* of the colour-valence association, but the study by Schietecat et al. (2018b) left open if the *absence* of colour opposition diminishes the colour-valence congruence effects in the predicted way. If polarity correspondence is a decisive factor in the congruence of colour-valence associations, then congruence effect sizes should differ, depending on the presence versus absence of structural overlap between dimensions.

In the current study, the first test of contributions of polarity correspondence to colour-valence congruence effects therefore compared performance in blocks in which we presented red and green colour stimuli intermixed (mixed-colour condition) with the performance in blocks with only a single colour for all stimuli (monochromatic condition), expecting a stronger colour-valence congruence effect in the mixed-colour condition than in the monochromatic condition.

As our second test of the polarity-correspondence principle, we checked our data for the distinct *asymmetrical* response (time) pattern that is typically predicted in similar bimanual cross-modal classification tasks (Lakens, 2012). The reason of the unequal congruence effect sizes

[3] Some studies that tested not exactly colour-emotion congruence effects themselves but related effects, such as influences of specific colour on

(*footnote continued*)
attractiveness ratings of faces of varying attractiveness rather than of different valence stimuli, used rating tasks and allowed even intermediate or neutral judgments (e.g., Elliot & Niesta, 2008). Though in these tasks, the mere presence of stimuli on opposite poles of dimensions (e.g., of + and – pole stimuli on an attractiveness dimension) might have been critical for the found colour-based effects, this question is beyond the scope of the present study and not central to our major tests.

can be explained by the following rationale: to comply with any arbitrary binary classification instructions, participants can assign *priority* to one of the two available alternatives; that is, they can look for the presence of the prioritized target characteristic (say, for stimuli above as + pole attribute), responding "yes" to the + pole attribute when the search criterion is fulfilled (stimulus *is* above) and responding "no" to the + pole when the searched-for feature is absent (stimulus *is not* above); or participants can instead switch to the alternative task of searching for (and responding "yes" to) the less prioritized characteristic (stimulus *is below* fixation). In either case, for – pole but not for + pole stimuli, switching between responses (see Bertelson, 1965) or between tasks (Rogers & Monsell, 1995; Wolfe, Butcher, Lee, & Hyle, 2003) should delay responses. In line with this assumption, where participants consistently assign + poles and – poles to different specific target alternatives, + pole responses (e.g., to stimuli above fixation) are faster than – pole responses (e.g., to stimuli below fixation; see Clark, 1969; Lakens, 2012; Proctor & Cho, 2006). In the following, we refer to this influence brought about by the polarity-correspondence principle as + *pole response facilitation*.

Importantly in the context of the present investigation of colour-valence associations, participants apparently assign poles to additional conceptual dimensions of the targets besides their relevant features, as long as these additional conceptual dimensions also vary in a binary way. This is what Proctor and Cho (2006) refer to as structural overlap, which occurs when an experimental design allows for a symmetrical mapping across (conceptual/perceptual) dimensions. Returning to the previous example: when participants classify target words appearing above or below a fixation point (variation on the spatial dimension) as positive or negative (variation on the valence dimension), responses are faster to positive targets above fixation than to negative targets above fixation (Meier & Robinson, 2004). Theoretically, this could be due to context-independent semantic priming of the task-irrelevant stimulus position by the relevant emotional dimension (i.e., 'good is up' and 'bad is not up/bad is down'). However, in line with the polarity-correspondence principle, these space-valence congruence effects can be missing or even reversed for + pole words below fixation (technically an incongruent condition) compared to – pole words below fixation (technically a congruent condition). As explained by Lakens (2012), + pole response facilitation according to the polarity-correspondence principle can account for this asymmetric congruence effect (i.e., the presence of the congruence effect for + pole stimuli but its absence or diminution for – pole stimuli). First of all, indicative of additional facilitation through the + pole responses on top of the facilitation by polar congruence alone, Lakens found that responses to congruent +/+ pole stimuli (here: positive words above fixation) were significantly faster than to congruent −/− pole stimuli (here: negative words below fixation). Secondly, however, in the case of Lakens' (2012) meta-analysis of experiments on space-valence associations, this + pole response facilitation was so strong as to even facilitate responses to incongruent +/− pole stimuli (here: positive words below fixation) relative to congruent −/− pole stimuli (here: negative words below fixation). This abolished the congruence effect under – pole conditions. This asymmetry of the strength of the congruence effect is, thus, also an indication of the polarity-correspondence principle, as it directly follows from + pole response facilitation. To test if polarity correspondence contributes to the *colour*-valence congruence effect like it does to *position*-valence congruence, in the current study, we investigated if we find a similar asymmetry of the congruence effect for red and green positive and negative word targets.

1.3. Congruence sequence effect

For our test of the asymmetry of the colour-valence congruence effect, we also wanted to rule out one alternative explanation: Past research suggested that congruence in the preceding trial ($n - 1$) and emotional valences can jointly influence the congruence effect in a current trial (n). In general, congruence effects are stronger following congruent than following incongruent trials (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Gratton, Coles, & Donchin, 1992; Huber-Huber & Ansorge, 2017; Puccioni & Vallesi, 2012). These *congruence sequence effects* are at least partly brought about by cognitive control: an active suppression of the irrelevant dimension and more attention to the relevant dimension following an incongruent/conflicting trial than following a congruent trial without conflict (e.g., Huber-Huber & Ansorge, 2017). Importantly, this kind of cognitive control critically depends on the negative affect elicited by the conflict (Hobson, Saunders, Al-Khindi, & Inzlicht, 2014; Inzlicht, Bartholow, & Hirsh, 2015; Schouppe et al., 2015; van Steenbergen, Band, & Hommel, 2009, 2010). First of all, conflict elicits negative affect as indicated by priming of negative judgments through conflicting stimuli (cf. Dreisbach & Fischer, 2012; Fritz & Dreisbach, 2013; Goller, Kroiss, & Ansorge, 2019). Secondly, modifying the affective response to the conflicting stimulus changes the amount of exerted cognitive control (e.g., De Bruijn, Hulstijn, Verkes, Ruigt, & Sabbe, 2004; Hobson et al., 2014; van Steenbergen et al., 2009, 2010). For example, administration of an anxiolytic drug diminishes the error-related negativity (ERN) that follows an erroneous response in a conflict task (de Bruijn et al., 2004). Likewise, van Steenbergen et al. (2009) found reduced congruence sequence effects if participants received a monetary gain. The upshot of such investigations is that the valence of the target stimulus presented following a conflicting trial could theoretically influence the size of conflict regulation and, thus, the average size of the congruence effect, too. Following a preceding incongruent trial n-1, presenting a negative target word in a current trial n could semantically prime conflict-elicited negative affect and, hence, conflict regulation. In contrast, presenting a positive target word in a current trial n would not semantically prime and possibly even counter-act conflict-elicited negative affect and, hence, diminish cognitive control. As a consequence, the asymmetry of the congruence effect predicted by the polarity-correspondence principle – stronger congruence effects with positive than with negative words – could also be due to more cognitive control and resulting smaller average congruence effects with negative than positive target words. This would be reflected in an interaction between (1) valence, (2) congruence in the preceding trial, and (3) congruence in the current trial: lower congruence effects following incongruent trials, more so with negative than with positive words. Importantly, such an influence goes hand in hand with the predicted two-way interaction between word valence and congruence (in the current trial): a diminution of the average congruence effect for negative words compared to positive words, but for different reasons than assumed by the polarity-correspondence principle.

In an explorative analysis, this potential alternative explanation of the congruence-effect asymmetry was therefore tested, too. To that end, we analysed congruence effects as a function of congruence in a preceding trial (congruence in $n - 1$). Only when this variable is included in our analysis can we eventually decide if the polarity-correspondence principle or the valence-dependency of conflict regulation that we sketched above accounts better for the expected asymmetry of the congruence effect. Our criterion for this test will be the presence versus absence of the significant three-way interaction between target valence, congruence in trial $n - 1$, and congruence in trial n. This interaction should be significant if congruence between the affective state and preceding conflict account for the asymmetry of the congruence effect. In contrast, the three-way interaction should be absent when the polarity-correspondence principle accounts for the expected asymmetry of the congruence effect. According to our knowledge, no polarity-correspondence study has tested for this potential alternative explanation to date. This is why we found it necessary to analyse congruence sequence effects as a function of valence.

2. Methods

2.1. Participants

Based on effect sizes in a previous pilot similar to the current study (Lohmann & Jorschick, 2015) of Cohen's $f = 0.86$, for a power of 0.95, $\alpha = 0.05$, the suggested minimum sample size (using the G*Power tool; Faul, Erdfelder, Buchner, & Lang, 2009) is 20 participants (for the mixed colour condition). Taking into consideration that in the monochromatic blocks, Colour was a between-participants, not a within-participant variable, we estimated that 40–60 participants per condition would be sufficient to gauge whether any differences are present in the conditions. Keeping the participant number comparable across the three conditions, we aimed for a total of 120–180 participants.[4]

In total, 145 students at University of Vienna (27 men, $M_{age} = 20.86$ years, $SD_{age} \pm 2.85$) participated in exchange for course credit. Participants were randomly assigned to one of the block order conditions (see Table 1).

Table 1
Block order and number of participants assigned to each experimental condition.

Condition	1st block	2nd block	3rd block	*N*
1	Mono red	Mono green	Mixed	50
2	Mono green	Mono red	Mixed	47
3	Mixed	Mono red	Mono green	23
4	Mixed	Mono green	Mono red	24
Total				145

Note. Our main interest lay in effects in the first experimental block, in which case Conditions 3 and 4 were identical (both starting with mixed blocks). We therefore collected for Conditions 3 and 4 in sum approximately the same number of participants as we did for Condition 1 and Condition 2 respectively.

2.2. Materials

We selected 60 positive and 60 negative German words from the Berlin Affective Word List Reloaded (BAWL-R; Võ et al., 2009). We classified words as positive when the mean emotion value was ≥ 0.6 and as negative when the mean emotion value was ≤ -0.6. The number of nouns, verbs and adjectives was balanced, and values for word arousal, imageability, letter count and word frequency (Brysbaert et al., 2011) were kept constant across the two valence categories (see Table 2). A list with all stimuli used is provided in Appendix A. To ensure equiluminance of stimulus colour, brightness values for red, green, and grey (background colour) were measured with a spectrophotometer (X-Rite i1XTreme, Grand Rapids, MI, USA) for each of the five monitors used and colour values were selected accordingly (see Table 3 for the L*C*h* values, as referenced in the CIE LCh colour space; a detailed list with the values per monitor is provided in the Materials section on osf.io/c4zry). Instruction colour, which was used for all text not constituting experimental stimuli, was in a light blue hue, and was, depending on the luminance-controlled colours picked for each monitor, calculated by averaging the reds and greens and inverting the result for better readability against the grey background. The size of the coloured word stimuli was set to 50 pixels (angular size 1.45°), with a fixed viewing distance between eye and centre display of 60 cm, assured via the utilization of chin rests.

2.3. Apparatus

During the study, stimuli were presented on five 19-inch LCD monitors with a 1280 × 1024 resolution and a refresh rate of 60 Hz. Responses and associated response times (RTs) were registered via serial QWERTZ computer keyboards.

2.4. Procedure

Testing was conducted in a dimly lit laboratory room. Upon arrival, participants were welcomed and placed at one of five identical PCs, running PsychoPy2 software (Peirce, 2007). After signing the consent forms, participants were instructed to place their chin on the chin rest and start the experiment. Instructions were presented on screen. The study consisted of two tasks: an initial valence-rating task and a subsequent binary valence-classification task. The experimental session ended with a short test for colour deficiency.

2.4.1. Valence-rating task

To ensure that the valence provided in the BAWL-R database matched participants' perceived valence, a rating task (‘*Please rate the valence of the word.*’) was employed before the binary classification task. Participants judged each of 120 potential target words once. Per each trial, one word was presented in the instruction colour, in capital letters, and in randomised order, centred above a 10-point rating scale with the endpoints being labelled *very negative* (‘sehr negativ’) on the left side of the screen and *very positive* (‘sehr positiv’) on the right side. Participants had to move the mouse cursor to the tick mark corresponding with their rating and confirm the selected valence value with a left-click. Words that received a rating below six were classified as ‘rated negative’ and appeared in a text box on the left side of the screen, below the negative scale label; words with a rating of six or higher were coded as ‘rated positive’ and appeared in a text box on the right side of the screen, below the positive scale label. Each word was presented for a variable duration (until the judgment was made). There were no time constraints for this task and participants were informed thusly. After participants' rating of all words, the 50 most positively and the 50 most negatively rated words were selected for the upcoming valence-categorisation task. If the number of words per category was lower than 50, the shortest word list (list of words rated positive or list of words rated negative by the participant) determined the overall number of stimuli that appeared in the subsequent task. For an equal number of positive and negative items, the word list of the remaining valence category was trimmed accordingly. For example, if a participant rated 43 of the 60 positive words (per BAWL-R database) as positive and 52 of the 60 negative words as negative, then the 43 positive-rated words were selected for the participant's valence-categorisation task, as well as the same number of negative-rated words, in which case the total number of items would be 43 + 43 = 86 items.[5] The maximum of stimuli presented in the categorisation task was 100, while there was no minimum of items specified in the experiment. For data analysis, however, we set a lower boundary at a minimum of 40 words as participant exclusion criterion (see Data collection and analysis).

[4] In the preregistration, we stated that we opened 180 slots for participation (i.e., giving a maximum of 180 people the possibility to take part in our experiment). Beforehand, it was not possible to say how many participants would sign up in the university's participant-registration system, but typically, at least around two thirds of all slots are filled up. The smallest targeted sample size (see Methods) was approximately 40 participants per group. Testing took place within a span of 3 days. Out of the 180 open spaces during these 3 testing days, a total of 145 people could be tested in actuality.

[5] Note that the number of positively rated (e.g., 43) and negatively rated words (e.g., 52) does not necessarily have to add up to 120. In the example, 43 out of 60 positive words were “correctly” judged by the participant as, in fact, positive, while 17 of the 60 positive words were judged as negative. Furthermore, the participant judged 52 out of the 60 negative words as, in fact, negative, but judged eight of the negative words as positive (e.g., ‘naïve’ is a negative term according to the BAWL-R database, but could be judged as a more positive concept by some participants).

Table 2
Number of positive and negative stimuli per word category and mean values for valence, arousal, imageability and letters taken from the Berlin Affective Word List Reloaded (BAWL-R) database, and log word frequencies (Subtlex Lg) taken from the SUBTLEX database.

	Emotion value	Arousal value	Image value	Letters	Frequency	Nouns	Verbs	Adjectives
Positive	1.67	2.98	3.94	6.83	2.37	20	20	20
Negative	−1.53	3.12	3.98	6.77	2.29	20	20	20

Table 3
Average values for lightness (L*), chroma (C*), and hue (h* specified in degrees) (according to the CIE LCh colour model) of the red, green, and grey colour used in the experiment. Saturation (sat) is computed as C*/L* (see Wilms & Oberfeld, 2018).

	L*	C*	h*	Sat
Red	51.17	102	40.14	1.99
Green	51.81	74.03	134.39	1.43
Grey	35.37	0	145.10	0

2.4.2. Binary valence-classification task

The binary classification task started after participants had read the instructions on the monitor, which informed them that, per each trial, they would be presented with a single word (which they had previously seen in the valence-rating task) at screen centre. Each target was shown for a maximum of 2 s. For each word, participants judged the valence (*Is this word positive or negative*?) and indicated their choice by pressing either the 'E' (for positive valence) or the 'I' (for negative valence) key on a standard QWERTZ keyboard. Key assignment was balanced across participants. We did not mention to participants beforehand that words would be presented in different colours. In addition to an initial 10-trial practice, the binary classification task comprised three experimental blocks in total: one block with each word (out of the max. 100) presented once in only red colour ('mono red'), one block with each word presented once in green ('mono green'), and a mixed-colour block with each word presented twice, once in red and once in green font colour. Depending on the participants' assigned condition, order of blocks varied (see Table 1). Stimuli were presented in randomised order, with the restrictions that not more than five words of the same valence (in all blocks) and colour (in the mixed blocks) were presented in a row. Participants completed the entire experiment in < 30 min.

2.4.3. Test for colour deficiency

After participants completed the valence rating and all three blocks of the binary valence-classification task, they were asked to enter the numbers printed on three colour plates, which were displayed on the computer screen (pictures are provided as supplementary material under osf.io/c4zry). These pictures were computer-screen versions of the Ishihara colour plates. Note that this is a non-standardized assessment of the Ishihara test for colour deficiency and may not give accurate results as for the presence or absence of colour-blindness. However, since our experimental setup also made use of computer-display colours, we opted to include this control measure, in addition to self-reported colour deficiencies. Upon entering all three numbers, participants were debriefed in written form and the experiment ended.

2.5. Data collection and analysis

Data from the previous rating task was not analysed further, since it only served to ensure the participants' valence judgment. Data from all three blocks of the valence-classification task was analysed, but analyses reported below were restricted to the first experimental blocks from the valence-categorisation task because by the time of commencing mixed-colour blocks, colour polarity was established and blocks in the monochromatic condition were no longer free from potential transfer confounds. However, we provide analyses of the data from all three experimental blocks on the project's OSF page osf.io/c4zry. We also ran a more conservative analysis of the mixed colour blocks, where colour switch trials (a red word following a green word, or vice versa) were excluded from the data. We did this to ensure that congruence effects do not only result from colour switch trials but are also present in colour-repetition trials.

Criteria for data exclusion were as follows: Data from participants who did not reach the full score in our colour-deficiency test or were self-reportedly colour-blind were discarded (four participants). Furthermore, if participants, during the rating task, classified < 40 items in each valence category in accordance with the valence given in the BAWL-R database, their data was not analysed further (one participant). The exclusion criterion for participants' accuracy was 75%, but since no participant had an overall accuracy rate lower than this threshold, no further participants had to be excluded, leaving a total of $n = 140$ participants. Mean error rate (ER) over all blocks of the categorisation task was 6.0%. Trials with RTs below 150 ms and above 2 s were excluded from analysis (< 1%). In the congruence sequence effect analyses, trials with previously erroneous or too slow answers were excluded (e.g., if the response to trial *n-1* was incorrect and thus excluded from analysis, trial *n* was excluded as well). For RT analyses, correct median RTs were taken, as the median is less sensitive to disproportionately slow responses than the mean. For error analyses, arcsine square root transformed error rates were used, as this stabilises and normalises variances of proportional data (Sokal & Rohlf, 1981). Additionally, in Tables 4 and 5, we report the untransformed mean ERs for an easier understanding of means over conditions.

2.5.1. Effect sizes and statistical significance

To demonstrate the magnitude of the observed effects, partial eta-squared (η_p^2) values, 90% confidence intervals (CI), and generalised eta-squared (η_G^2) are reported for *F*-tests (Steiger, 2004). We report Bayes factors (as BF_{10} when supporting difference, and BF_{01} when supporting equivalence) using the default *r*-scale of 0.707 (Morey & Rouder, 2018). In case of analyses of variance (ANOVAs), we report inclusion BFs based on matched models (Makowski, Ben-Shachar, & Lüdecke, 2019; Mathôt, 2017). Where applicable, we report Welch-corrected *t*-tests (Delacre, Lakens, & Leys, 2017) with corresponding Cohen's *d* values (Lakens, 2013). We used the conventional alpha level of 0.05 for all statistical significance tests. All analyses were conducted in R (R Core Team, 2019; via: Kelley, 2019; Lawrence, 2016; Lukács, 2019; Morey & Rouder, 2018).

Table 4
Means (*SD* in parentheses) of the median reaction times (MRT; in ms) and mean errors (MER; in %) of the first experimental block per Colour and Valence conditions.

Colour	Valence	Monochromatic		Mixed Colour	
		M_{RT} (*SD*)	M_{ER} (*SD*)	M_{RT} (*SD*)	M_{ER} (*SD*)
Green	Positive	689 (86)	4.8 (4.2)	691 (78)	3.0 (4.1)
	Negative	718 (85)	8.1 (5.3)	755 (82)	9.8 (7.8)
Red	Positive	722 (93)	4.9 (3.9)	738 (82)	8.1 (5.8)
	Negative	739 (93)	6.6 (5.4)	725 (87)	3.1 (3.0)

Table 5
Means (SD in parentheses) of the median reaction times (MRTs, in ms) and the untransformed error rates (MERs, in %) in the mixed-colour condition per valence, N congruence, and *N − 1* congruence.

Valence	Trial *N − 1*	Trial *N*	M_{RT} (*SD*)	M_{ER} (*SD*)
Positive	Congruent	Congruent	679 (81)	2.7 (3.9)
	Incongruent	Congruent	702 (85)	3.6 (6.0)
	Congruent	Incongruent	737 (88)	9.3 (7.5)
	Incongruent	Incongruent	734 (101)	7.0 (6.5)
Negative	Congruent	Congruent	706 (81)	1.9 (2.8)
	Incongruent	Congruent	736 (97)	4.4 (4.7)
	Congruent	Incongruent	752 (80)	11.9 (9.3)
	Incongruent	Incongruent	749 (120)	7.5 (8.2)

3. Results

Five participants were excluded because they violated the inclusion criteria for data analysis as stated above. Key assignment influenced neither RTs nor ERs (all $F < 1$) and is not discussed further.

We conducted the preregistered analyses with minor, necessary modifications explicitly specified for the respective tests. The preregistration and all supplementary materials, together with the statistical analyses of all-block data and a more conservative analysis of the mixed-colour blocks including only colour-repetition trials (no colour switch trials), are available under the project's Open Science Framework (Foster & Deardorff, 2017) page osf.io/c4zry. By and large, the more conservative analysis of only the colour repetition trials in the mixed-colour condition as well as the all-block analysis confirmed the results of the present analyses.

3.1. Monochromatic condition

If polarity on two dimensions is required for congruence effects to emerge, we would expect no congruence effect in the red and green monochromatic conditions, where colour does not vary across trials. In contrast, if polarity correspondence has no effect on the weight of the colours as modifiers of valence processing, then a more typical colour-valence congruence effect, with green-positive and maybe weaker red-negative associations, should show in the monochromatic data. Means of median correct RTs and ERs were calculated per valence, colour, and block condition and are given in Table 4. Figs. 1 and 2 illustrate the RTs and ERs, respectively, with results for the monochromatic blocks printed on the left side, and results for the mixed-colour blocks printed on the right.

3.1.1. Reaction times (RTs)

Our analysis was restricted to the participants that started with either the red or the green monochromatic block (see Table 1, Conditions 1 and 2) because the performance of these participants was free from any potential carry-over of alternative polar assignments of the colours from a preceding mixed colour-block. We conducted a two (between-participants Colour: red vs. green) × two (within-participant Valence: positive vs. negative) mixed-design ANOVA on the correct binary classification times (RTs) of the monochromatic (mono red, mono green) blocks. Colour did not influence RTs significantly, $F(1, 93) = 2.29$, $p = 0.134$, $\eta_p^2 = 0.024$, 90% CI [0, 0.095], $\eta_G^2 = 0.022$, $BF_{10} = 1.40$. There was a significant main effect for Valence, $F(1, 93) = 22.79$, $p < 0.001$, $\eta_p^2 = 0.197$, 90% CI [0.088, 0.308], $\eta_G^2 = 0.016$, $BF_{10} = 2766.49$, with positive words being responded to faster ($M_{RT} = 706$ ms) than negative words ($M_{RT} = 729$ ms), a finding in line with + pole response facilitation (Lakens, 2012).[6] The interaction between Valence and Colour was not significant, $F(1, 93) = 1.41$, $p = 0.237$, $\eta_p^2 = 0.015$, 90% CI [0, 0.078], $\eta_G^2 = 0.001$, $BF_{01} = 2.98$. This is in line with the predicted absence of polarity-correspondence effects where only *one* of the involved dimensions (i.e., valence) varied but the other (i.e., colour) did not.

RTs to red positive words were slower than RTs to green positive words, but RTs to red targets still showed the typical + pole response facilitation for positive red relative to negative red words. Thus, we do not find any evidence for an automatic red-negativity link. This could be due to a lack of any colour-valence congruence effect in the absence of opposing green colours in the monochromatic red blocks. However, this could have also reflected a general red-positivity association of maybe somewhat smaller size than in the case of green stimuli. To investigate, if there could be red-positivity and green-positivity links, we ran two paired *t*-tests, one for red-monochromatic and one for the green-monochromatic conditions, where we compared the RTs for negative versus positive words.

Despite our decent sample size, in the red-monochromatic condition, there was only weak evidence for a difference in RTs, $t(49) = 2.41$, $p = 0.020$, $d_{within} = 0.34$, 95% CI [0.05, 0.62], $BF_{10} = 2.11$. In the green-monochromatic condition, evidence for a difference in RTs was strong, $t(44) = 4.72$, $p < 0.001$, $d_{within} = 0.70$, 95% CI [0.37, 1.03], $BF_{10} = 861.62$.

In light of the relatively weak evidence in the red-monochromatic condition, we cannot confidently corroborate the hypothesis that red is associated with positive valence. This lack of an automatic red-positivity or an automatic red-negativity link in a non-polarity context could be an indication for an ambiguous role of the colour red when it comes to valence associations, when it is perceived in isolation instead of in opposition to another colour. For an illustration of the RT data from the monochromatic conditions see Fig. 1, left panel.

3.1.2. Error rates (ERs)

We computed an ANOVA with the same variables on the arcsine transformed ERs. Results corroborated the RT findings. Colour did not influence ERs significantly, $F(1, 93) = 0.64$, $p = 0.426$, $\eta_p^2 = 0.007$, 90% CI [0, 0.059], $\eta_G^2 = 0.005$, $BF_{01} = 3.31$. There was a significant main effect for Valence, $F(1, 93) = 21.01$, $p < 0.001$, $\eta_p^2 = 0.184$, 90% CI [0.078, 0.295], $\eta_G^2 = 0.066$, $BF_{10} = 1238.00$, with less accurate responses to negative ($M_{ER} = 7.4\%$) than to positive stimuli ($M_{ER} = 4.9\%$). The interaction between Colour and Valence was not significant, $F(1, 93) = 2.37$, $p = 0.127$, $\eta_p^2 = 0.025$, 90% CI [0, 0.097], $\eta_G^2 = 0.008$, $BF_{01} = 1.69$. For an illustration of the ER data from the monochromatic conditions see Fig. 2, left panel.

3.2. Mixed-colour condition

Above, we have shown that the Colour × Valence interaction is *not* significant in case of monochromatic blocks. Now we tested the same interaction in case of mixed-colour blocks (again, only for the conditions where the mixed-colour block was the first experimental block, i.e., Conditions 3 and 4). Since both colours are present during the mixed-colour block, colour and valence were within-participant variables. Contrary to the monochromatic blocks, in the mixed-colour condition, colour polarity is present, since stimuli are presented intermixed in either red or green. In such a context, the polarity correspondence principle (Lakens, 2012) predicts the prevalence of congruence effects (i.e., better performance in congruent than incongruent conditions), in the way that through polar assignments, green should share the + pole with positive valence and red should share the – pole with negative valence. Such a congruence effect should then be

[6] For the + pole response facilitation, it is not necessary that both dimensions contributing to a polarity correspondence effect are realized within the same context. As long as there are polar opposites regarding one dimension, here of valence, + pole facilitation for this dimension can be observed.

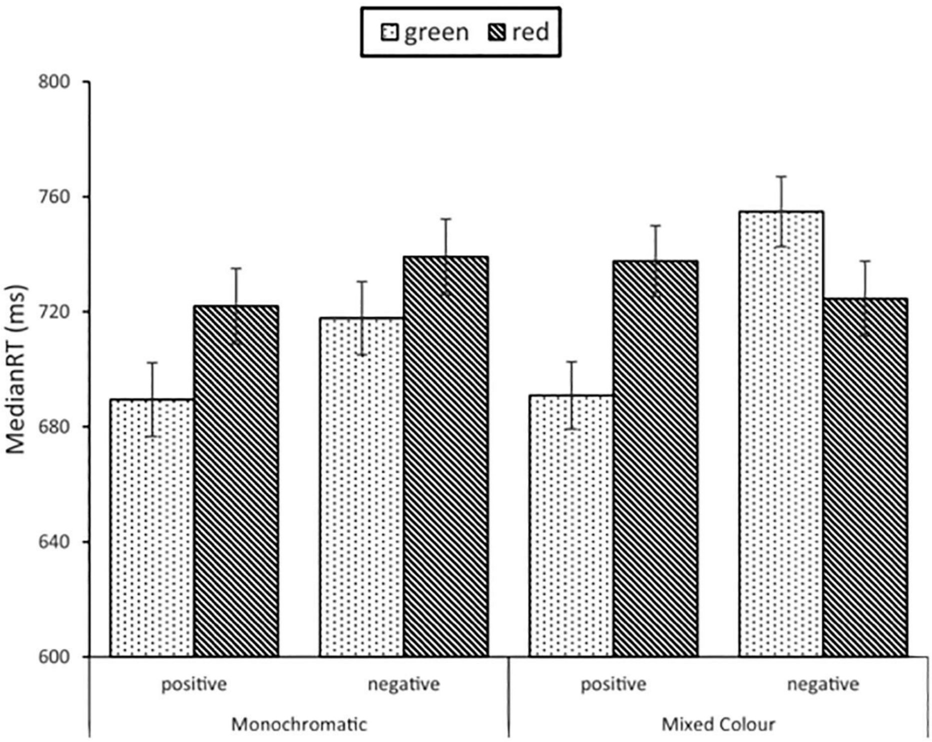


Fig. 1. Median reaction times (RTs) for positive and negative words presented in red and green font colour in the mixed-colour versus monochromatic conditions. Error bars show the *SE* of the means.

reflected in the form of a significant interaction between Colour and Valence.[7]

3.2.1. Reaction times (RTs)

The results showed a main effect for Valence, $F(1, 44) = 30.71$, $p < 0.001$, $\eta_p^2 = 0.411$, 90% CI [0.220, 0.545], $\eta_G^2 = 0.024$, $BF_{10} = 1601.38$, and a weaker main effect for Colour, $F(1, 44) = 5.08$, $p = 0.029$, $\eta_p^2 = 0.103$, 90% CI [0.006, 0.252], $\eta_G^2 = 0.003$, $BF_{01} = 2.30$. In line with the expectations based on the influence of the polarity-correspondence principle, the Colour × Valence interaction was significant, $F(1, 44) = 61.81$, $p < 0.001$, $\eta_p^2 = 0.584$, 90% CI [0.413, 0.684], $\eta_G^2 = 0.053$, $BF_{10} = 4.26 \times 10^{11}$. Two paired *t*-tests showed the predicted congruence effect in line with previous research: In green colour, positive words were classified faster than negative words, $M_{RT} \pm SD = 690.91 \pm 78.43$ vs. 754.81 ± 81.69, respectively, with $t(44) = 10.39$, $p < 0.001$, $d_{within} = 1.55$, 95% CI [1.11, 1.98], $BF_{10} = 3.88 \times 10^{10}$, while in red colour, negative words were classified faster than positive words, $M_{RT} \pm SD = 724.58 \pm 87.31$ vs. 737.62 ± 82.46, respectively, with $t(44) = -1.81$, $p = 0.078$, $d_{within} = -0.27$, 95% CI [−0.57, 0.03], $BF_{01} = 1.39$. Although the latter post-hoc test failed to reach significance, the trend was in the expected direction, in line with our hypothesis of red-negative stimuli being responded to faster than red-positive stimuli, when opposed to green colour. Together, these stronger congruence effects for the green than for the red target words in the mixed-colour condition confirmed the presence of a congruence-effect asymmetry with stronger effects for the + pole than for the – pole stimuli.

The results for the RT data from the mixed-colour conditions are illustrated in Fig. 1, right panel.

3.2.2. Error rates (ERs)

We conducted an ANOVA with the same variables on the arcsine transformed ERs. Neither Valence nor Colour elicited a significant main effect, $F(1, 44) = 1.22$, $p = 0.275$, $\eta_p^2 = 0.027$, 90% CI [0, 0.142], $\eta_G^2 = 0.005$, $BF_{01} = 4.33$, $F(1, 44) = 0.02$, $p = 0.882$, $\eta_p^2 < 0.001$, 90% CI [0, 0.036], $\eta_G^2 < 0.001$, $BF_{01} = 6.23$, respectively. The interaction between Colour and Valence was significant, $F(1, 44) = 78.39$, $p < 0.001$, $\eta_p^2 = 0.640$, 90% CI [0.483, 0.727], $\eta_G^2 = 0.248$, $BF_{10} = 1.04 \times 10^{14}$. We performed two paired *t*-tests on the arcsine transformed ERs. They corroborated the expected congruence effect: In green colour, positive words were classified more accurately than negative words, $M_{ER\ asin} \pm SD = 0.13 \pm 0.12$ vs. 0.29 ± 0.15, respectively, with $t(44) = 6.65$, $p < 0.001$, $d_{within} = 0.99$, 95% CI [0.63, 1.34], $BF_{10} = 3.62 \times 10^5$, while in red colour, negative words were classified more accurately than positive words, $M_{ER\ asin} \pm SD = 0.15 \pm 0.10$ vs. 0.27 ± 0.11, respectively, with $t(44) = -5.94$, $p < 0.001$, $d_{within} = -0.89$, 95% CI [−1.23, −0.54], $BF_{10} = 3.84 \times 10^4$. Unlike the RTs, error rates showed a quite symmetrical cross-interaction between Colour and Valence, and consequently, the size of the congruence effects. The results for the error data from the mixed-colour conditions are illustrated in Fig. 2, right panel.

3.3. Congruence effect sizes

To quantify the impact of polarity on the congruence effect, we ran two separate analyses, one for red stimuli and one for green stimuli; this means, we restricted the data taken from the mixed-colour blocks to only one of the colours in turn and contrasted it with the data from the

[7] This ANOVA for the mixed blocks was not preregistered. Originally, we preregistered a three-way mixed model ANOVA on the conjoined first block data (from the mixed and monochromatic conditions), the variables being Valence (within participant), Colour-Valence Congruence (within) and Colour Condition (monochromatic vs. mixed, between) if the first ANOVA for the monochromatic blocks shows a Colour × Valence interaction (i.e., a congruence effect). The reasoning was that the effects for the mixed blocks are discernible from the two preregistered analyses of only the monochromatic condition and mono- and mixed-coloured condition together. However, this analysis proved to be not feasible, as it relies partly on within-participant and partly on between-participants data from the same participants. For this reason, we opted for the two separate analyses.

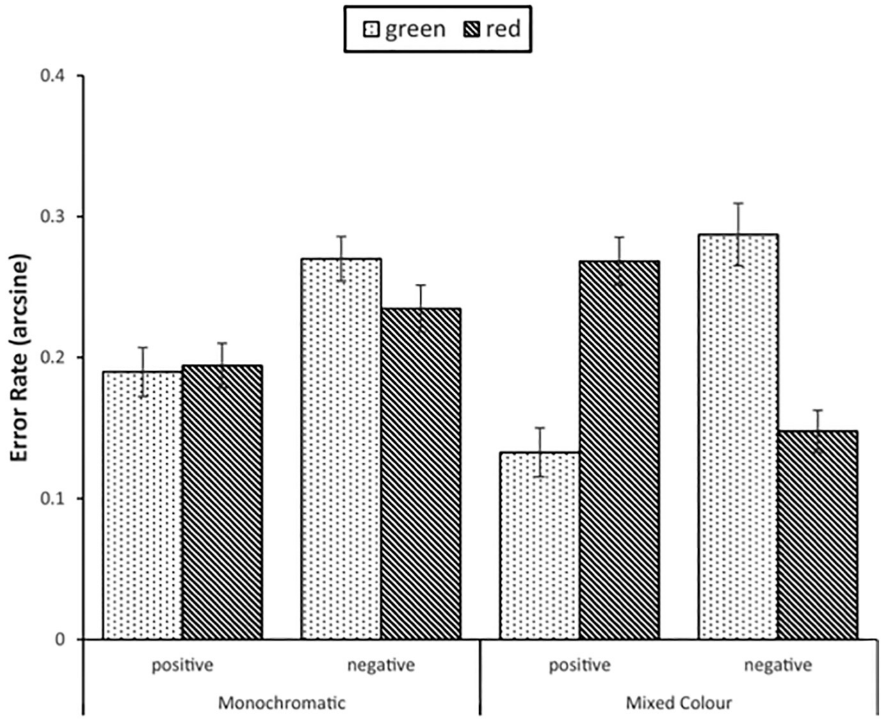


Fig. 2. Arcsine transformed error rates for positive and negative words presented in red and green font colour in the mixed-colour versus monochromatic conditions. Error bars show the *SE* of the means.

monochromatic condition of the corresponding colour (green-monochromatic versus green-mixed trials; red-monochromatic vs. red-mixed trials). This was done, as the separate ANOVAs for monochromatic blocks and mixed-colour blocks reported above differed from each other by how colour was operationalised – as a between-participants variable (in the monochromatic blocks) or as a within-participant variable (in the mixed-colour blocks) – and as this might have affected the power or sensitivity of the designs differently in the two blocks (with arguably more sensitivity for any interactions in the mixed-colour than in the monochromatic blocks). Following the results from Analysis 3.2, for each colour, we coded the trials for Congruence in the way that for green, positive trials were congruent and negative trials were incongruent; conversely, for red, negative trials were congruent trials and positive trials were incongruent.

If polarity, which is only present in the mixed-colour but not in the monochromatic condition, affects responses to stimuli, we would again expect trials from the mixed-colour condition to show a larger congruence effect (difference between incongruent and congruent trials) than from the monochromatic condition, respectively, for each of the two colours.

3.3.1. *Red-monochromatic versus red-mixed trials*

3.3.1.1. Reaction times (RTs). A two (within-participant Congruence: congruent vs. incongruent) × two (between-participants Condition: monochromatic-red vs. mixed-red) mixed variables ANOVA was conducted on the correct RTs from all red-coloured target trials from the first experimental block, comparing Condition 1 (monochromatic red) to Conditions 3 and 4 (mixed-coloured). The results showed no main effect for Congruence, $F(1, 93) = 0.31$, $p = 0.576$, $\eta_p^2 = 0.003$, 90% CI [0, 0.047], $\eta_G^2 < 0.001$, $BF_{01} = 5.63$, and no main effect for Condition, $F(1, 93) = 0.00$, $p = 0.972$, $\eta_p^2 < 0.001$, 90% CI [0, < 0.001], $\eta_G^2 < 0.001$, $BF_{01} = 2.43$. Importantly, despite partly resting on a between-participants variable, the Congruence × Condition interaction was significant, $F(1, 93) = 8.83$, $p = 0.004$, $\eta_p^2 = 0.087$, 90% CI [0.017, 0.185], $\eta_G^2 = 0.007$, $BF_{10} = 9.79$. As for congruence effects' sizes: Incongruent minus congruent RTs showed an effect of 13 ms for red in the mixed-colour condition, demonstrating shorter RTs to congruent (negative) than incongruent (positive) stimuli. In contrast, an “inverted congruence effect” of −17 ms, with faster responses to incongruent than to congruent trials, was found in the red monochromatic condition. The negative effect reflected the typical + pole response facilitation of the valence targets in the absence of the antagonistic colour pole (here: green), showing faster responses to positive (here incongruent) than negative (here congruent) words. Alternatively, as mentioned above, the responses to red targets in the monochromatic blocks might have reflected a red-positivity association that can only be unveiled once the context of an antagonistic colour pole is missing.

3.3.1.2. Error rates (ERs). ERs showed no significant main effect for Condition, $F(1, 93) = 0.13$, $p = 0.723$, $\eta_p^2 = 0.001$, 90% CI [0, 0.036], $\eta_G^2 = 0.001$, $BF_{01} = 4.59$, but a significant main effect for Congruence, $F(1, 93) = 6.80$, $p = 0.011$, $\eta_p^2 = 0.068$, 90% CI [0.009, 0.161], $\eta_G^2 = 0.026$, $BF_{10} = 1.99$. Importantly, the interaction between Congruence and Condition was significant, $F(1, 93) = 34.14$, $p < 0.001$, $\eta_p^2 = 0.269$, 90% CI [0.147, 0.379], $\eta_G^2 = 0.118$, $BF_{10} = 3.36 \times 10^5$, demonstrating also in the error data a larger congruence effect in the red trials from the mixed-colour condition, with a difference of 5.0%, than in the red trials from the monochromatic condition, with a difference of −1.7%.

3.3.2. *Green-monochromatic versus green-mixed trials*

3.3.2.1. Reaction times (RTs). A two (within-participant Congruence: congruent vs. incongruent) × two (between-participants Condition: monochromatic-green vs. mixed-colour green) mixed variables ANOVA was conducted on the RTs from all green-coloured trials from the first experimental block comparing Condition 2 (monochromatic green) to Conditions 3 and 4 (mixed-coloured). The main effect for Condition was not significant, $F(1, 88) = 1.29$, $p = 0.258$, $\eta_p^2 = 0.014$, 90% CI [0, 0.079], $\eta_G^2 = 0.014$, $BF_{01} = 1.51$. The main effect for Congruence was significant, $F(1, 88) = 115.14$, $p < 0.001$, $\eta_p^2 = 0.567$, 90% CI [0.451, 0.646], $\eta_G^2 = 0.073$, $BF_{10} = 6.53 \times 10^{12}$. Importantly, the interaction between Congruence and Condition was also significant, $F(1, 88) = 17.12$, $p < 0.001$, $\eta_p^2 = 0.163$, 90% CI [0.060, 0.275], $\eta_G^2 = 0.012$, $BF_{10} = 222.53$. RT difference comparison (mean of incongruent minus mean of congruent trials) showed that the size of the congruence effect in the mixed-colour condition was 64 ms, while in the green monochromatic condition it was only 29 ms. This latter RT difference was probably no congruence effect at all, as it could have reflected the typical + pole response facilitation of the positive stimuli in general. However, the residual facilitation of RTs to green positive relative to green negative words that we found in the monochromatic blocks could have likewise reflected the more consistent coding of the colour green as positive, regardless of whether or not the context included an antagonistic colour pole (here: red).

3.3.2.2. Error rates (ERs). ERs showed a similar pattern. The main effect for Condition was not significant, $F(1, 88) = 0.87$, $p = 0.355$, $\eta_p^2 = 0.010$, 90% CI [0, 0.069], $\eta_G^2 = 0.007$, $BF_{01} = 3.01$. The main effect for Congruence was significant, $F(1, 88) = 64.49$, $p < 0.001$, $\eta_p^2 = 0.423$, 90% CI [0.292, 0.523], $\eta_G^2 = 0.189$, $BF_{10} = 2.07 \times 10^9$, showing more accurate responses to congruent stimuli (i.e., positive stimuli in green). Importantly, the interaction between Congruence and Condition was significant, $F(1, 88) = 6.44$, $p = 0.013$, $\eta_p^2 = 0.068$, 90% CI [0.008, 0.164], $\eta_G^2 = 0.023$, $BF_{10} = 3.50$. Mean ERs corroborated the larger congruence effect when polarity was present, with a difference of 6.8% in the green mixed-colour condition versus 3.3% in the green monochromatic condition. Again, the difference in the monochromatic block could be entirely due to the + pole advantage or reflect a residual context-independent green-positivity link.

Overall, the results show that for red and green colour alike, congruence is significantly larger and, in the current study at least, even only definitely present in colour-polarity-affected (mixed-colour) blocks as compared to colour-polarity-unaffected (monochromatic) blocks. Especially when looking at the response pattern for red coloured stimuli, polarity-induced congruence was evidently strong enough to override + pole response facilitation, showing faster and more accurate responses to (red-) negative than (red-)positive stimuli, and an inverted pattern, that is, the typical valence asymmetry with + pole response facilitation, when polarity is not present.

3.4. Congruence sequence effects and valence modulation

On the basis of the polarity-correspondence explanation, we expected that, where a colour-valence congruence effect is found, it is asymmetrical: stronger for the + pole than for the − pole stimuli (Lakens, 2012). As explained above, in the current study, this prediction concerned the mixed-colour blocks in which we found that colour-valence congruence effects were indeed asymmetrical. However, in the Introduction, we explained that it is important to consider more conflict control and stronger diminution of congruence effects following incongruent trials for negative targets as an alternative explanation. Therefore, we analysed performance in a preregistered but exploratory analysis from the mixed-colour condition for congruence sequence effects and its valence modulation.

For the factorial analysis, Congruence was coded as described in the previous section (congruent trials: red-negative and green-positive; incongruent trials: red-positive and green-negative).[8] We took into account not only the congruence of the 'current' trial *n* (*N* Congruence: congruent vs. incongruent), but also, if this trial *n* was preceded by an incongruent trial or not (*N − 1* Congruence: *n − 1* congruent vs. *n − 1* incongruent). A sequence effect would be reflected in an ANOVA as a significant *N* Congruence × *N − 1* Congruence interaction, its modulation would be reflected in a three-way interaction with Valence, with stronger congruence sequence effects (i.e., stronger *N* Congruence × *N − 1* Congruence interaction) for negative than for positive word targets. Again, only the data from the condition were analysed, where the mixed block was presented as the first block (Conditions 3 and 4, see Table 1). In all congruence sequence effect analyses, first trials were excluded, since they did not have a value for *N − 1* Congruence. Incorrect trials were excluded in the RT analysis, as well as those trials that followed an incorrect answer, since in such cases it is unclear if the preceding (in)congruence was perceived as such. In Table 5, we provide the means of median correct RTs and mean ERs.

3.4.1. Reaction times (RTs)

A two (Valence) × two (*N* Congruence) × two (*N − 1* Congruence) repeated-measures ANOVA was performed on the RT data of the mixed condition data. In addition to the main effects for Valence, $F(1, 44) = 23.13$, $p < 0.001$, $\eta_p^2 = 0.345$, 90% CI [0.158, 0.489], $\eta_G^2 = 0.015$, $BF_{10} = 1261.36$, and *N* Congruence, $F(1, 44) = 33.45$, $p < 0.001$, $\eta_p^2 = 0.432$, 90% CI [0.241, 0.562], $\eta_G^2 = 0.039$, $BF_{10} = 9.80 \times 10^8$, which have also been shown previously (see mixed-colour condition analysis), there was a marginally significant main effect for *N − 1* Congruence, $F(1, 44) = 3.39$, $p = 0.072$, $\eta_p^2 = 0.072$, 90% CI [0, 0.211], $\eta_G^2 = 0.004$, $BF_{10} = 1.21$, with faster reactions following *n − 1* congruent than following *n − 1* incongruent trials (see Table 5 for means). The interaction between *N* Congruence and Valence was significant, $F(1, 44) = 4.70$, $p = 0.036$, $\eta_p^2 = 0.096$, 90% CI [0.003, 0.243], $\eta_G^2 = 0.002$, $BF_{01} = 2.24$, illustrating once more the congruence-effect asymmetry – stronger congruence effects (incongruent minus congruent RTs) for positive (45 ms) than for negative words (29 ms). The interaction between *N* Congruence and *N − 1* Congruence was also significant, $F(1, 44) = 8.07$, $p = 0.007$, $\eta_p^2 = 0.155$, 90% CI [0.026, 0.310], $\eta_G^2 = 0.006$, $BF_{10} = 7.28$, demonstrating a clear congruence sequence effect – a stronger congruence effect (incongruent minus congruent trials) after *n − 1* congruent trials (52 ms) than after *n − 1* incongruent trials (22 ms). The Valence × *N − 1* Congruence interaction was not significant, $F(1, 44) = 0.18$, $p = 0.670$, $\eta_p^2 = 0.004$, 90% CI [0, 0.080], $\eta_G^2 < 0.001$, $BF_{01} = 5.95$. Critically, the three-way interaction between *N − 1* Congruence, *N* Congruence and Valence was not significant, $F(1, 44) = 0.11$, $p = 0.746$, $\eta_p^2 = 0.002$, 90% CI [0, 0.068], $\eta_G^2 < 0.001$, $BF_{01} = 4.24$. This means that the asymmetry of the congruence effect in the RTs cannot be explained by more conflict regulation with negative targets. RTs for the *n* congruent and *n* incongruent trials per Valence and *N − 1* Congruence condition are illustrated in Fig. 3.

3.4.2. Error rates (ERs)

An ANOVA with the same variables was performed for the arcsine transformed error rates. The main effect for *N* Congruence was significant, $F(1, 44) = 72.79$, $p < 0.001$, $\eta_p^2 = 0.623$, 90% CI [0.461, 0.714], $\eta_G^2 = 0.177$, $BF_{10} = 1.47 \times 10^{17}$. Main effects for Valence and *N − 1* Congruence were not significant, $F(1, 44) = 0.81$, $p = 0.373$, $\eta_p^2 = 0.018$, 90% CI [0, 0.123], $\eta_G^2 = 0.002$, $BF_{01} = 4.60$, and $F(1, 44) = 0.83$, $p = 0.368$, $\eta_p^2 = 0.018$, 90% CI [0, 0.124], $\eta_G^2 = 0.001$, $BF_{01} = 6.50$, respectively. The interaction between *N* Congruence and Valence was not significant in the ERs, $F(1, 44) = 0.00$, $p = 0.952$, $\eta_p^2 < 0.001$, 90% CI [0, < 0.001], $\eta_G^2 < 0.001$, $BF_{01} = 6.85$, and neither was the interaction between *N − 1* Congruence and Valence, $F(1, 44) = 0.15$, $p = 0.697$, $\eta_p^2 = 0.003$, 90% CI [0, 0.076], $\eta_G^2 < 0.001$, $BF_{01} = 5.97$. However, the *N* Congruence × *N − 1* Congruence interaction, showing the congruence sequence effect, was significant, $F(1, 44) = 23.62$, $p < 0.001$, $\eta_p^2 = 0.349$, 90% CI [0.162, 0.493], $\eta_G^2 = 0.042$, $BF_{10} = 2500.10$. Paired comparisons on the

[8] The preregistered analysis for congruence sequence effects in the mixed-colour condition comprised the factors *N* Congruence, *N − 1* Congruence, Valence and, additionally, Colour. This was a logical fallacy, since Congruence itself already describes a specific directional relationship between Colour and Valence. Thus, in the scope of this analysis, Colour was not a necessary variable and therefore not included.

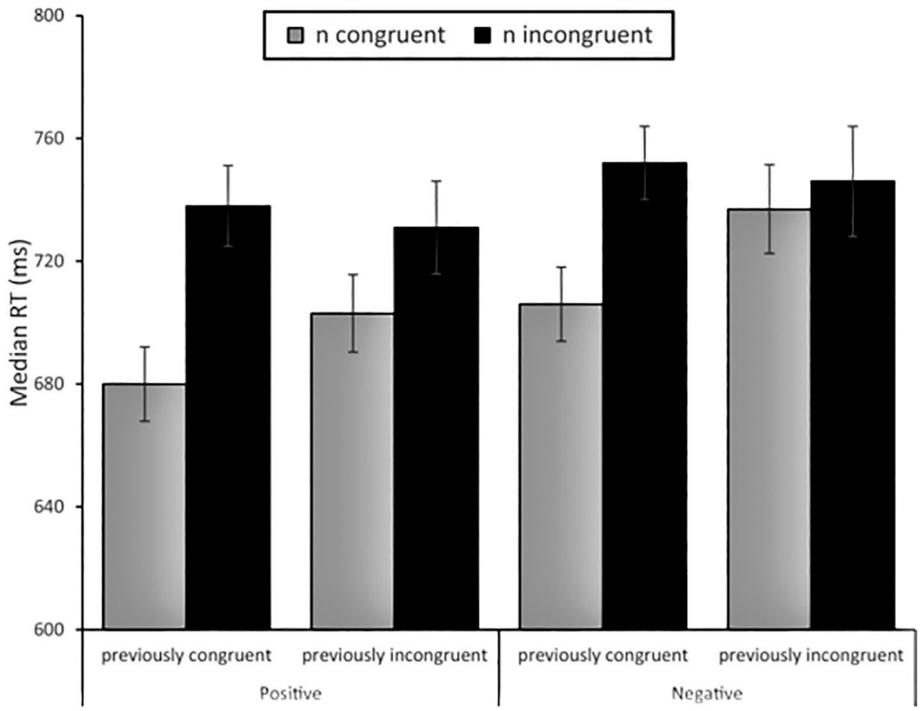


Fig. 3. Means of the median correct reaction times (RTs) for positive and negative words from the first mixed-colour block in current n congruent/incongruent and previously ($n-1$) congruent/incongruent conditions. Error bars show the *SE* of the means.

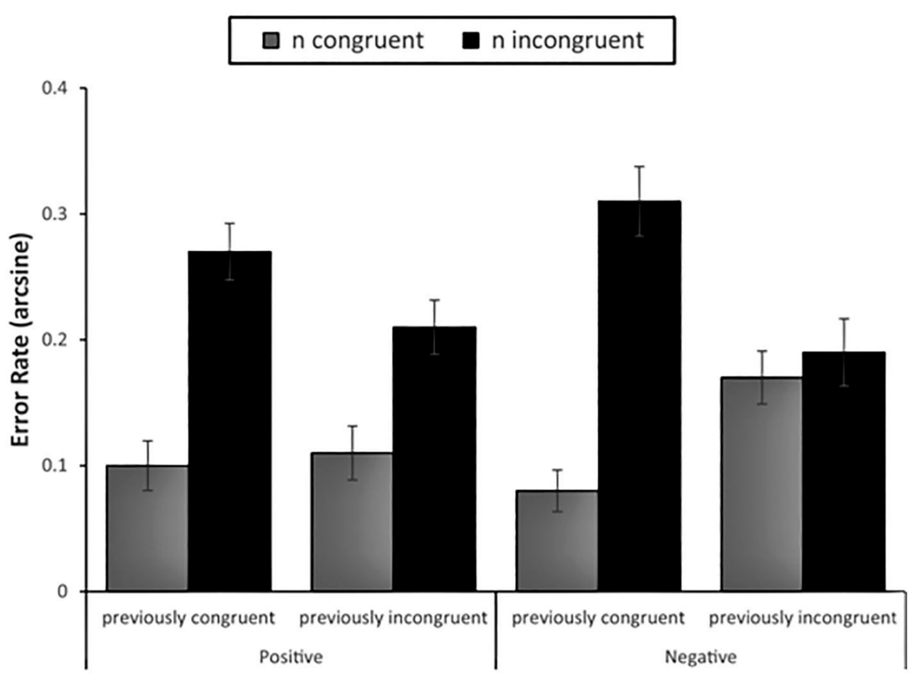


Fig. 4. Arcsine transformed error rates for positive and negative words from the first mixed-colour block in current n congruent/incongruent and previously ($n-1$) congruent/incongruent conditions. The three-way interaction did not lead to an on average just smaller congruence effect in trials following a previously incongruent negative-target trial. Error bars show the *SE* of the means.

arcsine transformed ERs confirmed a larger congruence effect following preceding congruent (8.3%) than preceding incongruent trials (2.7%). The *N* Congruence × *N* − *1* Congruence interaction was also modulated by Valence, $F(1, 44) = 5.92$, $p = 0.019$, $\eta_p^2 = 0.119$, 90% CI [0.011, 0.270], $\eta_G^2 = 0.012$, $BF_{10} = 2.26$. This significant three-way interaction showed that the congruence sequence effect in ERs was stronger for negative words than for positive words. However, this three-way interaction was *not* responsible for a congruence effect asymmetry in the ERs, as on average the congruence effects for – pole and + pole stimuli were about the same (no *N* Congruence × Valence interaction). The three-way interaction in the ERs is illustrated in Fig. 4.

Taken together, the findings show clear congruence sequence effects in RTs and ERs. While valence influenced the size of the congruence effect, but not the congruence sequence effect in the RTs, valence did influence the size of congruence sequence effects in the ERs, but since the congruence effects observed in the error rates were quite symmetrical between valence categories (as illustrated in Fig. 2, right panel), valence-dependent congruence-sequence effects cannot serve as an explanation for congruence-effect asymmetry.

4. Discussion

This study investigated the role of polarity, specifically the presence of one versus two colours – red and green – for colour-valence congruence effects in a binary valence-classification task. Whereas previous studies manipulated the variable colour *either* within-participant or between-participants, making it hard to paint a unified picture of the nature of influences on colour-valence associations, we methodically manipulated if colours were presented "in opposition" to one another or in isolation: Valence categorisation was performed on positive versus negative words which were either presented in red and green colours intermixed (mixed-colour) or in only one colour, red or green (monochromatic).

Results showed clear differences in congruence effects between polarity-affected and polarity-unaffected trials. To this end, we coded green-positive as well as red-negative conditions as congruent and green-negative and red-positive conditions as incongruent and found no clear-cut evidence of a congruence effect in monochromatic blocks where different colour polarities were *not* present. In these blocks, the remaining advantages for green-positive relative to green-negative and for red-positive relative to red-negative stimuli could have reflected a + pole advantage rather than a colour-valence congruence effect. In contrast, when polarity was present in the mixed-colour blocks, we found congruence effects similar to those reported in other experiments with red-green colour opposition (Kuhbandner & Pekrun, 2013; Moller et al., 2009; Schietecat et al., 2018b): Responses to congruent trials (red-negative; green-positive) were faster and more accurate than to incongruent trials (red-positive; green-negative). That these differences were indeed due to colour-valence association and not only due to a + pole advantage was reflected in a significant increase of the advantage in green-positive relative to green-negative conditions from monochromatic to mixed-colour blocks and in an advantage of red-negative relative to red-positive conditions in mixed-colour blocks that was missing in the monochromatic blocks. These findings in support of the polarity-correspondence principle are also in line with findings on congruence or compatibility effects in general, where it is often found that the presence of opposite poles of dimensions is necessary for the compatibility effects (cf. Ansorge & Wühr, 2004; Wühr et al., 2008).

We also corroborated that polarity differences in the valence dimension affect congruence effects (Lakens, 2012; Proctor & Cho, 2006). Our data showed larger congruence effects for positive than for negative words in the RTs. Such polarity-dependent differences have not yet been reported for (chromatic) colour-valence associations. These congruence-effect asymmetries were predicted by + pole response facilitation, a faster processing of + pole than of – pole stimuli (see Lakens, 2012), which we also found in most of our present analyses for the valence targets–with the exception of a polarity-induced red-negative congruence effect, that was pronounced enough to override the advantage for positive words in mixed-colour blocks.

In an additional analysis, we could also rule out that the congruence-effect asymmetry was brought about by a possible conflict control mechanism involving a selectively stronger diminution of congruence effects in negative-target trials, leading to an overall decrease of the congruence effect for negative targets.

Our study shows that red is not linked to negativity in a (monochromatic) speeded binary valence categorisation task, where alternative polarities in the colour dimension are absent. This is in line with an increasing number of studies showing the context-dependence of colour-valence congruence effects and demonstrated that polarity correspondence is one factor responsible for how strongly colour-valence associations impact on stimulus processing and overt behaviour. To be precise, we found no evidence that the negativity of red is automatically activated in contexts in which an alternative colour green did not provide an opposite polarity in the colour dimension. From our analysis of the red-monochromatic blocks, it seems that red is not dominantly associated with either positive or negative valence, since evidence for a difference in classification speed of red-positive versus red-negative words was relatively weak. In fact, this finding rather suggests a more neutral if not even slightly positive valence of the colour red without a context of an antagonistic colour pole. Only when green was present alongside red and a bipolar mapping was encouraged, we observed for the colour red a polarity correspondence contribution in the way that responses to negative words were faster than to positive words. All in all, the fact that for red colour, we were able to observe (1) generally relatively small differences between positive and negative valence responses, and (2) an inversion from a (marginal) processing advantage of red-positive (over red-negative) words when polarity is *absent* to a (marginal) processing advantage of red-negative (over red-positive) words when polarity is present, lend support to the hypothesis that red is ambiguous in nature when it comes to its affective associations. In contrast, for the colour green data might have been more in line with a consistent association between green and positive affect. However, since (1) the difference between green-positive and green-negative responses increased drastically when the counter-colour red was present in the mixed-condition and (2) response patterns did not differ significantly between the monochromatic-red versus the monochromatic-green condition, we cannot rule out that the performance advantage in the monochromatic blocks reflected a general, colour-independent + pole facilitation for the positive words.

What the present study demonstrates is that polarity correspondence is one factor that mediates the connection between the conceptual associations and the overt responses. What the polarity correspondence principle by itself is not able to explain, is, especially in the case of colour, which is not a binary dimension, why polar assignments are made in this way. Apart from the discussion about the potentially ambiguous role of red, the fact that at least a majority of all subjects must have coded green as positive demonstrates that some conceptual knowledge underlying polarity correspondence effects cannot be explained by the polarity correspondence principle itself. However, it is the task of future studies to state more clearly which other factors besides polarity correspondence might account for how much weight a conceptual association gains for a behavioural decision (see the *dimension-specificity hypothesis* by Schietecat, Lakens, IJsselsteijn, & De Kort, 2018a; Schietecat et al., 2018b, for a recent addition on how context predicts cross-modal associations).

A possible limitation of our study originates from the fact that our colour stimuli were controlled for brightness (luminance), but varied on their average chroma values (C*), and in effect, in saturation (red: L* = 51.17, C* = 102, h* = 40.41, sat = 1.99; green: L* = 51.81, C* = 74.03, h* = 134.39, sat = 1.43, see Table 3). While the influence of brightness on the emotional valence dimension is widely corroborated by various studies and ruled out as a factor in the present study (see Lakens et al., 2012; Lakens, Fockenberg, Lemmens, Ham, & Midden, 2013; Meier, Robinson, & Clore, 2004; Specker et al., 2018; Specker & Leder, 2018), variation on other dimensions of colour is an undesirable potential confound. As the study by Wilms and Oberfeld (2018) indicates, highly saturated red might be perceived as neutral, while medium saturated green might be perceived as more positive. Thus, in the case of the mixed blocks, where red was opposed to green, we cannot rule out that the responses to the emotional word stimuli were affected by different saturations for the red versus the green stimuli. However, some studies suggest that saturation has a stronger behavioural and/or physiological effect on the arousal or activity dimension of emotions, less so on the valence dimension (Schietecat et al., 2018a; Valdez & Mehrabian, 1994; Zieliński, 2016). Still, in the case of the mixed blocks, even if the saturation difference between our red and green is with approximately 0.56 relatively small, we cannot rule out that the response times to the emotional word stimuli were affected by different saturations for the red versus the green stimuli.

Another limitation is that it is not easy to study the generality of the polarity correspondence principle in the realm of colour-emotion associations. The reason for this is that colour can be a difficult research subject if we want to exert proper control over the responsible factors by which colours differ from one another. For instance, while it would be interesting to test green against less ambiguously associated colours than red, the questions as to what colour combination could be tested is not easy to answer: Firstly, colours may have all sorts of associations, but positive feelings seem to be predominant, which, on a related note, is reflected in valence judgments for colours. As mentioned, in Wilms and Oberfeld (2018), some colours receive higher valence ratings than others, but no colour is judged as negative per se. Secondly, while achromatic colours are often described as less positive than chromatic ones (Adams & Osgood, 1973; Kaya & Epps, 2004), equating the colourimetrics is impossible in this case.

5. Conclusion

Our study is the first in colour-valence-association research to directly oppose polarity-affected to polarity-unaffected trials. Our results underscore once more the green-positive and red-negative associations, but demonstrate one important influence on whether or not the associations modulate behaviour: the presence of polarities on both valence and colour dimensions. The structural overlap in the mapping of colour and valence concepts on + and − poles is, thus, one factor for if or if not these associations really play out. Our study is also the first to rule out an alternative origin of the asymmetry of the polarity-correspondence effect in terms of more conflict regulation with – pole than with + pole stimuli.

Open practices statement

The study was preregistered via https://osf.io/c4zry (Foster & Deardorff, 2017). The source code for the experiment is available via the same link, as well as all behavioural data collected.

CRediT authorship contribution statement

Claudia Kawai: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing - original draft.**Gáspár Lukács:**Methodology, Investigation, Software, Writing - review & editing.**Ulrich Ansorge:**Supervision, Writing - review & editing.

Acknowledgments

Claudia Kawai and Gáspár Lukács are recipients of a DOC Fellowship of the Austrian Academy of Sciences at the Department of Cognition, Emotion, and Methods in Psychology at the University of Vienna. The research was also supported by Dr. Franziska Pinsker.

Appendix A

Positive words		
SPANNEND	TRIUMPH	SCHLAU
ERINNERN	FEIERN	VERTRAUT
WÜNSCHEN	HELD	BELOHNEN
AKTIV	PARTY	TOLERANT
TAPFER	BEGEISTERN	FREIZEIT
REIZVOLL	RETTEN	ANMUTIG
ENTDECKEN	MUTIG	LEUCHTEN
MÖGEN	FREUEN	SCHÜTZEN
GENIE	FLIRT	BUNT
TALENT	EKSTASE	BEQUEM
STARK	SPENDE	SCHENKEN
GRANDIOS	VERBESSERN	BEIFALL
LEBENDIG	ZUSTIMMEN	UMARMEN
INTENSIV	ANGEBOT	ERFINDER
ANSPORN	MAGISCH	OPTIMIST
STAUNEN	APPETIT	HEITER
REICHTUM	SANFT	LUXUS
GRINSEN	STÄRKEN	HEIMKEHR
MUNTER	DUFTEN	LOB
GEBURTSTAG	GRATIS	BEGRÜSSEN

Negative words		
NAIV	AAS	UNBEQUEM
BEREUEN	SPUCKEN	VERBIETEN
MIETE	SCHMUTZ	PEINLICH
UNFAIR	BALLAST	ZWINGEN
SCHWACH	MAUL	HEUCHELN
FORDERUNG	TRETEN	TRAUERN
PRIMITIV	STOTTERN	DROHEN
ALLEIN	WARTEREI	GEFÄNGNIS
VERZICHT	ABSTIEG	PANISCH
MONOTON	REGLOS	LEIDEN
MÜSSEN	STRAFE	TROSTLOS
FEHLEN	GRAUSAM	VERLETZEN
KÜNDIGEN	SCHEUSAL	BOSHAFT
ZÖGERN	IDIOT	BLAMAGE
BITTER	MANGEL	VERBRECHEN
NARBIG	KORRUPT	HEULEN
SCHLAPP	ARROGANZ	KONFLIKT
SCHLIESSEN	GEMEIN	GIERIG
HÄNGEN	STÖREN	TROTZIG
LOCH	VERSAGEN	KÄFIG

References

Adams, F. M., & Osgood, C. E. (1973). A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology, 4*(2), 135–156. https://doi.org/10.1177/002202217300400201.

Akers, A., Barton, J., Cossey, R., Gainsford, P., Griffin, M., & Micklewright, D. (2012). Visual color perception in green exercise: Positive effects on mood and perceived exertion. *Environmental Science & Technology, 46*(16), 8661–8666. https://doi.org/10.1021/es301685g.

Ansorge, U., Kunde, W., & Kiefer, M. (2014). Unconscious vision and executive control: How unconscious processing and conscious action control interact. *Consciousness and Cognition, 27*, 268–287. https://doi.org/10.1016/j.concog.2014.05.009.

Ansorge, U., & Wühr, P. (2004). A response-discrimination account of the Simon effect. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 365–377. https://doi.org/10.1037/0096-1523.30.2.365.

Bertelson, P. (1965). Serial choice reaction-time as a function of response versus signal-and-response repetition. *Nature, 206*(4980), 217. https://doi.org/10.1038/206217a0.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624.

Briki, W., & Hue, O. (2016). How red, blue, and green are affectively judged: Affective judgments of colors. *Applied Cognitive Psychology, 30*(2), 301–304. https://doi.org/10.1002/acp.3206.

Briki, W., Rinaldi, K., Riera, F., Trong, T. T., & Hue, O. (2015). Perceiving red decreases motor performance over time: A pilot study. *European Review of Applied Psychology, 65*(6), 301–305. https://doi.org/10.1016/j.erap.2015.09.001.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123.

Buechner, V. L., & Maier, M. A. (2016). Not always a matter of context: Direct effects of red on arousal but context-dependent moderations on valence. *PeerJ, 4*, Article e2515. https://doi.org/10.7717/peerj.2515.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review, 76*(4), 387–404. https://doi.org/10.1037/h0027578.

Clarke, T., & Costall, A. (2008). The emotional connotations of color: A qualitative investigation. *Color Research & Application, 33*(5), 406–410. https://doi.org/10.1002/col.20435.

Crozier, W. R. (1999). The meanings of colour: Preferences among hues. *Pigment & Resin Technology, 28*(1), 6–14. https://doi.org/10.1108/03699429910252315.

De Bruijn, E. R., Hulstijn, W., Verkes, R. J., Ruigt, G. S., & Sabbe, B. G. (2004). Drug-induced stimulation and suppression of action monitoring in healthy volunteers. *Psychopharmacology, 177*(1–2), 151–160. https://doi.org/10.1007/s00213-004-1915-6.

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology, 30*(1), 92. https://doi.org/10.5334/irsp.82.

Dreisbach, G., & Fischer, R. (2012). Conflicts as aversive signals. *Brain and Cognition, 78*(2), 94–98. https://doi.org/10.1016/j.bandc.2011.12.003.

Elliot, A. J., & Maier, M. A. (2012). Color-in-context theory. *Advances in experimental social psychology*. *Vol. 45*. *Advances in experimental social psychology* (pp. 61–125). . https://doi.org/10.1016/B978-0-12-394286-9.00002-0.

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*(1), 154–168. https://doi.org/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*(5), 1150–1164. https://doi.org/10.1037/0022-3514.95.5.1150.

Elliot, A. J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., & Liu, H. (2010). Red, rank, and romance in women viewing men. *Journal of Experimental Psychology: General, 139*(3), 399–417. https://doi.org/10.1037/a0019689.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.

Fetterman, A. K., Robinson, M. D., & Meier, B. P. (2012). Anger as "seeing red": Evidence for a perceptual association. *Cognition & Emotion, 26*(8), 1445–1458. https://doi.org/10.1080/02699931.2012.673477.

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association, 105*(2), 203. https://doi.org/10.5195/jmla.2017.88.

Francis, G. (2013). Publication bias in "Red, rank, and romance in women viewing men," by Elliot et al. (2010). *Journal of Experimental Psychology: General, 142*(1), 292–296. https://doi.org/10.1037/a0027923.

Fritz, J., & Dreisbach, G. (2013). Conflicts as aversive signals: Conflict priming increases negative judgments for neutral stimuli. *Cognitive, Affective, & Behavioral Neuroscience, 13*(2), 311–317. https://doi.org/10.3758/s13415-012-0147-1.

Goller, F., Kroiss, A., & Ansorge, U. (2019). Conflict-elicited negative evaluations of neutral stimuli: Testing overt responses and stimulus-frequency differences as critical side conditions. *Frontiers in Psychology, 10*, 2204. https://doi.org/10.3389/fpsyg.2019.02204.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121*(4), 480.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464.

Hobson, N. M., Saunders, B., Al-Khindi, T., & Inzlicht, M. (2014). Emotion down-regulation diminishes cognitive control: A neurophysiological investigation. *Emotion, 14*(6), 1014–1026. https://doi.org/10.1037/a0038028.

Huber-Huber, C., & Ansorge, U. (2017). The role of RT carry-over for congruence sequence effects in masked priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(5), 757–780. https://doi.org/10.1037/xlm0000331.

Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive control. *Trends in Cognitive Sciences, 19*(3), 126–132. https://doi.org/10.1016/j.tics.2015.01.004.

Jonauskaite, D., Wicker, J., Mohr, C., Dael, N., Havelka, J., Papadatou-Pastou, M., & Zhang Oberfeld, D. (2019). A machine learning approach to quantify the specificity of colour–emotion associations and their cultural differences. *Royal Society Open Science, 6*(9), 190741. https://doi.org/10.1098/rsos.190741.

Kaya, N., & Epps, H. H. (2004). Relationship between color and emotion: A study of college students. *College Student Journal, 38*(3), 396–405.

Kelley, K. (2019). MBESS: The MBESS R Package. R package version 4.5.1. Retrieved from https://CRAN.R-project.org/package=MBESS.

Kuhbandner, C., & Pekrun, R. (2013). Joint effects of emotion and color on memory. *Emotion, 13*(3), 375–379. https://doi.org/10.1037/a0031821.

Labrecque, L. I., & Milne, G. R. (2012). Exciting red and competent blue: The importance of color in marketing. *Journal of the Academy of Marketing Science, 40*(5), 711–727. https://doi.org/10.1007/s11747-010-0245-y.

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 726–736. https://doi.org/10.1037/a0024955.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863.

Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J., & Midden, C. J. H. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion, 27*(7), 1225–1246. https://doi.org/10.1080/02699931.2013.781501.

Lakens, D., Semin, G. R., & Foroni, F. (2012). But for the bad, there would not be good: Grounding valence in brightness through shared relational structures. *Journal of Experimental Psychology: General, 141*(3), 584–594. https://doi.org/10.1037/a0026468.

Lawrence, M. A. (2016). Ez: Easy analysis and visualization of factorial experiments. R package version 4.4-0. Retrieved from https://CRAN.R-project.org/package=ez.

Lebois, L. A., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science, 39*(8), 1764–1801. https://doi.org/10.1111/cogs.12174.

Lehmann, G. K., & Calin-Jageman, R. J. (2017). Is red really romantic?: Two pre-registered replications of the red-romance hypothesis. *Social Psychology, 48*(3), 174–183. https://doi.org/10.1027/1864-9335/a000296.

Lehmann, G. K., Elliot, A. J., & Calin-Jageman, R. J. (2018). Meta-analysis of the effect of red on perceived attractiveness. *Evolutionary Psychology, 16*(4), Article 147470491880241. https://doi.org/10.1177/1474704918802412.

Lichtenfeld, S., Elliot, A. J., Maier, M. A., & Pekrun, R. (2012). Fertile green: Green facilitates creative performance. *Personality and Social Psychology Bulletin, 38*(6), 784–797. https://doi.org/10.1177/0146167212436611.

Lichtenfeld, S., Maier, M. A., Elliot, A. J., & Pekrun, R. (2009). The semantic red effect: Processing the word red undermines intellectual performance. *Journal of Experimental Social Psychology, 45*(6), 1273–1276. https://doi.org/10.1016/j.jesp.2009.06.003.

Lohmann, C., & Jorschick, A. (2015). *Studie in Rot-Grün – Farben aktivieren emotionale Wortbedeutungen [Study in red-green – Colors activate emotional word meanings]. Abstract presented at the TEAP 2015 Conference, Hildesheim, Germany.* March.

Lukács, G. (2019). neatStats: An R Package for neat and painless statistical reporting. R package version 0.3.1. Retrieved from https://github.com/gasparl/neatstats.

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software, 4*(40), 1541. https://doi.org/10.21105/joss.01541.

Mathôt, S. (2017). Bayes like a Baws: Interpreting Bayesian repeated measures in JASP. May 15 https://www.cogsci.nl/blog/interpreting-bayesian-repeated-measures-in-jasp.

Mehta, R., & Zhu, R. (2009, February). Blue or red? Exploring the effect of color on cognitive task performances. *Science, 323*(5918), 1226–1229. https://doi.org/10.1126/science.1169144.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science, 15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x.

Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white: Automatic inferences about stimulus valence based on brightness. *Psychological Science, 15*(2), 82–87. https://doi.org/10.1111/j.0963-7214.2004.01502002.x.

Moller, A. C., Elliot, A. J., & Maier, M. A. (2009). Basic hue-meaning associations. *Emotion, 9*(6), 898–902. https://doi.org/10.1037/a0017811.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297–326. https://doi.org/10.1037/0033-2909.132.2.297.

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-4.2. Retrieved from https://CRAN.R-project.org/package=BayesFactor.

Pazda, A. D., Elliot, A. J., & Greitemeyer, T. (2012). Sexy red: Perceived sexual receptivity

mediates the red-attraction relation in men viewing woman. *Journal of Experimental Social Psychology, 48*(3), 787–790. https://doi.org/10.1016/j.jesp.2011.12.009.
Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017.
Peperkoorn, L. S., Roberts, S. C., & Pollet, T. V. (2016). Revisiting the red effect on attractiveness and sexual receptivity: No effect of the color red on human mate preferences. *Evolutionary Psychology, 14*(4), Article 147470491667384. https://doi.org/10.1177/1474704916673841.
Pravossoudovitch, K., Cury, F., Young, S. G., & Elliot, A. J. (2014). Is red the colour of danger? Testing an implicit red–danger association. *Ergonomics, 57*(4), 503–510. https://doi.org/10.1080/00140139.2014.889220.
Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132*(3), 416–442. https://doi.org/10.1037/0033-2909.132.3.416.
Puccioni, O., & Vallesi, A. (2012). Sequential congruency effects: Disentangling priming and conflict adaptation. *Psychological Research, 76*(5), 591–600. https://doi.org/10.1007/s00426-011-0360-5.
R Core Team (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.
Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207.
Saito, M. (1996). Comparative studies on color preference in Japan and other Asian regions, with special emphasis on the preference for white. *Color Research and Application, 21*(1), 35–49. https://doi.org/10.1002/(SICI)1520-6378(199602)21:1 < 35::AID-COL4 > 3.0.CO;2-6.
Schietecat, A. C., Lakens, D., IJsselsteijn, W. A., & De Kort, Y. A. W. (2018a). Predicting context-dependent cross-modal associations with dimension-specific polarity attributions. Part 1 – Brightness and aggression. *Collabra: Psychology, 4*(1), 14. https://doi.org/10.1525/collabra.110.
Schietecat, A. C., Lakens, D., IJsselsteijn, W. A., & De Kort, Y. A. W. (2018b). Predicting context-dependent cross-modal associations with dimension-specific polarity attributions. Part 2: Red and valence. *Collabra: Psychology, 4*(1), 21. https://doi.org/10.1525/collabra.126.
Schloss, K. B., Strauss, E. D., & Palmer, S. E. (2013). Object color preferences. *Color Research and Application, 38*(6), 393–411. https://doi.org/10.1002/col.21756.
Schouppe, N., Braem, S., De Houwer, J., Silvetti, M., Verguts, T., Ridderinkhof, K. R., & Notebaert, W. (2015). No pain, no gain: The affective valence of congruency conditions changes following a successful response. *Cognitive, Affective, & Behavioral Neuroscience, 15*(1), 251–261. https://doi.org/10.3758/s13415-014-0318-3.
Smajic, A., Merritt, S., Banister, C., & Blinebry, A. (2014). The red effect, anxiety, and exam performance: A multistudy examination. *Teaching of Psychology, 41*(1), 37–43. https://doi.org/10.1177/0098628313514176.
Sokal, R. R., & Rohlf, R. F. (1981). *Biometry: The principles and practice of statistics in biological research* (2nd ed.). San Francisco, CA: Freeman.
Soriano, C., & Valenzuela, J. (2009). Emotion and colour across languages: implicit associations in Spanish colour terms. *Social Science Information, 48*(3), 421–445. https://doi.org/10.1177/0539018409106199.
Specker, E., & Leder, H. (2018). Looking on the bright side: Replicating the association between brightness and positivity. *Collabra: Psychology, 4*(1), 34. https://doi.org/10.1525/collabra.168.
Specker, E., Leder, H., Rosenberg, R., Hegelmaier, L. M., Brinkmann, H., Mikuni, J., & Kawabata, H. (2018). The universal and automatic association between brightness and positivity. *Acta Psychologica, 186*, 47–53. https://doi.org/10.1016/j.actpsy.2018.04.007.
Steele, K. M. (2014). Failure to replicate the Mehta and Zhu (2009) color-priming effect on anagram solution times. *Psychonomic Bulletin & Review, 21*(3), 771–776. https://doi.org/10.3758/s13423-013-0548-3.
Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*(2), 164–182. https://doi.org/10.1037/1082-989X.9.2.164.
Terwogt, M. M., & Hoeksma, J. B. (1995). Colors and emotions: Preferences and combinations. *The Journal of General Psychology, 122*(1), 5–17. https://doi.org/10.1080/00221309.1995.9921217.
Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. Journal of experimental psychology. *General, 123*(4), 394. https://doi.org/10.1037/0096-3445.123.4.394.
van Steenbergen, H., Band, G. P., & Hommel, B. (2009). Reward counteracts conflict adaptation: Evidence for a role of affect in executive control. *Psychological Science, 20*(12), 1473–1477. https://doi.org/10.1111/j.1467-9280.2009.02470.x.
van Steenbergen, H., Band, G. P., & Hommel, B. (2010). In the mood for adaptation: How affect regulates conflict-driven control. *Psychological Science, 21*(11), 1629–1634. https://doi.org/10.1177/0956797610385951.
Võ, M., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M., & Jacobs, A. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods, 41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534.
Wilms, L., & Oberfeld, D. (2018). Color and emotion: Effects of hue, saturation, and brightness. *Psychological Research, 82*(5), 896–914. https://doi.org/10.1007/s00426-017-0880-8.
Wilson, G. D. (1966). Arousal properties of red versus green. *Perceptual and Motor Skills, 23*(3), 947–949. https://doi.org/10.2466/pms.1966.23.3.947.
Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 483–502. https://doi.org/10.1037/0096-1523.29.2.483.
Wu, Y., Lu, J., van Dijk, E., Li, H., & Schnall, S. (2018). The color red is implicitly associated with social status in the United Kingdom and China. *Frontiers in Psychology, 9*, 1902. https://doi.org/10.3389/fpsyg.2018.01902.
Wühr, P., Biebl, R., & Ansorge, U. (2008). The impact of stimulus and response variability on SR correspondence effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(3), 533–545. https://doi.org/10.1037/0278-7393.34.3.5.
Zieliński, P. (2016). An arousal effect of colors saturation: A study of self-reported ratings and electrodermal responses. *Journal of Psychophysiology, 30*(1), 9–16. https://doi.org/10.1027/0269-8803/a000149.

2.2 **Kawai, C.**, Lukács, G., & Ansorge, U. (2021). A new type of pictorial database: The Bicolor Affective Silhouettes and Shapes (BASS). *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01569-7

Behavior Research Methods
https://doi.org/10.3758/s13428-021-01569-7

A new type of pictorial database: The Bicolor Affective Silhouettes and Shapes (BASS)

Claudia Kawai[1] · Gáspár Lukács[1] · Ulrich Ansorge[1,2,3]

Accepted: 19 February 2021

Abstract

We introduce the Bicolor Affective Silhouettes and Shapes (BASS): a set of 583 normed black-and-white silhouette images that is freely available via https://osf.io/anej6/. Valence and arousal ratings were obtained for each image from US residents as a Western population (n = 777) and Chinese residents as an Asian population (n = 869). Importantly, the ratings demonstrate that, notwithstanding their visual simplicity, the images represent a wide range of affective content (from very negative to very positive, and from very calm to very intense). In addition, speaking to their cultural neutrality, the valence ratings correlated very highly between US and Chinese ratings. Arousal ratings were less consistent between the two samples, with larger discrepancies in the older age groups inviting further investigation. Due to their simplistic and abstract nature, our silhouette images may be useful for intercultural studies, color and shape perception research, and online stimulus presentation in particular. We demonstrate the versatility of the BASS by an example online experiment.

Keywords Affective rating · Silhouette · Emotion · Database · Valence · Arousal · Online research

Images are an excellent medium to convey information, simple or complex. They serve as stimuli in a wide range of research areas, such as emotion, attention, or aesthetics research, among others (e.g., Bradley & Lang, 2007; Huston et al., 2015; Lindsay, 2020). In the present article, we introduce an open access database of normed Bicolor Affective Silhouettes and Shapes (BASS).[1] Each of the BASS images consists of 300 × 300 black and white pixels, providing a computationally economical and visually uniform layout.[2]

New freely available normed images enrich research by providing greater diversity of choices for research, including potential replications of findings with different kinds of images to ensure generalizability. Additionally, participants may get habituated to the images in existing databases, meaning they may process them differently on repeated exposures (Baker et al., 2010; Foa & Kozak, 1986; Ramaswami, 2014). The present BASS database may be used, for instance, for conceptual replications and novel research in such diverse areas as space-valence congruence effects (e.g., Meier & Robinson, 2004), priming of evaluations (e.g., Fazio, 2001),

[1] Some of the BASS images are not *silhouettes* in the strictest sense, but rather *silhouette-like shapes* because they also illustrate hollow parts or outlines (e.g., the edges and dots on dice). Nonetheless, for simplicity, in the present article, we refer to all the BASS images as "silhouettes."

[2] Several databases of shapes or objects do exist, like MultiPic (Duñabeitia et al., 2018), ALOI (Geusebroek et al., 2005), the sets by Nishimoto et al. (2012), by Brady et al. (2013), or by Snodgrass and Vanderwart (Rossion & Pourtois, 2004; Snodgrass & Vanderwart, 1980). However, these were not created (let alone verified or normed) with the intention of eliciting affect: The majority of these images depict everyday objects or animals. With a few exceptions, like dangerous versus cute animals, or weapons, it is unlikely that they would elicit strong affective responses or emotions. In the OASIS, for example, people and scene pictures were rated as more emotional than objects (Kurdi et al., 2017, p. 463), which suggests that these databases are not ideal for emotion research purposes.

✉ Claudia Kawai
claudia.kawai@univie.ac.at

1 Department of Cognition, Emotion, and Methods in Psychology, University of Vienna, Vienna, Austria

2 Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria

3 Research Platform Mediatized Lifeworlds, University of Vienna, Vienna, Austria

Published online: 07 May 2021

affective priming (e.g., Hermans et al., 2001), or emotional facilitation (e.g., Schupp et al., 2003).

More importantly, the BASS also has very specific advantages compared to other available stimulus sets: Silhouette images allow for the study of meaning-related processing without many of the confounds that are present in words and pictures. Whereas words carry inherent differences, such as their length and their phonological transparency (or orthographic depth, see Aro & Wimmer, 2003; Frost et al., 1987; Schmalz et al., 2016) for which one would have to control, these confounding differences are absent in silhouettes. Likewise, whereas pictures differ from one another in terms of color heterogeneity, depth cues, spatial perspective, or feature complexity, silhouettes are more uniform and less complex, and can therefore be more easily equated for these factors, without corrupting their meaning entirely.

The main purpose of the present research article is to demonstrate that the BASS images evoke a very wide range of affective representations on the side of the participants—despite the images' relative simplicity. We demonstrate this primarily by collecting valence and arousal ratings in a large and nationally representative US (Western) sample. We provide further evidence from analogous ratings from an extensive (Eastern) sample from Mainland China and from demonstrating characteristic valence-color congruence effects in an example response time experiment using BASS images (in an Austrian sample).

This methodological approach validates several important characteristics of the BASS images, relevant to but often not investigated in image sets used in psychological research. First, typically, normative ratings of affective picture databases like the Geneva Affective Picture Database (GAPED), the Open Affective Standardized Image Set (OASIS), the International Affective Picture System (IAPS), or the Nencki Affective Picture System (NAPS) (Dan-Glauser & Scherer, 2011; Kurdi et al., 2017; Lang et al., 1997; Marchewka et al., 2014, respectively) are largely based on mono-cultural samples, with ratings predominantly collected in Western populations only. Moreover, within Western populations, samples were often restricted to student participants.[3] Yet, ideally, behavioral image or stimulus sets should be applicable in different cultures in order to allow investigation of universal psychological phenomena and topics. The IAPS, for example, has been generally found to be valid among Western cultures (Deák et al., 2010; Grühn & Scheibe, 2008; Verschuere et al., 2001, to name a few), but less so among Asian cultures (Gong & Wang, 2016; Huang et al., 2015). Therefore, from the start, for the current BASS images, we intended to collect normative rating data from at least two large, representative samples of different cultural origin: one Western (US) and one Eastern (Mainland China) sample. Silhouette images theoretically carry potential for more culturally neutral affective representations than photographic images, due to the latter's culture-specific details (such as depictions of regionally different vegetation, landscapes, food, people, architecture, traffic etc.), which are often subtle and difficult to control for. We tested this conceptual notion by directly comparing the consistency of the ratings from our Western and Asian samples.

Second, the bicolor silhouettes tackle one particular virulent problem posed by online studies: accurate color rendition of heterogeneous colors on different physical displays. By implication of their high realism, color photos of natural scenes carry substantial color heterogeneity and, in fact, their realism relies in part on this color heterogeneity (cf. Hansen & Gegenfurtner, 2009). It has indeed been shown that differences in the display of photographs, such as varying brightness or resolution, can substantially affect the evaluation of emotional content (e.g., Lakens et al., 2013; Mould et al., 2012). However, the huge variety of physical devices used to display images (e.g., smartphones, tablets, and laptops by different manufacturers) plus an increased variance in idiosyncratic monitor settings and lighting conditions on the side of the viewer compromise accurate color rendition, especially in online studies (Anwyl-Irvine et al., 2020)—a problem which increases with the complexity of the image's color palette. In contrast, the black-and-white silhouettes of the BASS database were already created and selected (from existing silhouettes) under the premise of having to convey their affective meanings largely by their contours alone and less by their exact color rendition. In this way, the BASS images reduce the heterogeneity of the necessary color palette for their rendition and the problems associated with this color heterogeneity. We demonstrated this by collecting participants' affective ratings using the BASS images in inverse colors (pilot study).

However, this is not to say that colors could not be manipulated and used to study their effects (e.g., on affective ratings) in studies with BASS images. For photographs, manipulations of color rendition can easily corrupt image meaning altogether (cf. Oliva & Schyns, 2000; Mould et al., 2012). Such manipulation requires complicated technical procedures (e.g., Orzan et al., 2007), and, even more fundamental, one can hardly be sure what manipulation of colors in a photography affects emotional image content in what manner: Setting up a proper balance of hue, saturation, and brightness is difficult enough when calibrating the display of a single color (Wilms & Oberfeld, 2018), let alone for a photograph with thousands of interacting colors. In contrast, the BASS images were created with the intention of keeping an unambiguous meaning by way of their contours, even if presented in varying colors. Thus, BASS images offer an extremely easy way to manipulate and study or control color: They can be converted effortlessly to any other two colors[4]—and, as long as the two colors are discernible from each other, the

[3] IAPS: US undergraduate students, OASIS: more diverse but predominantly white US participants, GAPED: South/West-European psychology students, NAPS: Polish and other mostly European students.

[4] An R script for easy instant conversion of any number of images to any chosen color is available via https://gasparl.github.io/BASS/.

content remains unambiguous. Below, we also demonstrate how such color manipulations of the BASS images could be used in research by investigating known color-valence congruence effects using variously colored BASS images for an example experiment.

Third, large photographic images are generally not optimal for fast decision tasks based on their diverse visual characteristics and complex semantics: Various depicted details and their many subtle visual properties might affect each participant very differently, influencing processing time and potentially confounding results (especially where small effects are expected and processing-time differences in fractions of seconds can affect outcomes). In contrast, the contents of the less ambiguous (few or single) objects in each BASS image is easy to grasp quickly and better suited for a task involving fast decisions, unaffected by visual or semantic noise. In addition, their simplistic and abstract nature makes BASS silhouettes more resistant than photographs to incidental but potentially meaning-corrupting influences such as display size or viewing distance (Anwyl-Irvine et al., 2020).[5] Big file sizes of photographic images used in experiments typically pose technical problems too, especially in connection with online research, which is becoming more and more important in psychological science. For one, common browsers are not optimized for precisely timing image presentation, and physical display times can be strongly affected by loading large size images (e.g., Garaizar & Reips, 2019). For another, some participants might be reluctant (or even unable, where infrastructure is weak) to download many large files. It is not unusual for an experiment to require several hundreds of different stimuli (especially when each stimulus is presented repeatedly in different forms, e.g., varying hue, etc.; e.g., Huang et al., 2015; Kawai et al., 2020; Oliva & Schyns, 2000). Calculating with the average file size of IAPS images (262 KB; cf. averages of 92 KB for OASIS and 789 KB for GAPED), just 100 images take 26 MB—already a notable hurdle for potential participants. Participant dropout due to inadequate internet speed could even introduce confounds via selective attrition (Zhou & Fishbach, 2016). With an extremely economical average file size of 2.2 KB (max. 6.0 KB), the BASS images minimize the issues of display timing and download. Our follow-up experiment on color-valence congruence effects also serves to demonstrate these usability advantages in fast decision tasks online.

Fourth, silhouettes, contours, and shapes themselves constitute a field of scientific interest (e.g., Bar, 2007; Biederman, 1987; Marr, 1982), and the affective qualities associated with geometric contour properties may be of interest to future researchers (e.g., Leder et al., 2011), although they have not been a major focus of research on contours so far. This could be particularly interesting when comparing human object recognition with artificial object recognition (e.g., Rajalingham et al., 2018): While objects as simple as silhouettes may be processed by machines similarly to humans, affective responses are (currently) only evoked in humans.

[5] The contents of most BASS images can be easily recognized even when they are strongly blurred or scaled to icon size.

Creating the BASS

Before carrying out the large-scale ratings for the BASS, we assessed several issues in a pilot study ($n = 180$, students of the University of Vienna). Detailed information on the pilot study can be found in the online supplementary material available on https://osf.io/anej6/. Firstly, researchers suspected, but never tested, possible contamination of affective ratings when participants had to rate valence and arousal in one experiment (Kurdi et al., 2017), since prior studies showed that question order in surveys can influence judgements (Lau et al., 1990; Wilcox & Wlezien, 1993). We put this assumption to the test and split participants into two groups: one rating only valence (or only arousal) and another rating images on both valence and arousal. Results indeed confirmed contamination in affective ratings when two affective ratings were required in the same experiment. Thus, we decided to use only one rating measure (either valence or arousal) per participant in the main study. In the pilot study, we simultaneously tested for the relative-color independence of the silhouette ratings by studying possible effects of color-inversion of silhouettes (black-on-white vs. white-on-black) on ratings. High positive correlations between the (original) black-on-white images and the inverted white-on-black images suggested that there is no apparent preferred color mode for ratings of our silhouettes (clarity ratings: $r[618] = .866$, 95% CI [.845, .885]; valence: $r[618] = .918$, 95% CI [.905, .930]; arousal: $r[618] = .719$, 95% CI [.678, .755]). The ratings from the pilot study, given by Austrian psychology students, also helped us to decide on a reasonable exclusion threshold for intrarater correlation, which we subsequently used as an individual reliability measure for data quality control in the main study. Since intrarater correlation turned out to be lower in the arousal rating than in the valence rating condition, we determined different exclusion thresholds depending on the rating condition (see Data Exclusion).

Clarity ratings provided by the first as well as a second online prestudy ($n = 50$, Chinese participants) helped to identify unclear or ambiguous images in the original set of 620. After excluding those with the lowest clarity ratings, the subset that was used in the large-scale ratings in the US and China comprised a total of 583 silhouettes.

Method

Participants

For the US rating, a total of 806 participants was recruited via Prolific (www.prolific.co). The study was divided into two parts: one sub-study comprised the valence rating (n = 402), the other the arousal rating (n = 404) task. The sample was collected as a "representative sample for the United States of America" (sex, age, ethnicity, according to Simplified US Census; excluding people under the age of 18) and participants received £2.17 upon study completion.

The experiment had to be completed at a desktop computer. Participants could read all relevant information on the welcome page and consented to participate by clicking the consent button at the bottom.

After exclusion (see Data Exclusion), valid data from 777 participants remained (arousal rating: 386; valence rating: 391), out of which 402 participants were female, 375 were male. Ages ranged from 18 to 78 years, with a mean age of 45.19 years (SD = 16.09).[6] Completion took an average of 17 min (SD = 4 min, median = 15 min).

For the Chinese sample, the recruitment (and compensation) of online participants in China was managed by Dynata. In the supplementary materials, we provide more information on Dynata's incentive and compensation system. Chinese participants were asked to provide demographic data after consenting to the study. We only allowed participation for those who indicated (1) their age being between 18 and 95 years, (2) Mainland China as their country of residence, and (3) any gender information. A total of 1341 registered participants submitted their rating data, of which 472 had to be excluded due to poor data quality (for more details see Data Exclusion). We aimed for a representative sample (sex; age, excluding people under the age of 18 years) for the People's Republic of China, but since our study required PC and Internet access, relatively younger age groups are over-represented.

After exclusion, valid data from 869 participants remained (arousal rating: 455; valence rating: 414), out of which 419 participants were female and 450 were male. Ages ranged from 18 to 81 years, with a mean age of 34.5 years (SD = 9.4). Completion took on average 25 min (SD = 116 min, median = 14 min).[7]

An overview of the demographic data from all included participants (US and China) is provided in Table 1 (Appendix A).

Materials

The majority of the silhouettes were acquired using Google Images (images.google.com) under the creative commons license, by searching either directly for silhouettes or for photographic pictures that we edited into black-and-white silhouettes using GNU Image Manipulation Program (The GIMP Development Team, 2019) and R (R Core Team, 2020).

We restricted our search to images labeled "available for reuse with modification." Most silhouettes are from pixabay.com (n = 444), cleanpng.com (n = 33), and svgsilh.com (n = 29). Eighteen additional silhouettes were created for this project by a professional illustrator. We collected a wide range of images, depicting humans, animals, objects, and scenes, which we indicated by a category column in the database for easier stimulus selection. Special focus was put on collecting images of a wide range of valence (positive, negative, and neutral) as well as arousal (low, medium, and high) levels. Every image was scaled and/or cropped to a size of 300 × 300 pixels. All images were checked (or converted) to contain only fully black (RGB: 0, 0, 0) and fully white (RGB: 255, 255, 255) pixels.

Procedure

Online Rating Study – US After giving their informed consent, participants saw the instruction page with an explanation of the terms valence and/or arousal (see Figs. 11 and 12 in Appendix A), depending on the participants' rating condition, and three example images (selected based on the pilot studies as average in both arousal and valence; none of these images were used in the subsequent task for the given participant). The experiment started as soon as the "Start" button on the bottom of the instruction page was pressed. Each black-and-white silhouette stimulus (300 × 300 px) was presented on a grey background (RGB: 128, 128, 128) for 2 s. When this time had elapsed, the image disappeared and was replaced with the rating scale, a line ranging from "very low" on the left to "very high" on the right, with nine equally spaced tick marks (see Fig. 13 in Appendix A). Participants could enter their ratings by clicking on the scale(s) and submit their rating by pressing a button that appeared below, or they could skip the rating. After confirming their choice, the next image appeared on the screen, and so on. A total of 145-146 silhouettes were presented to each participant. Additionally, we incorporated two attention checks. Hyperlinks to the original experimental websites are available via https://osf.io/anej6/.

Attention checks After rating every stimulus in the list, a black-and-white image was shown after which the participant had to indicate what was depicted on it out of a list of answers (this was the same item for all participants, a clearly discernible black-and-white image of a car). In addition, we assessed participants' attention via intrarater reliability. Following the first attention check, the participant was presented with five of their most highly and five of their most lowly rated images (in randomized order). We calculated Spearman's correlation coefficient between the first and second ratings for these ten items.

Online Rating Study – China The procedure of the experiment was identical to that used for the US sample, with the exception

[6] For two US participants, we do not know their definite ages, only which age group they belong to. Mean age and SD are calculated without these two participants.

[7] Duration ranged from 10 min to 53 hours. The latter was an extreme outlier that is no doubt due to a measurement error: For example, the participant may have opened the link to the task but only began completing it after a very long pause.

that Chinese participants were asked to provide demographic data (gender, age, country of residence) after giving consent to the study.[8] The experimental website including the informed consent and instruction pages was translated into Mandarin Chinese by a native speaker (a psychologist) and independently double-checked by another native speaker (a linguist).

Data exclusion

Participants were excluded according to the following criteria: (1) the rate of skipped answers was above 25%; (2) no intrarater correlation could be calculated;[9] (3) for arousal ratings, a failed Attention Check 1 and an intrarater correlation of less than .77 or otherwise a passed Attention Check 1 but an intrarater correlation of less than .67; (4) for valence ratings, a failed Attention Check 1 and an intrarater correlation of less than .80; or otherwise a passed Attention Check 1 but an intrarater correlation of less than .70.[10] In addition, in the Chinese rating, data from participants was excluded if they did not provide a personal identification code (generated by Dynata).

Surprisingly, exclusion rates for the China data (472/1341 = 35.2%) were much larger than that for the US data (29/806 = 3.6%). This may have happened for various reasons, but a major difference is that Prolific (US data) specializes in collecting data for scientific purposes, while Dynata serves mainly as a provider of market research data.

Results

Each image received between 91 and 119 ratings. For each country, we computed mean valence and mean arousal ratings per image.

US rating

Reliability For valence and arousal ratings, separately, we generated 1000 random split halves of our sample and calculated the Spearman–Brown reliability coefficient of the mean ratings. The mean of the correlation coefficients in the valence dimension was $\overline{R}_{val}$ = .986 ($SD = 5.51 \times 10^{-5}$, range: R_{min} = .985 and R_{max} = .986), and in the arousal dimension $\overline{R}_{aro}$ = .921 ($SD = 3.50 \times 10^{-4}$, range: R_{min} = .920 and R_{max} = .929), demonstrating extremely high reliability in both affective dimensions.

[8] The demographic data for people who participated in the representative US rating was provided by Prolific.

[9] This can either happen when the repetition items needed to calculate the correlation were skipped or when the same rating, for example, “5”, was given on every trial.

[10] Exclusion criteria 3–4 were empirically determined by pilot testing in the pilot study, which is described in more detail in the supplementary online material.

Relationship between valence and arousal Mean valence ratings from the US participants lie between 1.11 and 8.35 (M = 5.64, SD = 1.76); mean arousal ratings lie between 2.51 and 7.23 (M = 5.00, SD = 1.02). The correlation between valence and arousal ratings was significant, with $r(581) = -.323$, 95% CI [− .394, − .248], $p < .001$, $BF_{10} = 5.76 \times 10^{12}$. However, the Pearson correlation coefficient shows a relatively weak linear relationship and a look at the distributions of the US mean ratings, as shown in Fig. 1, suggests a U-shaped relationship between valence and arousal in the BASS images, where arousal dips for neutral images but then increases with rising valence.[11]

Effects of demographic variables Prior research has identified differences in emotion processing based on factors like gender (Bradley et al., 2001; Proverbio et al., 2009; Sabatinelli et al., 2004; Wrase et al., 2003) and age (Bradley & Lang, 2007; Grühn & Scheibe, 2008; Pôrto et al., 2010). To investigate potential gender-related differences in our affective ratings, we calculated the individual means for valence and arousal ratings for male and female US participants.

The correlation between the mean ratings made by male and female participants was very high, for both valence, $r(581) = .972$, 95% CI [.967, .976], $p < .001$, $BF_{10} > 10^{200}$, and arousal, $r(581) = .877$, 95% CI [.857, .895], $p < .001$, $BF_{10} = 2.31 \times 10^{182}$.

For age-wise comparison, we performed a median split for the demographic variable age on both of our US subsamples (valence, arousal) and correlated the respective mean ratings of the halves. For the valence rating group (median age = 46 years), correlation between the younger (< 46 years) and the older (≥ 46 years) age group was very high, with $r(581)$ = .976, 95% CI [.972, .979], $p < .001$, $BF_{10} > 10^{200}$. The same was true for the arousal rating group (median age = 44.5 years), with a correlation between the younger (< 45 years) and the older (≥ 45 years) age group of $r(581) = .906$, 95% CI [.891, .920], $p < .001$, $BF_{10} > 10^{200}$. Finally, Fig. 2 shows that the valence-arousal pattern is very similar for all age groups (divided into five brackets following Prolific’s age stratification, see Table 1; see also Fig. 14 in Appendix A).

Chinese rating

Reliability The mean of the correlation coefficients in the valence dimension was $\overline{R}_{val}$ = .980 ($SD = 7.64 \times 10^{-5}$, range: R_{min} = .979 and R_{max} = .980), and in the arousal dimension $\overline{R}_{aro}$ = .921 ($SD = 2.87 \times 10^{-4}$, range: R_{min} = .917 and R_{max} = .923). Interrater reliability for both affective dimensions was again extremely high and almost identical to the US sample.

[11] The same U-shaped relationship was found in the IAPS and OASIS. Our observed valence-arousal correlation is very similar to that of the IAPS, which shows $r(1192) = -.289$, 95% CI [− .340, − .236], $p < .001$, $BF_{10} = 2.16 \times 10^{21}$, while the correlation in the OASIS was not significant, with $r(898) = -.058$, 95% CI [− .123, .007], $p = .082$, $BF_{01} = 2.84$.

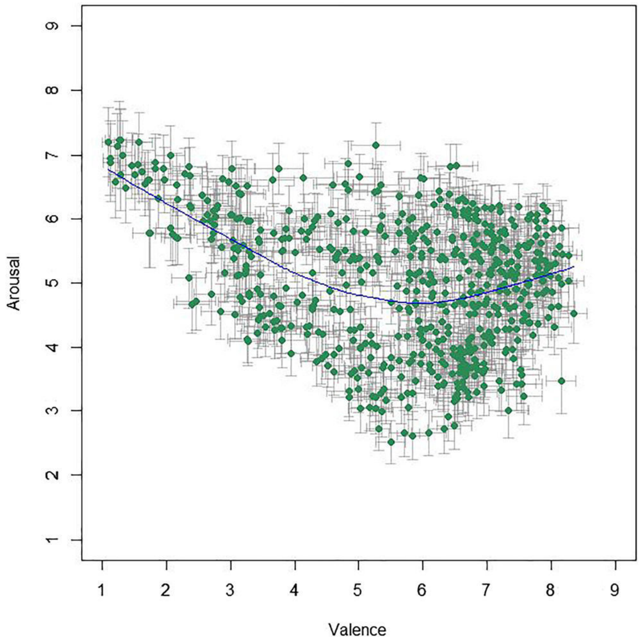


Fig. 1 Mean ratings (95% CI error bars) per image for valence and arousal from US participants. To illustrate the tendency of the relation, the *blue line* depicts locally weighted regression (LOWESS)

Relationship between valence and arousal Mean valence ratings from the Chinese participants lie between 1.47 and 7.84 ($M = 5.46$, $SD = 1.37$); mean arousal ratings lie between 3.65 and 6.75 ($M = 5.31$, $SD = 0.64$). Unexpectedly, and different from the results in the US sample, mean arousal and mean valence ratings showed a strong positive correlation, with $r(581) = .805$, 95% CI [.774, .832], $p < .001$, $BF_{10} = 2.33 \times 10^{129}$. The distributions of the Chinese mean valence and arousal ratings are shown in Fig. 3, which illustrates that negative images received relatively low arousal ratings while positive images received relatively high arousal ratings (as opposed to the US ratings, where arousal ratings were distributed fairly evenly between positive and negative images).

Effects of demographic variables Similarly to the US sample, mean ratings between Chinese male and female participants were highly correlated for valence, $r(581) = .979$, 95% CI [.975, .982], $p < .001$, $BF_{10} > 10^{200}$, and arousal, $r(581) = .856$, 95% CI [.832, .876], $p < .001$, $BF_{10} = 4.11 \times 10^{163}$.

For age-wise comparison, we again performed a median split for age on both Chinese subsamples (valence, arousal). For the valence rating group (median age = 32 years), correlation between the younger (< 32 years) and the older (≥ 32 years) age group was extremely high, with $r(581) = .974$, 95% CI [.970, .978], $p < .001$, $BF_{10} > 10^{200}$. For the arousal rating group (median age = 33 years), the agreement between the younger (< 33 years) and the older (≥ 33 years) age group was still very high, although somewhat lower, $r(581) = .785$, 95% CI [.752, .814], $p < .001$, $BF_{10} = 3.41 \times 10^{118}$. Note that the median age was much lower in the Chinese compared to the US sample: The Chinese sample was split in the early 30s while the US sample was split in the mid-40s.

As for the US sample, five age brackets were created—however, the ratings by Chinese age group "58+" are hardly reliable since this sample consists of a mere 14

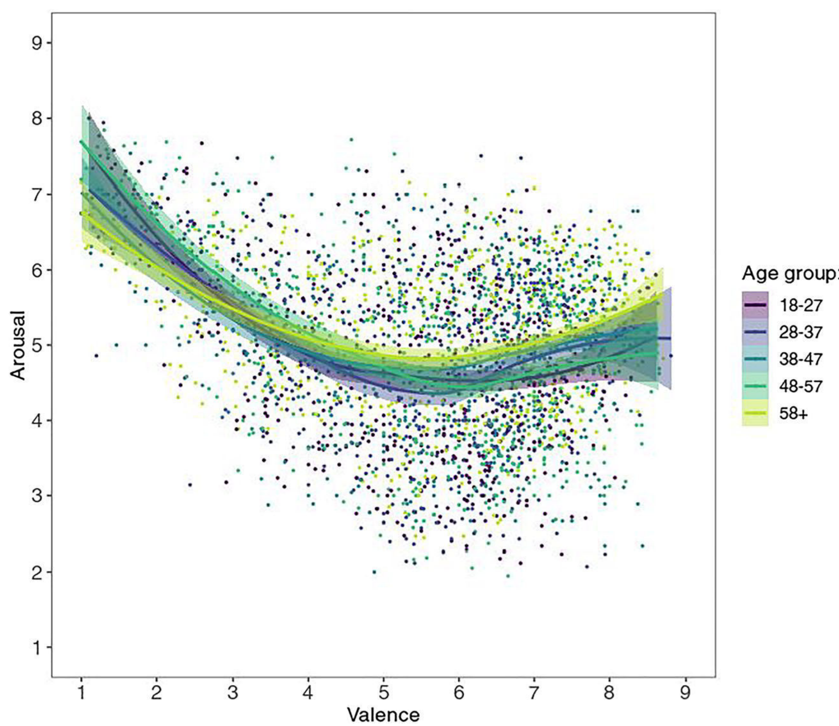


Fig. 2 Mean valence by mean arousal ratings per picture by age group from US participants, with corresponding LOWESS lines

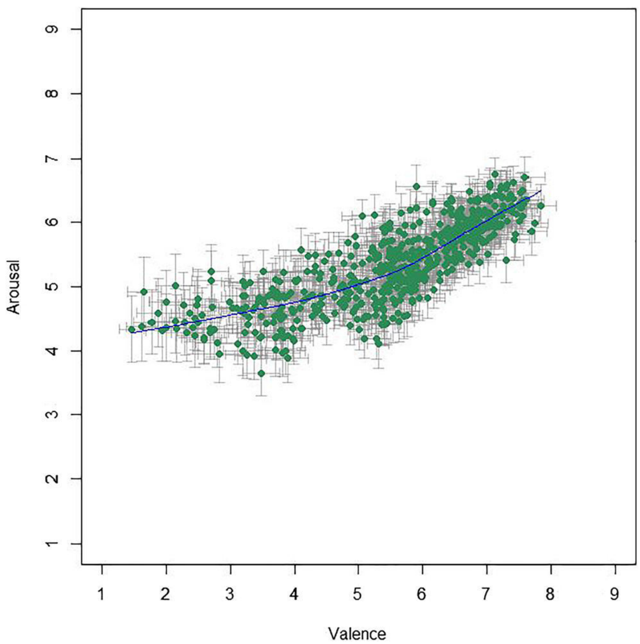


Fig. 3 Mean ratings (95% CI error bars) per image for valence and arousal from Chinese participants with the LOWESS line in *blue*

participants—therefore they are omitted from the present figures. (They are nonetheless largely in line with the rest of the data; see supplementary figures at https://osf.io/anej6/). Most interestingly, the different age groups show somewhat different patterns (see Fig. 4), with ratings from younger age groups, especially the youngest (ages 18–27), tending to resemble more to the US sample and the U-shape often observed in Western samples.

Comparison of US and Chinese ratings

Mean ratings in the valence category from the US sample showed extremely high positive correlation with those from the Chinese rating, $r(581) = .936$, 95% CI [.925, .946], $p < .001$, $BF_{10} > 10^{100}$ (see Fig. 5). A paired *t* test showed that Americans rated images more positive than Chinese, with a statistically significant but practically negligible difference of 0.18, 95% CI [0.12, 0.23] (US: 5.64 ± 1.76, China: 5.46 ± 1.37), $t(582) = 6.35$, $p < .001$, $d = 0.26$, 95% CI [0.18, 0.35], $BF_{10} = 1.21 \times 10^7$. Both these findings are very similar for all age groups and both sexes (see Fig. 6; Table 2).

The mean ratings for arousal showed only a weak, albeit significant, correlation, $r(581) = .282$, 95% CI [.206, .355], $p < .001$, $BF_{10} = 2.30 \times 10^9$ (see Fig. 7). The lower correlation in the arousal (as compared to the valence) dimension relates to the fact that, as mentioned above, the data from US raters showed a U-shaped valence-arousal-relationship, but the Chinese average affective ratings displayed a linear relationship. However, just as the ratings by younger age groups showed more visually similar valence-arousal patterns (see Fig. 4), they also demonstrated substantially higher correlation with US arousal ratings, as most clearly reflected in the youngest age group (see Fig. 8; $r = .53$ as opposed to .22, .17, and –.07 for the other three age groups, see Table 2; $p < .001$ for all *z*-test comparisons; Diedenhofen & Musch, 2015). This

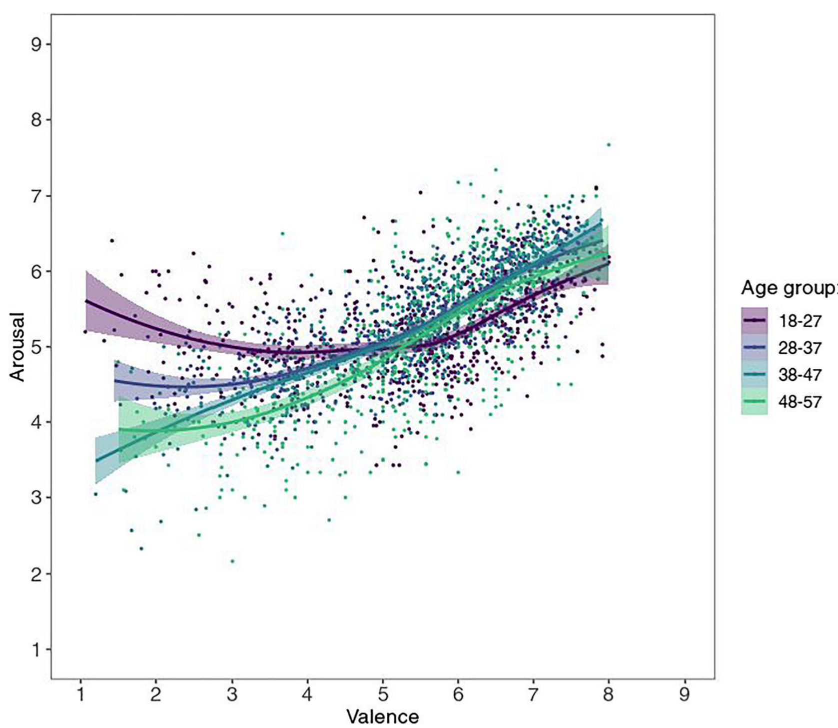


Fig. 4 Mean valence by mean arousal ratings per picture by age group from Chinese participants, with corresponding LOWESS lines

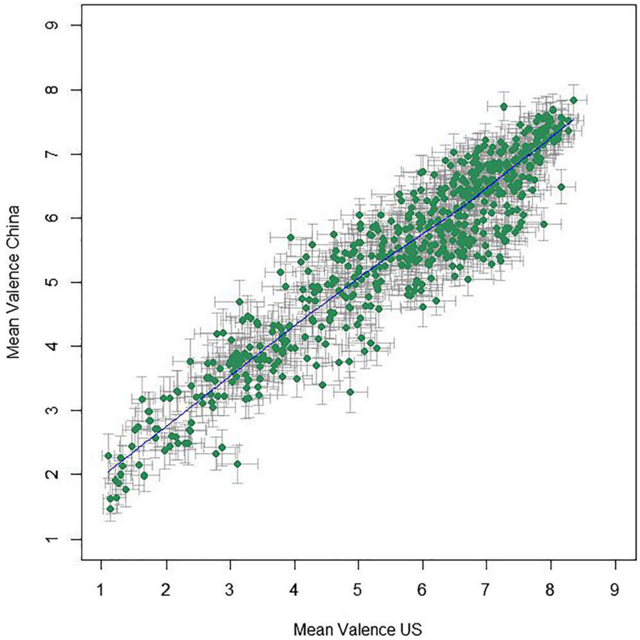


Fig. 5 Relationship between US and Chinese mean valence ratings (95% CI error bars) per image, with the LOWESS line in *blue*

pattern of age group influence is similar for both males and females, though the correlation was generally higher for males (r = .39 vs .20; Table 2; see also all figures via https://osf.io/anej6/ with male and female ratings in separate panels).

A paired t test showed that arousal ratings by Americans were lower than those by Chinese, again a statistically significant yet very small difference, − 0.31, 95% CI [− 0.40, − 0.23] (mean rating US: 5.00 ± 1.02, China: 5.31 ± 0.64), $t(582) = -7.30$, $p < .001$, $d = -0.30$, 95% CI [− 0.39, − 0.22], $BF_{10} = 4.95 \times 10^9$. This relation is similar for all age groups and both sexes (Table 2).

All in all, we may conclude that the consensus for valence ratings was generally higher than for arousal ratings, both within a sample of a given culture (see correlation tests for interrater reliability, gender, and age) and between the two cultures tested here.

Example experiment: Color-valence congruence effects in the BASS

We also included an example experiment, in which we tested the suitability of the BASS images for online research. Here, we manipulated color to demonstrate one major advantage of the BASS images: their easy rendition in a different color. Content-wise, it has been repeatedly shown that the stimulus valence of words interacts with perceptual features like color and brightness: Positive stimuli are associated with green (rather than red) and with white (rather than black; Kawai et al., 2020; Kuhbandner & Pekrun, 2013; Lakens et al., 2012; Meier et al., 2004, 2015; Moller et al., 2009). As a validation of the suitability of the BASS for online research, in a web-based fast decision task, we tested if the BASS silhouettes can induce similar effects. Details about the

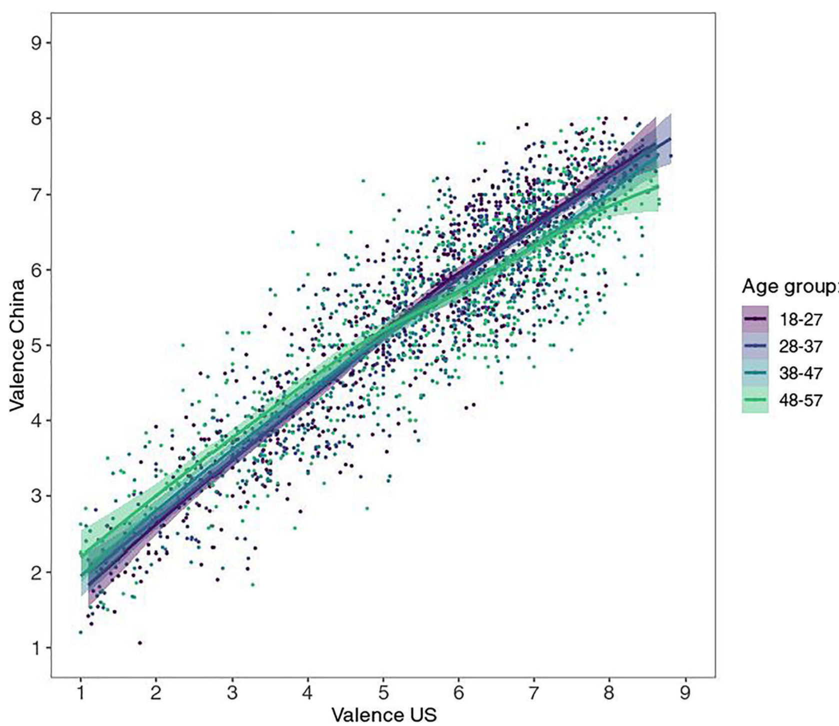


Fig. 6 Relationship between US and Chinese mean valence ratings per image by age group, with corresponding LOWESS lines

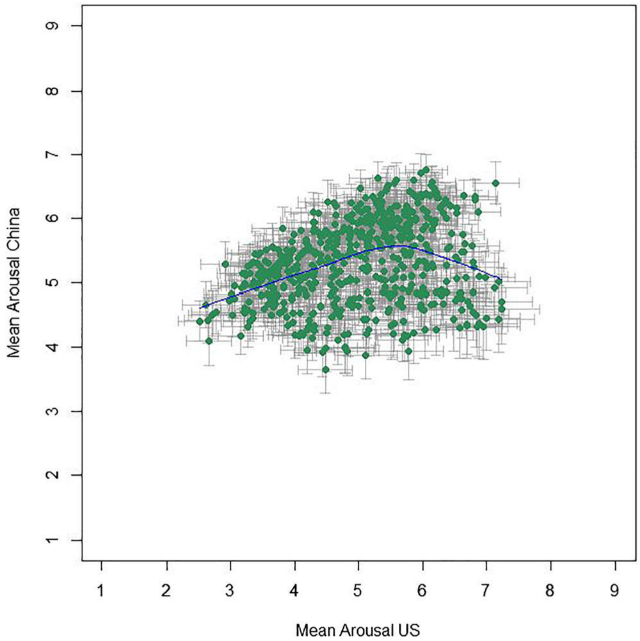


Fig. 7 Relationship between US and Chinese mean arousal ratings (95% CI error bars) per image, with the LOWESS line in *blue*

experiment can be found in the online supplementary material. Below we provide only a brief summary.

Method We selected a subset of 60 positive and 60 negative silhouettes from the BASS (a list of filenames is provided in Table 4 in Appendix C). These 120 images were presented to 90 students (age = 23.0 ± 3.5; 27 male) from the University of Vienna in two experimental blocks: a "color" block, containing each silhouette once in red and once in green color; and a "brightness" block, containing each silhouette once in black and once in white (with silhouette background always in gray, see Appendix C, Fig. 15, for examples). The task was to categorize the valence of each silhouette via key press as either positive or negative. For each block, we ran a repeated-measures analysis of variance (ANOVA) on the correct mean response times (RTs) and another on the mean error rates (ERs).

Results According to expectations, the two (Color: red vs. green; within) × two (Valence: positive vs. negative; within) ANOVA in the color block showed a significant interaction between color and valence for RTs, $F(1, 89) = 91.46, p < .001, \eta^2_p = .507$, 90% CI [.383, .595], $\eta^2_G = .022, BF_{10} = 3.10 \times 10^9$, and also for ERs, $F(1, 89) = 83.67, p < .001, \eta^2_p = .485$, 90% CI [.359, .576], $\eta^2_G = .106, BF_{10} = 2.30 \times 10^{15}$. Positive silhouettes were categorized faster and more accurately when presented in green rather than red, and, inversely, negative silhouettes were categorized faster and more accurately when presented in red rather than green. Also, in the brightness block, the silhouettes elicited the expected brightness–valence interactions in both RTs, $F(1, 89) = 18.68, p < .001, \eta^2_p = .173$, 90% CI [.068, .286], $\eta^2_G = .005, BF_{10} = 65.74$, and ERs, $F(1, 89) = 8.76, p = .004, \eta^2_p = .090$, 90% CI [.017, .191], $\eta^2_G = .011, BF_{10} = 7.55$. Positive silhouettes were categorized faster and more

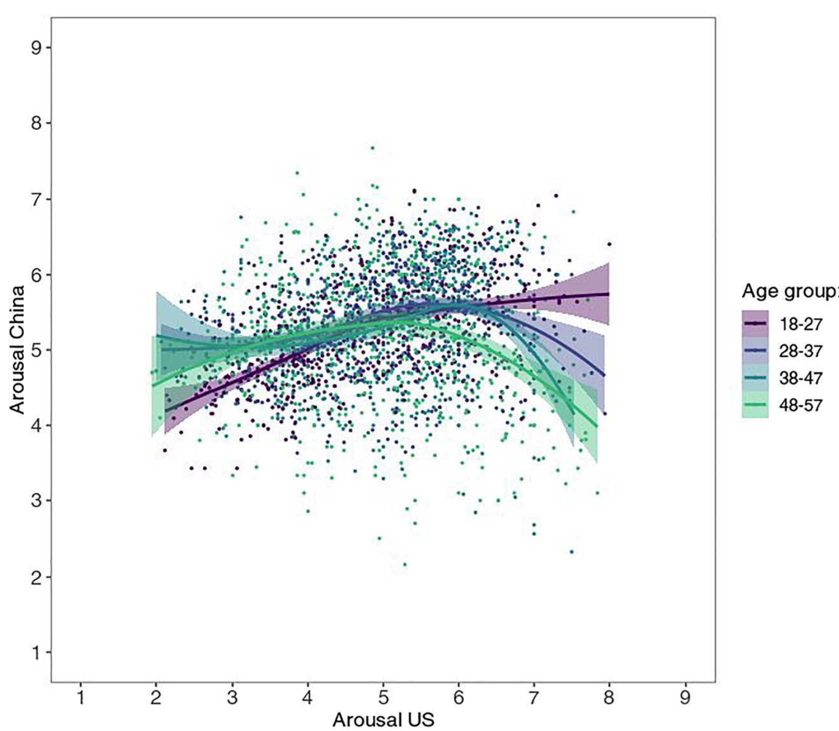


Fig. 8 Relationship between US and Chinese mean arousal ratings per image by age group, with corresponding LOWESS lines

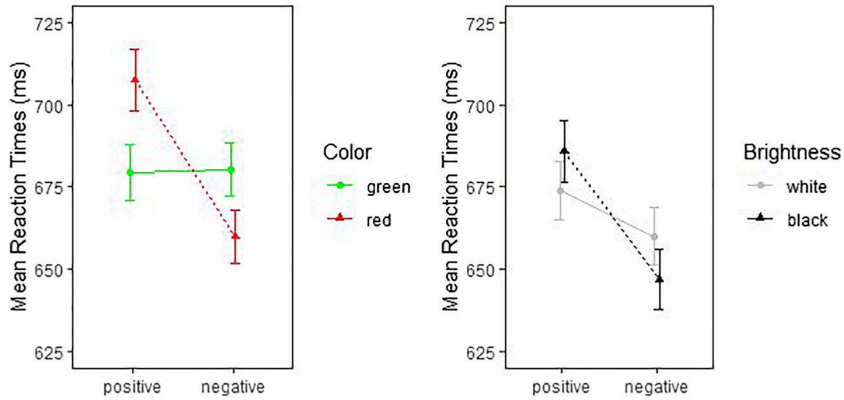


Fig. 9 Mean reaction times for positive and negative silhouettes in the color block (*left*) and the brightness block (*right*). Error bars indicate SEMs

accurately when presented in white rather than black, and inversely, negative silhouettes were categorized faster and more accurately when presented in black rather than white.

Means and *SD*s for RTs and ERs are illustrated in Figs. 9 and 10, and the means can be found in Table 5.

Discussion

For the current study, we created an affective silhouette database with the specific goal of providing an easy and openly accessible lightweight stimulus pool capable of reliable elicitation of affective representations on a wide range of valence and arousal levels that was validated by extensive samples of two of the biggest culture groups studied in psychology. In the US sample, we found that the images cover a very wide range of affective representations from very negative to very positive (valence), and from very calm to very intense (arousal), indicating various potential applications (see Introduction). The Chinese sample was remarkably similar in valence ratings, covering a wide range with an extremely high correlation with the US sample.

However, the Chinese *arousal* ratings were only weakly correlated, and comparatively limited in range. This relates to the fact that data from the US participants demonstrated a U-shaped relationship between valence and arousal, while data from the Chinese participants, instead, displayed a more linear valence–arousal relationship (strong positive correlation between valence and arousal ratings) on average.

Studies that look at affective ratings for emotional *words* in Chinese find a U-shaped valence-arousal relation similar to the one apparent in the US ratings (Ho et al., 2015; Liu et al., 2018; Yao et al., 2017). It is noteworthy, that participants in the affective word studies were high school or university students. Our Chinese sample was more diverse: The age-group-wise analysis shows that younger Chinese display a more quadratic relationship which becomes more linear with increasing age. This provides a plausible reconciliation of these previous findings with our results in that it is only the relatively older Chinese generations whose perception substantially differs from those in Western populations, while the younger generations perceive images more in accordance with the Western population. We can only speculate about the reasons, but the more similar ratings of young US and Chinese participants could reflect a larger shared sphere of visual experiences, for instance, through digital social media (e.g., TikTok).

However, two studies that compared affective ratings for emotional *photographs* (a subset of the IAPS) between Chinese and Western samples interestingly seem to indicate

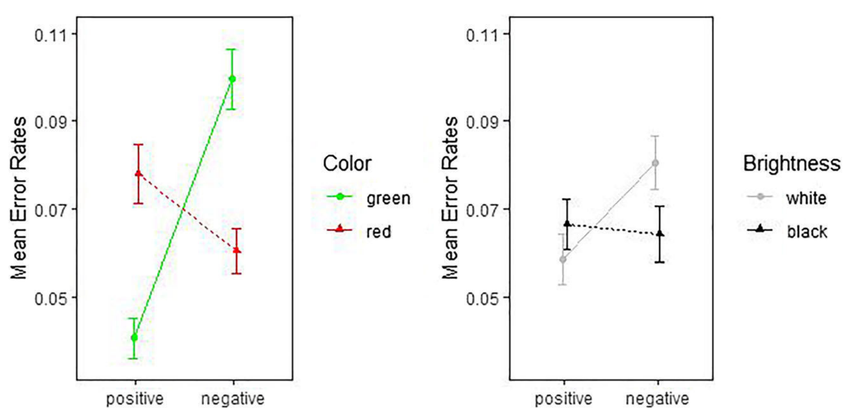


Fig. 10 Mean error rates for positive and negative silhouettes in the color block (*left*) and the brightness block (*right*). *Error bars* indicate the SEMs

the opposite phenomenon: Gong and Wang (2016) found a high correlation of valence and arousal ratings between Chinese and German older adults[12], and Huang et al. (2015) argued that Chinese young adults display a different rating behavior from US young adults for IAPS pictures.

Generally, more empirical data on cultural differences in emotion representation or judgments is necessary to establish how and when Chinese rating behavior differs from that of a Western culture. One possibility might be, for instance, that the lower arousal for negative affective silhouettes, which explains most of the difference in our BASS ratings (and the lack of U-shape among Chinese participants, especially with increasing age), might be a result of conflict avoidance or emotion control as a more generally accepted strategy among Chinese than Westerners (Gong & Wang, 2016; Tjosvold & Sun, 2002).

We also demonstrated how the BASS images could be used in online experiments. In our example experiment, we showed the typical color-congruence effects of stimulus color on speed and accuracy of judgments about image valences: Positive images were rated faster and more accurately if presented in green than in red, and if presented in white than in black, while the opposite held true of negative images.[13] This result replicated prior research and already proves at least one particular use of the BASS images in scientific research.

All in all, the quality of our samples' data (as measured by the respective interrater correlation) is outstanding and, where comparable, its general characteristics are in line with effects observed in notable affective databases like IAPS, OASIS, ANEW and BAWL-R (Võ et al., 2009), to name a few.

The BASS, however, aims only to complement, and not to replace, realistic picture databases, for research for which realistic images do not provide ideal stimuli for a given experiment for any of the reasons detailed in the Introduction (e.g., complications of color manipulation, file size, etc.).

While the simplicity of silhouettes can be advantageous for some research designs, their innate lack of context information may simultaneously lead to more (or less) ambiguity in their interpretations between participants than is the case for pictures that are rich in detail. Heterogeneous cultural backgrounds between participant groups could add to differences in content evaluation if contexts interact with objects in photorealistic image sets, but not or at least less so with silhouettes where context is sparse. However, while the direct interpretation of some of the silhouettes may, thus, be more or less ambiguous than that of a photographic image of the corresponding content (e.g., some participants may perceive a coyote's silhouette simply as a dog when presented without a fitting habitat as context), in the current study, the affective ratings of the silhouettes generally demonstrated very high consistencies (very high inter-rater correlation, low *SD*s, and, in case of valence, very high intercultural correlation), which attest to the silhouettes' emotional unambiguousness. Furthermore, several numeric measures in our BASS database, such as the rating differences between US and China per image, the corresponding *BF*s or residuals, or simply the *SD*s of the ratings in either sample, allow researchers to choose images of emotional clarity or culture-neutrality suiting their needs best.

Nonetheless, and although the rating data of our BASS images correspond well with the data obtained for photographic databases, we cannot rule out that the silhouettes evoke smaller emotional responses than contextually richer photographic pictures when using other measures (e.g., skin conductance, heart rate, facial muscles activities). Do photorealistic pictures elicit stronger emotion-related physiological responses than silhouettes, and if so, what are the driving forces in such a case (object detail, color, depth and perspective, etc.)? We hope that future research, facilitated by the BASS, will help to address these questions.

Conclusions

The BASS offers a novel database of visual silhouettes for psychological research optimally suited for the study of image-elicited affective representations in online experiments and validated by normative ratings by representative Western and Asian samples.

Acknowledgements Claudia Kawai and Gáspár Lukács are recipients of DOC Fellowships of the Austrian Academy of Sciences at the Department of Cognition, Emotion, and Methods in Psychology at the University of Vienna. The research was further supported by Dr. Franziska Pinsker, for which we thank her. We want to acknowledge the assistance of Yang Zhang, who translated the experiment to Mandarin and collected the online pilot data in China. We also thank Wenyi Chu for her feedback on clarity of the experimental procedure, and Sophie Hanke, Tobias Greif, and Bence Szaszkó for making browsing the BASS easier. The meticulous

[12] Gong and Wang (2016) compared their Chinese participant sample with a mean age of 67.3 ± 4.96 years to a German participant sample with a mean age of 69.61 ± 3.58 years (Grühn & Scheibe, 2008). Thus, in this comparison, participants were older than our oldest Chinese age group (58+) which had a mean age of 62.36 ± 5.70 years and included only 14 participants.

[13] The finding may seem to contradict the pilot study results showing that black-on-white and (inverted) white-on-black images are rated very similarly, demonstrating equal recognizability. However, in the pilot study, each participant was presented with only one image type (either all images black-on-white, or all white-on-black). Color-valence congruence effects can depend on dichotomous decision alternatives within the same task (cf. Kawai et al., 2020; Lakens et al., 2012), and, the present results are in line with this possibility and previous research showing that such congruence effects were only elicited in the example experiment (where each task contained contrasting color types alternating trial by trial), not during the affective rating study.

comments Anna Walker provided are greatly appreciated. Special thanks goes to the talented Anna Lumaca (allhailthesnail.com) for creating and contributing 20 much needed "ghastly" silhouettes to the BASS.

Availability of data, materials, and code The BASS and its accompanying data table are available for download from the OSF project page under https://osf.io/anej6/. The project page further contains all preregistrations, collected raw data, experimental and analysis scripts, a useful R-script to convert the colors of the silhouettes, as well as detailed reports on the procedures involved in the creation of the BASS and its follow-up experiment. A browsable and downloadable version of the complete database (silhouette images and data table) is also available via https://gasparl.github.io/BASS/.

Authors' contributions Concept by C.K. and U.A., images by C.K., study design and analysis by C.K. and G.L., presentation software by G.L. and C.K., manuscript by C.K., G.L., and U.A.

Funding Open access funding provided by University of Vienna. This research was funded by the Austrian Academy of Sciences (C.K.'s personal grant number 25068, G.L.'s personal grant number 24945).

Declarations

Ethics, consent to participate, and consent for publication The research was conducted according to the principles expressed in the Declaration of Helsinki. During online participation, no potentially identifying information such as IP addresses was collected and participants were informed that (i) their data would be treated anonymously, and (ii) they could unconditionally stop their participation. Participants were informed that their anonymized data would be published and were given the opportunity to prohibit the use of their data. Informed consent was obtained from all subjects.

Conflict of interest The authors have no competing interests.

Appendix A

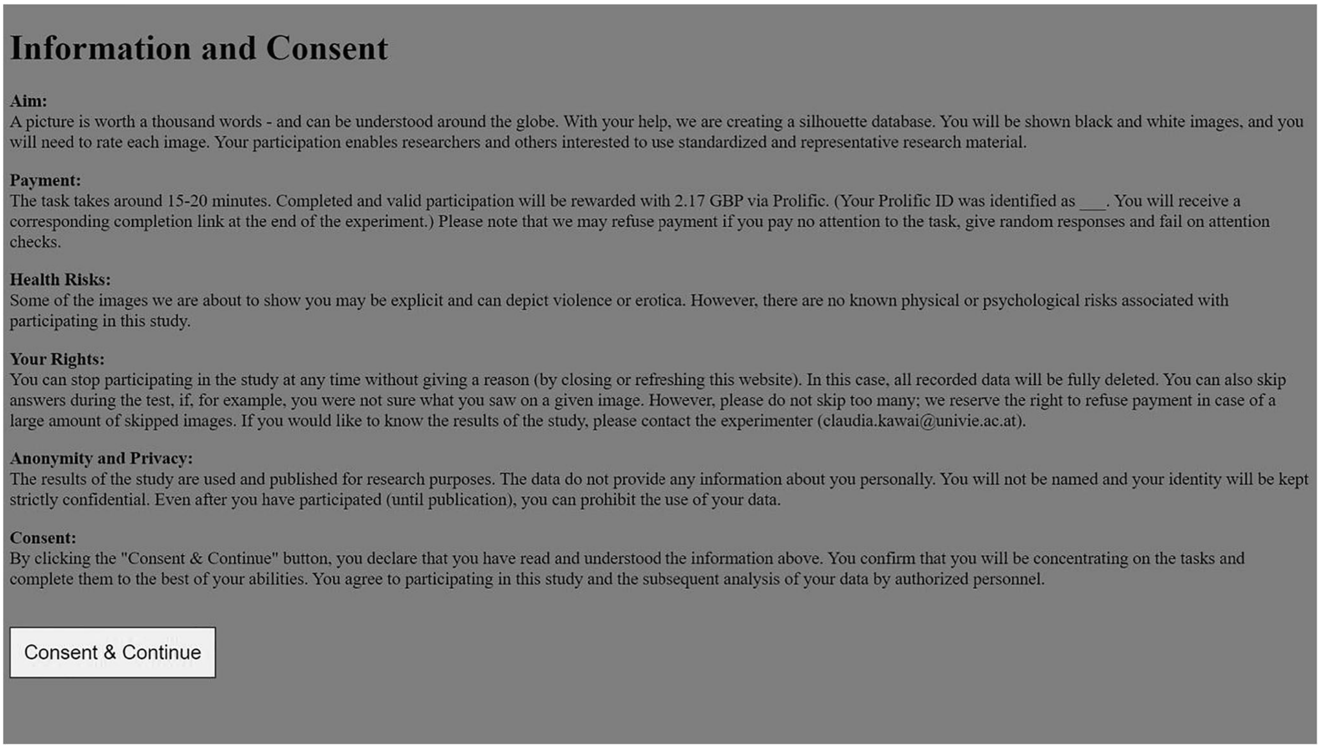
Table 1 Demographics of the included participants from the US and Chinese ratings with corresponding means and *SD* for valence and arousal

	US	Mean Valence Rating (*SD*)	Mean Arousal Rating (*SD*)	China	Mean Valence Rating (*SD*)	Mean Arousal Rating (*SD*)
Sex						
Female	402	5.70 (2.39)	5.00 (2.25)	419	5.49 (1.97)	5.25 (1.84)
Male	375	5.58 (2.17)	4.99 (2.28)	450	5.45 (1.94)	5.38 (1.82)
Sum	**777**			**869**		
Age						
18–27	138	5.41 (2.23)	4.85 (2.24)	186	5.42 (1.99)	5.23 (1.85)
28–37	147	5.55 (2.13)	4.90 (2.33)	419	5.48 (1.97)	5.38 (1.9)
38–47	132	5.66 (2.36)	5.04 (2.21)	176	5.47 (1.94)	5.32 (1.68)
48–57	134	5.70 (2.26)	4.96 (2.39)	74	5.50 (1.85)	5.20 (1.75)
58+	226	5.78 (2.37)	5.17 (2.16)	14	5.47 (1.78)	5.45 (1.74)
Ethnicity						
Asian	58					
Black	106					
Mixed	35					
Other	30					
White	548					

Table 2 US and Chinese valence and arousal rating correlation per sex and age group

	Valence correlation [95% CI]	Arousal correlation [95% CI]
Sex		
Female	.913 [.899, .926]	.198 [.119, .275]
Male	.942 [.933, .951]	.387 [.316, .454]
Age		
18–27	.910 [.895, .923]	.530 [.469, .586]
28–37	.921 [.908, .933]	.219 [.140, .295]
38–47	.899 [.882, .913]	.170 [.090, .248]
48–57	.856 [.833, .876]	– .069 [– .150, .012]
58+ *	.689 [.644, .730]	.251 [.173, .326]

* This oldest age group includes only 14 Chinese participants (cf. Table 1), hence the related correlations are not reliable.


Information and Consent

Aim:
A picture is worth a thousand words - and can be understood around the globe. With your help, we are creating a silhouette database. You will be shown black and white images, and you will need to rate each image. Your participation enables researchers and others interested to use standardized and representative research material.

Payment:
The task takes around 15-20 minutes. Completed and valid participation will be rewarded with 2.17 GBP via Prolific. (Your Prolific ID was identified as ___. You will receive a corresponding completion link at the end of the experiment.) Please note that we may refuse payment if you pay no attention to the task, give random responses and fail on attention checks.

Health Risks:
Some of the images we are about to show you may be explicit and can depict violence or erotica. However, there are no known physical or psychological risks associated with participating in this study.

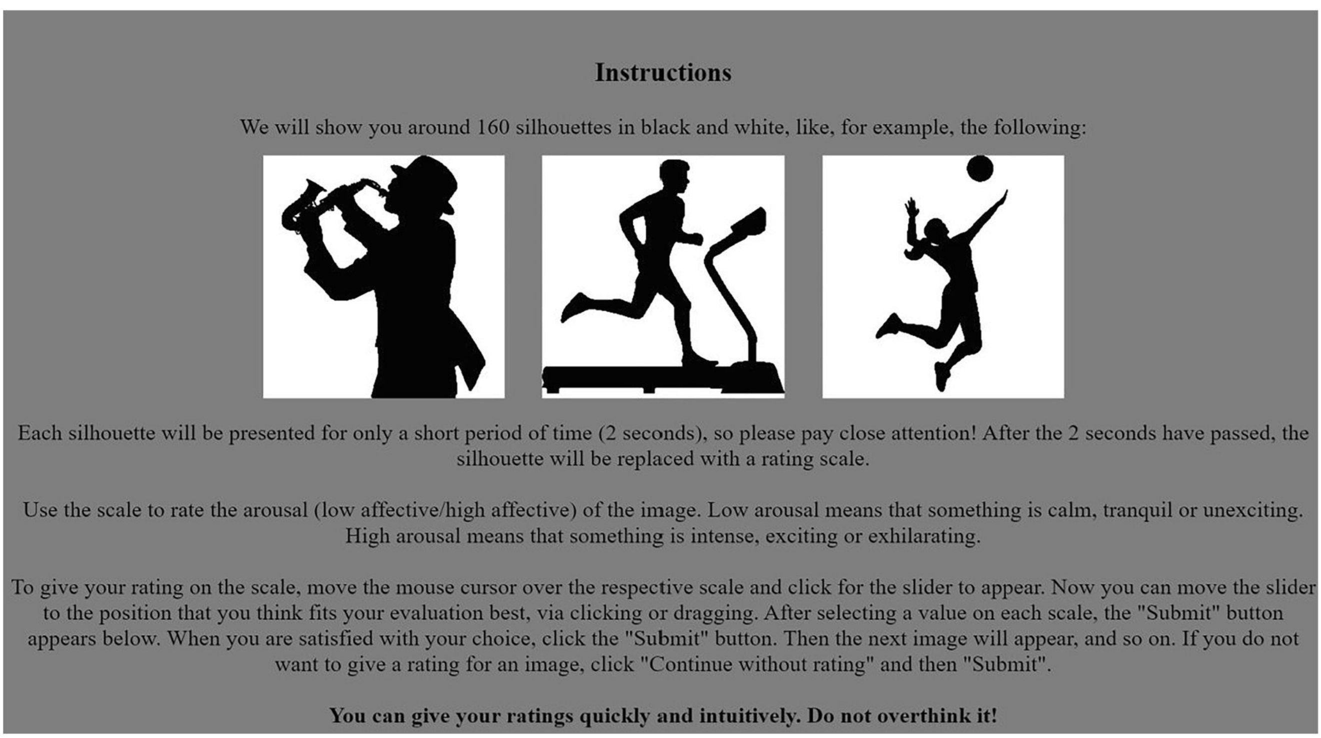
Your Rights:
You can stop participating in the study at any time without giving a reason (by closing or refreshing this website). In this case, all recorded data will be fully deleted. You can also skip answers during the test, if, for example, you were not sure what you saw on a given image. However, please do not skip too many; we reserve the right to refuse payment in case of a large amount of skipped images. If you would like to know the results of the study, please contact the experimenter (claudia.kawai@univie.ac.at).

Anonymity and Privacy:
The results of the study are used and published for research purposes. The data do not provide any information about you personally. You will not be named and your identity will be kept strictly confidential. Even after you have participated (until publication), you can prohibit the use of your data.

Consent:
By clicking the "Consent & Continue" button, you declare that you have read and understood the information above. You confirm that you will be concentrating on the tasks and complete them to the best of your abilities. You agree to participating in this study and the subsequent analysis of your data by authorized personnel.

Consent & Continue

Fig. 11 Screenshot of the experiment's welcome page


Instructions

We will show you around 160 silhouettes in black and white, like, for example, the following:

Each silhouette will be presented for only a short period of time (2 seconds), so please pay close attention! After the 2 seconds have passed, the silhouette will be replaced with a rating scale.

Use the scale to rate the arousal (low affective/high affective) of the image. Low arousal means that something is calm, tranquil or unexciting. High arousal means that something is intense, exciting or exhilarating.

To give your rating on the scale, move the mouse cursor over the respective scale and click for the slider to appear. Now you can move the slider to the position that you think fits your evaluation best, via clicking or dragging. After selecting a value on each scale, the "Submit" button appears below. When you are satisfied with your choice, click the "Submit" button. Then the next image will appear, and so on. If you do not want to give a rating for an image, click "Continue without rating" and then "Submit".

You can give your ratings quickly and intuitively. Do not overthink it!

Fig. 12 Screenshot of the instruction page after consenting (example from the arousal substudy)

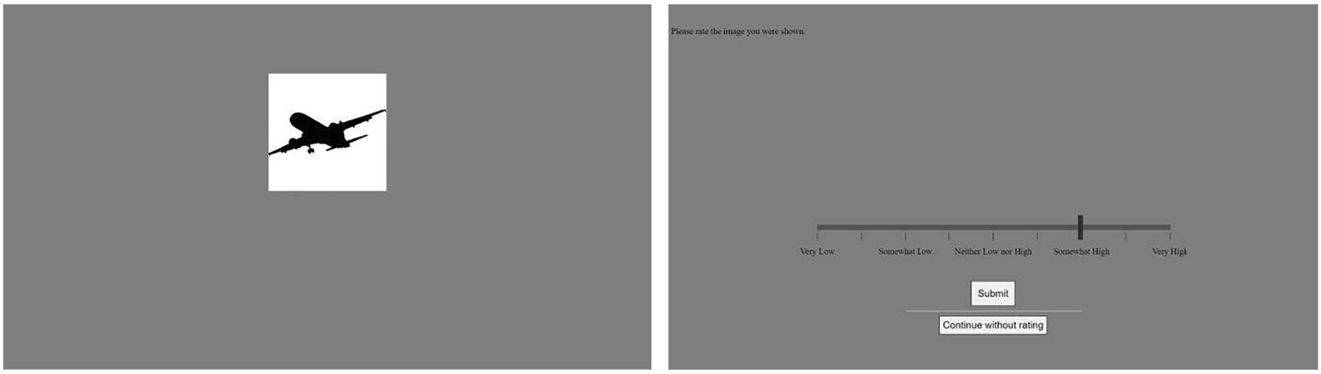


Fig. 13 Screenshot of the presentation of the silhouette stimulus (start of the trial), followed by the appearance of the rating scale (example from the arousal substudy)

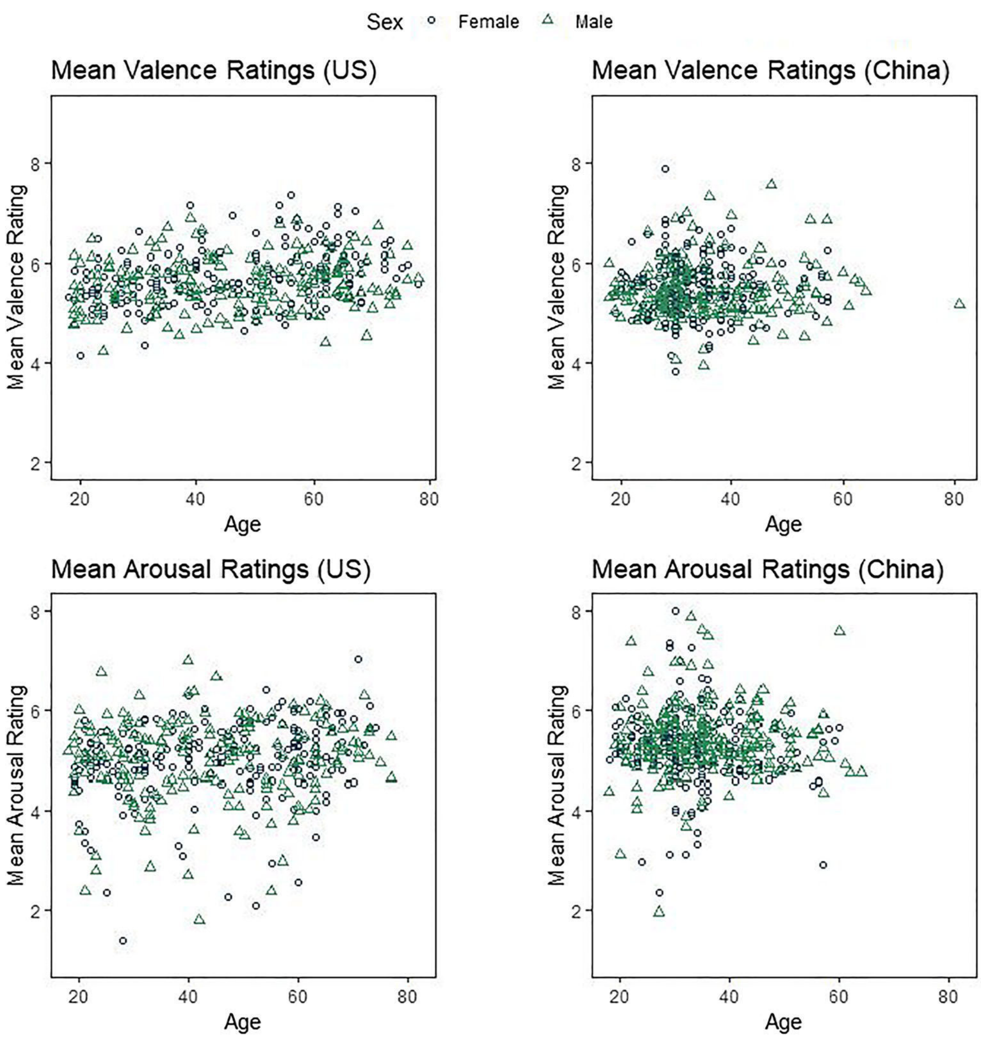


Fig. 14 Mean valence and mean arousal ratings per participant according to their country, sex, and age group

Appendix B

Table 3 Bicolor Affective Silhouettes and Shapes (BASS) datatable manual

Column	Explanation
file_name	The image file name.
val_mean_us	The mean of all valence ratings.
aro_mean_us	The mean of all arousal ratings.
val_sd_us	The *SD* of all valence ratings.
aro_sd_us	The *SD* of all arousal ratings.
val_n_us	The number of valence ratings.
aro_n_us	The number of arousal ratings.
val_cilo_us	Lower limit of 95% CI for the mean of all valence ratings.
val_ciup_us	Upper limit of 95% CI for the mean of all valence ratings.
aro_cilo_us	Lower limit of 95% CI for the mean of all arousal ratings.
aro_ciup_us	Upper limit of 95% CI for the mean of all arousal ratings.
val_diff	The difference between the means of US and Chinese valence ratings. (val_mean_us minus val_mean_ch).
aro_diff	The difference between the means of US and Chinese arousal ratings. (aro_mean_us minus aro_mean_ch).
val_bf_10	Wilcoxon Bayes factor in support of difference between US and Chinese. valence ratings.
val_bf_01	Wilcoxon Bayes factor in support of equivalence between US and Chinese. valence ratings.
aro_bf_10	Wilcoxon Bayes factor in support of difference between US and Chinese. arousal ratings.
aro_bf_01	Wilcoxon Bayes factor in support of equivalence between US and Chinese. arousal ratings.
val_resid	Linear regression residuals for valence ratings: observed values minus fitted. values for the US-Chinese mean ratings relations for the entire sample. A larger value means larger US mean valence rating than Chinese mean valence rating, relative to what is expected when values are perfectly correlated (linear fitted values). The closer this value is to zero, the more likely it is that the given image is suitable for cross-cultural measurement of valence. (Nonetheless, this value is quite close to val_diff, which is more straightforward to interpret.)
aro_resid	Linear regression residuals for arousal ratings (same as for valence, see above).
black_px	Number of black pixels in the image.
white_px	Number of white pixels in the image.
keywords	Keywords for the given image.
source	Source of the (original) image the silhouette is based on.

Note. The columns listed above with "_us" suffixes refer to rating data from the US sample. The data referring to the Chinese ratings is indicated with "_ch" suffixes – here omitted for brevity. (E.g., val_mean_ch: The mean of all valence ratings from China, etc.)

Appendix C

Table 4 Filenames of the positive and negative silhouettes used in the verification study

Positive silhouettes		Negative silhouettes	
trapeze4	badminton	injury	handinjury
wedding2	jump	attack2	grenade
love	highjump4	execution3	robber
singer3	highjump	abuse2	bazooka
couple6	acrobat3	abuse	pollution2
romance	couple5	hanging	gunman
volleyball	skiing2	drunkard	drunk
champion	dance2	vomit2	tick
euphoria	ferriswheel	argument4	axe3
iceskater3	concert	handcuff2	scorpion2
success2	iceskater	noose	mosquito
affection2	skateboard	argument3	gasmask2
cardio	trapeze3	argument2	handcuff
handball	tennis	dispute	knife2
kiss4	handball2	argument6	pollution
ballet	winner	hazard	mosquito2
basketball4	rockstar	hitman	knife4
propose	carefree	sorrow2	snake2
kiss3	basketball	murder2	wasp
joy	breakdance	gunman2	machete
acrobat	skiing3	beggar2	tank2
trapeze	soccer	skull2	soldier2
jump2	aerobics	scolding2	gun
happiness	highjump2	gunman3	kalashnikov
surfer2	snowboard2	knife	tank
skiing	couple7	axe	wildwest
soccer2	gymnastics3	accident2	sniper
father2	highjump3	roach2	rifle
propose3	dog3	vomit	skull
skateboard2	poledance	falling6	snake

Fig. 15 Example for color-manipulated silhouettes (from *left to right*): original, red-gray, green-gray, black-gray, and white-gray (as used in the example experiment)

Table 5 Means and *SD*s for reaction times (RTs) and error rates (ERs) per color and valence categories in the Bicolor Affective Silhouettes and Shapes (BASS) validation experiment

Color	Valence	RT $M \pm SD$	ER $M \pm SD$
Red	Negative	659.96 ± 200.61	0.06 ± 0.24
Red	Positive	708.00 ± 233.20	0.08 ± 0.27
Green	Negative	680.91 ± 207.14	0.10 ± 0.30
Green	Positive	679.24 ± 217.30	0.04 ± 0.20
Black	Negative	647.60 ± 202.37	0.06 ± 0.25
Black	Positive	686.14 ± 217.78	0.07 ± 0.25
White	Negative	660.16 ± 202.40	0.08 ± 0.27
White	Positive	674.30 ± 210.28	0.06 ± 0.24

References

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2020). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01501-5

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*(4), 621–635. https://doi.org/10.1017/S0142716403000316

Baker, A., Mystkowski, J., Culver, N., Yi, R., Mortazavi, A., & Craske, M. G. (2010). Does habituation matter? Emotional processing theory and exposure therapy for acrophobia. *Behaviour Research and Therapy*, *48*(11), 1139–1143. https://doi.org/10.1016/j.brat.2010.07.009

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11*(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115–147. https://doi.org/10.1037/0033-295X.94.2.115

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science, 24*(6), 981-990. https://doi.org/10.1177/0956797612465439

Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: Sex differences in picture processing. *Emotion*, *1*(3), 300–319. https://doi.org/10.1037/1528-3542.1.3.300

Bradley, M. M., & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In *Handbook of emotion elicitation and assessment* (pp. 29–46). Oxford University Press.

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, *43*(2), 468–477. https://doi.org/10.3758/s13428-011-0064-1

Deák, A., Csenki, L., & Révész, G. (2010). Hungarian ratings for the International Affective Picture System (IAPS): A cross-cultural comparison. *Empirical Text and Culture Research*, *4*(8), 90–101.

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE, 10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion, 15*(2), 115-141. https://doi.org/10.1080/02699930042000024

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin*, *99*(1), 20–35. https://doi.org/10.1037/0033-2909.99.1.20

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 104–115. https://doi.org/10.1037/0096-1523.13.1.104

Garaizar, P., & Reips, U.-D. (2019). Best practices: Two web-browser-based methods for stimulus presentation in behavioral experiments with high-resolution timing requirements. *Behavior Research Methods, 51*(3), 1441–1453. https://doi.org/10.3758/s13428-018-1126-4

Geusebroek, J.-M., Burghouts, G. J., & Smeulders, A. W. M. (2005). The Amsterdam Library of Object Images. *International Journal of Computer Vision*, *61*(1), 103–112. https://doi.org/10.1023/B:VISI.0000042993.50813.60

Gong, X., & Wang, D. (2016). Applicability of the International Affective Picture System in Chinese older adults: A validation study: Cross-cultural validity of the IAPS. *PsyCh Journal*, *5*(2), 117–124. https://doi.org/10.1002/pchj.131

Grühn, D., & Scheibe, S. (2008). Age-related differences in valence and arousal ratings of pictures from the International Affective Picture System (IAPS): Do ratings become more extreme with age? *Behavior Research Methods*, *40*(2), 512–521. https://doi.org/10.3758/BRM.40.2.512

Hansen, T., & Gegenfurtner, K. R. (2009). Independence of color and luminance edges in natural scenes. *Visual Neuroscience, 26*(1), 35–49. https://doi.org/10.1017/S0952523808080796

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition & Emotion, 15*(2), 143-165. https://doi.org/10.1080/02699930125768

Ho, S. M. Y., Mak, C. W. Y., Yeung, D., Duan, W., Tang, S., Yeung, J. C., & Ching, R. (2015). Emotional valence, arousal, and threat ratings of 160 Chinese words among adolescents. *PLOS ONE*, *10*(7), e0132294. https://doi.org/10.1371/journal.pone.0132294

Huang, J., Xu, D., Peterson, B. S., Hu, J., Cao, L., Wei, N., Zhang, Y., Xu, W., Xu, Y., & Hu, S. (2015). Affective reactions differ between Chinese and American healthy young adults: A cross-cultural study using the international affective picture system. *BMC Psychiatry*, *15*(1), 60. https://doi.org/10.1186/s12888-015-0442-9

Huston, J. P., Nadal, M., Mora, F., Agnati, L. F., & Conde, C. J. C. (2015). *Art, aesthetics, and the brain*. OUP Oxford.

Kawai, C., Lukács, G., & Ansorge, U. (2020). Polarities influence implicit associations between colour and emotion. *Acta Psychologica*, *209*, 103143. https://doi.org/10.1016/j.actpsy.2020.103143

Kuhbandner, C., & Pekrun, R. (2013). Joint effects of emotion and color on memory. *Emotion*, *13*(3), 375–379. https://doi.org/10.1037/a0031821

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, *49*(2), 457–470. https://doi.org/10.3758/s13428-016-0715-3

Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J., & Midden, C. J. H. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion*, 27(7), 1225–1246. https://doi.org/10.1080/02699931.2013.781501

Lakens, D., Semin, G. R., & Foroni, F. (2012). But for the bad, there would not be good: Grounding valence in brightness through shared relational structures. *Journal of Experimental Psychology: General*, *141*(3), 584–594. https://doi.org/10.1037/a0026468

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*, 39–58.

Lau, R. R., Sears, D. O., & Jessor, T. (1990). Fact or artifact revisited: Survey instrument effects and pocketbook politics. *Political Behavior*, *12*(3), 217–242. https://doi.org/10.1007/BF00992334

Leder, H., Tinio, P. P. L., & Bar, M. (2011). Emotional valence modulates the preference for curved objects. *Perception*, *40*(6), 649–655. https://doi.org/10.1068/p6845

Lindsay, G. W. (2020). Attention in psychology, neuroscience, and machine learning. *Frontiers in Computational Neuroscience*, *14*, 29. https://doi.org/10.3389/fncom.2020.00029

Liu, P., Li, M., Lu, Q., & Han, B. (2018). Norms of valence and arousal for 2,076 Chinese 4-character words. In K. Hasida & W. P. Pa (Eds.), *Computational Linguistics* (Vol. 781, pp. 88–98). Springer Singapore. https://doi.org/10.1007/978-981-10-8438-6_8

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods*, *46*(2), 596–610. https://doi.org/10.3758/s13428-013-0379-1

Marr, D. (1982). *Vision. A computational investigation into the human representation and processing of visual information*. W. H. Freeman and Company.

Meier, B. P., Fetterman, A. K., & Robinson, M. D. (2015). Black and white as valence cues: A large-scale replication effort of Meier, Robinson, and Clore (2004). *Social Psychology*, *46*(3), 174–178. https://doi.org/10.1027/1864-9335/a000236

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science, 15*(4), 243-247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white: Automatic inferences about stimulus valence based on brightness. *Psychological Science*, *15*(2), 82–87. https://doi.org/10.1111/j.0963-7214.2004.01502002.x

Moller, A. C., Elliot, A. J., & Maier, M. A. (2009). Basic hue-meaning associations. *Emotion*, *9*(6), 898–902. https://doi.org/10.1037/a0017811

Mould, D., Mandryk, R. L., & Li, H. (2012). Emotional response and visual attention to non-photorealistic images. *Computers & Graphics, 36*(6), 658–672. https://doi.org/10.1016/j.cag.2012.03.039

Nishimoto, T., Ueda, T., Miyawaki, K., Une, Y., & Takahashi, M. (2012). The role of imagery-related properties in picture naming: A newly standardized set of 360 pictures for Japanese. *Behavior Research Methods*, *44*(4), 934–945. https://doi.org/10.3758/s13428-011-0176-7

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41*, 176–210. https://doi.org/10.1006/cogp.1999.0728

Orzan, A., Bousseau, A., Barla, P., & Thollot, J. (2007). Structure-preserving manipulation of photographs. *Proceedings of the 5th International Symposium on Non-Photorealistic Animation and Rendering - NPAR '07, 103*. https://doi.org/10.1145/1274871.1274888

Pôrto, W. G., Bertolucci, P. H. F., & Bueno, O. F. A. (2010). The paradox of age: An analysis of responses by aging Brazilians to International Affective Picture System (IAPS). *Revista Brasileira de Psiquiatria*, *33*(1), 10–15. https://doi.org/10.1590/S1516-44462010005000015

Proverbio, A. M., Adorni, R., Zani, A., & Trestianu, L. (2009). Sex differences in the brain response to affective scenes with or without humans. *Neuropsychologia*, *47*(12), 2374–2388. https://doi.org/10.1016/j.neuropsychologia.2008.10.030

Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *The Journal of Neuroscience*, *38*(33), 7255–7269. https://doi.org/10.1523/JNEUROSCI.0388-18.2018

Ramaswami, M. (2014). Network plasticity in adaptive filtering and behavioral habituation. *Neuron*, *82*(6), 1216–1229. https://doi.org/10.1016/j.neuron.2014.04.035

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217–236. https://doi.org/10.1068/p5117

Sabatinelli, D., Flaisch, T., Bradley, M. M., Fitzsimmons, J. R., & Lang, P. J. (2004). Affective picture perception: Gender differences in visual cortex? *NeuroReport*, *15*(7), 1109–1112. https://doi.org/10.1097/00001756-200405190-00005

Schmalz, X., Beyersmann, E., Cavalli, E., & Marinus, E. (2016). Unpredictability and complexity of print-to-speech correspondences increase reliance on lexical processes: More evidence for the orthographic depth hypothesis. *Journal of Cognitive Psychology*, *28*(6), 658–672. https://doi.org/10.1080/20445911.2016.1182172

Schupp, H. T., Markus, J., Weike, A. I., & Hamm, A. O. (2003). Emotional facilitation of sensory processing in the visual cortex. *Psychological Science, 14*(1), 7-13. https://doi.org/10.1111/1467-9280.01411

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215.

The GIMP Development Team. (2019). GIMP. Retrieved from https://www.gimp.org

Verschuere, B., Crombez, G., & Koster, E. (2001). The International Affective Picture System: A Flemish Validation Study. *Psychologica Belgica*, *41*(4), 205–217. https://doi.org/10.5334/pb.981

Tjosvold, D., & Sun, H. F. (2002). Understanding conflict avoidance: Relationship, motivations, actions, and consequences. *International Journal of Conflict Management, 13*(2), 142–164. https://doi.org/10.1108/eb022872

Võ, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

Wilcox, N., & Wlezien, C. (1993). The contamination of responses to survey items: Economic perceptions and political judgments. *Political Analysis*, *5*, 181–213. https://doi.org/10.1093/pan/5.1.181

Wilms, L., & Oberfeld, D. (2018). Color and emotion: Effects of hue, saturation, and brightness. *Psychological Research, 82*(5), 896–914. https://doi.org/10.1007/s00426-017-0880-8

Wrase, J., Klein, S., Gruesser, S. M., Hermann, D., Flor, H., Mann, K., Braus, D. F., & Heinz, A. (2003). Gender differences in the processing of standardized emotional visual stimuli in humans: A functional magnetic resonance imaging study. *Neuroscience Letters*, *348*(1), 41–45. https://doi.org/10.1016/S0304-3940(03)00565-2

Yao, Z., Wu, J., Zhang, Y., & Wang, Z. (2017). Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behavior Research Methods*, *49*(4), 1374–1385. https://doi.org/10.3758/s13428-016-0793-2

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111*(4), 493–504. https://doi.org/10.1037/pspa0000056

2.3 **Kawai, C.**, Zhang, Y., Lukács, G., Chu, W., Zheng, C., Gao, C., Gozli, D., Wang, Y., & Ansorge, U. (2021). *The Good, the Bad, and the Red: Implicit Color-Valence Associations Across Cultures* [Manuscript submitted for publication]. Department of Cognition, Emotion, and Methods in Psychology, University of Vienna.

The Good, the Bad, and the Red: Implicit Color-Valence Associations Across Cultures

Claudia Kawai[1], Yang Zhang[2], Gáspár Lukács[1], Wenyi Chu[3], Chaoyi Zheng[4], Cijun Gao[5], Davood Gozli[5], Yonghui Wang[2], and Ulrich Ansorge[1,6,7]

[1]Department of Cognition, Emotion, and Methods in Psychology, University of Vienna, Austria

[2]School of Psychology, Shaanxi Normal University, Xi'an, China

[3] Nexus International School, Putrajaya, Malaysia

[4]Department of Psychology, Central China Normal University, Wuhan, China

[5]Department of Psychology, University of Macau

[6]Vienna Cognitive Science Hub, University of Vienna

[7]Research Platform Mediatised Lifeworlds, University of Vienna

Author Note

The authors have no financial or personal conflicts of interest. Claudia Kawai is a recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Department of Cognition, Emotion, and Methods in Psychology at the University of Vienna. This funding source had no role or involvement related to this study. Supplementary material and data are available

under the OSF project page under osf.io/27gf8 (link masked until publication; for an anonymized link see https://osf.io/27gf8/?view_only=17708da2ad6b4d8c95cf10f5ac26356f).

Correspondence concerning this article should be addressed to Claudia Kawai, Faculty of Psychology, University of Vienna, Liebiggasse 5, A-1010 Vienna, Austria. E-mail: claudia.kawai@univie.ac.at. The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Contributions: concept, data acquisition, and statistical analyses by C.K.; study design and software by C.K. and G.L.; preregistrations by C.K. and U.A.; manuscript drafted by C.K., revised and proofread by U.A., G.L., and D.G.; material selection across countries by C.K., W.C., and Y.Z.; instruction translations to Cantonese by C.G., and into Mandarin by C.Z. and Y.Z.; data acquisition in Macau in the lab of D.G. was managed by C.K. and C.Z.; data acquisition in Xi’an in the lab of Y.W. was managed by C.K. and Y.Z.

Abstract

Cultural differences – as well as similarities – have been found in explicit color-emotion associations between Chinese and Western populations. However, implicit associations in a cross-cultural context remain an understudied topic, despite them being sensitive to more implicit knowledge. Moreover, they can be used to study color systems – that is, associations with one color in the context of an alternative color. Therefore, we tested the influence of two different color oppositions on affective stimulus categorization: red versus green and red versus white, in two experiments. In Experiment 1, stimuli comprised positive and negative words, and participants from the West (Austria/Germany), and the East (Mainland China, Macau) were tested in their native languages. The Western group showed a significantly stronger color-valence interaction effect than the Mainland Chinese (but not the Macanese) group for red-green but not for red-white opposition. To explore color-valence interaction effects independently of word stimulus differences between participant groups, we used affective silhouettes instead of words in Experiment 2. Again, the Western group showed a significantly stronger color-valence interaction than the Chinese group in red-green opposition, while effects in red-white opposition did not differ. These findings complement those from explicit association research in an unexpected manner: Where explicit measures showed similarities between cultures (associations for red and green), our results revealed differences; and where explicit measures showed differences (associations with white), our results showed similarities, underlining the value of applying comprehensive measures in cross-cultural research on cross-modal associations.

Keywords: polarity correspondence; congruence; color; emotion; cross-cultural comparison

The Good, the Bad, and the Red: Implicit Color-Valence Associations Across Cultures

Humans form associations between colors and affective valence, such as between black and negative valence versus white and positive valence (Lakens et al., 2013). These associations are important for many different applications. For instance, coloration of relevant signals for machine users (e.g., coloring of traffic lights, on-off buttons, etc.) could be designed in a more or less intuitive way (Garrido et al., 2019; Hochman et al., 2018). Critically, such signals are typically based on color systems, in which at least two different colors take on different – oftentimes even opposing – functions or meanings. For example, in some countries, white characters on blue or green backgrounds are used for traffic signs for guidance, while red circles are used for regulation signs. A very conventional system uses green or white traffic lights for signaling go and red for signaling stopping actions. In addition, buttons – physically on machines or virtually on the Internet and mobile applications – systematically use the opposing "call-to-action" colors red for *cancel*, *error*, and *decrease*, and green for *ok*, *success*, and *add*. It is also common to mark gaming tokens or playing cards by different colors (e.g., black and white chess pieces, or red vs. black for uneven vs. even numbers in roulette).

Importantly for the cross-cultural success of such applications, underlying associations between color and valence could apply universally or they could differ between cultures. Studies that used explicit ratings of color-emotion or color-valence associations found more cross-cultural similarities than differences (e.g., Adams & Osgood, 1973; Barchard et al., 2017; Hupka et al., 1997; Jonauskaite, Abu-Akel, et al., 2020; Volkova et al., 2012; T. Wang et al., 2014). Such universal color-valence or color-emotion associations could reflect communalities between globally shared knowledge (i.e., a global or globalized culture) but they could also be of an evolutionary origin (for the logic, see Darwin, 1872;

Ekman & Friesen, 1982). Take the example of the color red. In line with a cross-cultural similarity and an evolutionary origin of some color-valence associations, male human and nonhuman primates (Macaca Fuscata) show a preference for red over green contrast in lips or faces of conspecifics, presumably reflecting the fact that red lips look healthier to humans and that red skin signals fertility in macaques (e.g., Pflüger et al., 2014; Stephen & McKeegan, 2010; see also Elliot & Niesta, 2008; Elliot et al., 2010; but see Peperkoorn et al., 2016; Lehman et al., 2018). However, especially to red, connections to emotions of very different character seem to be prevalent. When asked to verbally name emotions associated with the color red, humans frequently also stated negative emotions such as anger besides positive ones like love (Jonauskaite, Abu-Akel, et al., 2020). Not all of these associations must be shared between cultures. Some of them might also be *culture-specific* (e.g., He, 2011; see also discussions below).

Critically, what is currently lacking are more implicit measures of the cultural universality of these color-valence associations to complement the explicitly obtained findings, which would allow for a better understanding of the conditions for the dominance of a particular color-valence association where there are several possible candidates, as in the case of the color red. To start with the importance of implicit measures, although explicit ratings and judgments carry a high face validity and allow for an economical and encompassing measure of the different emotions associated with each color, explicit ratings also have a number of drawbacks. Most notably, they depend on the participants' awareness of the associations, meaning also that participants need to remember these associations explicitly to report them in the first place (Squire, 1986). Explicit ratings are, thus, not always suited to tap into the more implicit forms of memory, including associations built on conditioning, for example (cf. Squire, 1986; Squire et al., 1993).

Regarding the much-researched color red: There are several findings from studies using implicit measures that corroborate results obtained from explicit measures. An association between red and negative valence does not only show in explicit naming or rating tasks but also in implicit tests like valence classifications. For example, if human participants have to categorize words as negative or positive, they are faster and more accurate to categorize negative words in red than in green or blue (e.g., Kawai et al., 2020; Moller et al., 2009; Schietecat et al., 2018b).

Even though red is associated with both positive (e.g., love) and negative (e.g., anger) associations, in the context of the usage of the color green as an alternative, at least in Western cultures, red-negativity associations can dominate behavior in the valence-categorization task: Responses are faster in congruent conditions, with negative words in red and positive words in green, than in an incongruent condition, with negative words in green and positive words in red (Kawai et al., 2020). The reason might be a general strong green-positivity association that was also found in cross-cultural (explicit association) studies (e.g., Jonauskaite, Abu Akel, et al., 2020) and that is also reflected in positive meanings of sayings such as *green is the color of hope* and positive connotations of metaphors surrounding nature and ecology such as *green medicine* or *green infrastructure* in various languages (Escobedo et al., 2019; Väliverronen & Hellsten, 2002). The negative valence of the color red might, thus, be particularly salient where the color green is used as an alternative in implicit measures such as the valence categorization of colored words (Kawai et al., 2020). However, while such a particularly shaped opposition seems to exist in Western cultures, this does not necessarily have to be the case in other cultures.

Exact color meanings and, thus, color contexts can differ between cultures. For example, in Chinese *red*, 红 (hóng), is used in the term *wedding*, 红事 (hóng shì, composed of ‘red’ + ‘matter’), but this is not the case in Western languages such as English, German, French, or Spanish. Maybe red for Chinese participants is comparatively more positively (or less negatively) associated than for Western participants? As no study ever tested cross-culturally which of the different red-valence associations is dominant in implicit color-valence categorization measures, it is also possible that we find differences between the cultures in the present study that more explicit measures might be insensitive to.

Furthermore, as explained, specific colors take on different roles or meanings in different contexts (cf. Elliot & Maier, 2012), and the specific color used as an alternative provides a particular context that could be decisive for which color-valence association dominates. To study the role of such color systems in a cross-culturally varying context, we also manipulated the alternative color to red. Specifically, He (2011) argued that in “Chinese culture, white is contrary to red” (p. 161). This color symbolism is reflected in tradition and language. In the Beijing opera, for example, the hero wears a red face mask and the adversary a white one (China National Tourist Office, 2020). In contrast to the term *red* being used in Chinese in the positive term *wedding* (see above), in Simplified Chinese *white*, 白 (bái), is used in the negatively evaluated term *funeral*, 白事 (bái shì, composed of ‘white’ + ‘matter’) (He, 2011). Again, as was the case with the terms *red* and *wedding*, no such connection between the terms *white* and *funeral* exists in Western languages such as English, German, French, or Spanish. In addition, explicit ratings of emotions confirmed the selective presence of negative emotions associated to the color white among Chinese compared to Western cultures (Jonauskaite et al., Abu Akel, et al., 2020).

This raises an additional important point: Without further testing, it is unclear which color-emotion (or in a broader sense color-valence) association is predominant in a particular context (cf. Elliot & Maier, 2012). Which association of red prevails in a situation could be a function of factors such as the specific culture of the users or the use of a particular color system. We also addressed this second open question, the possibility of a different and culturally more variable color dominance in an alternative color system, in our experiment: In a second color opposition condition, positive and negative stimuli were presented in red and white rather than in red and green to the Western (Austrian/German) and Chinese samples.

Regarding hypotheses, in the simplest case, we would expect an even greater similarity in size (and direction) of the color-valence congruence effects across cultures in implicit measures than has been found in explicit measures (Jonauskaite, Abu Akel, et al., 2020). This could be the case, for example, if the explicit measure reflects culture-specific associations whereas implicit measures tap more into phylogenetically shared roots of color-valence associations. In the current study this would mean across all tested cultures faster responses to congruent (e.g., green-positive, red-negative) than to incongruent (e.g., green-negative, red-positive) color-valence pairings as has been demonstrated in Western samples (e.g., Kawai et al., 2020).

It is, however, generally possible that our implicit measures reveal cross-cultural differences. If this difference is driven by *dissimilarities in color-valence associations*, following results from explicit measures, these differences (reflected in congruence effect size; i.e., incongruent performance minus congruent performance score) should concern the colors red and white (Jonauskaite, Abu Akel, et al., 2020). In the most extreme case, only among Chinese participants, but not among Western participants, would white be predominately negatively associated. This would then be reflected in inverse color-valence congruence effects for the color white in the Chinese as compared to the Western sample. However, given that cross-cultural similarities are reported for green and (to a lesser extent) for red in explicit associations (see above), we would expect congruence effects to be maybe more comparable in size between cultures with respect to green and red.

In addition, if a (cultural) color-emotion difference is driven by *dissimilarities in the effect of color-systems* (i.e., color-opposition), a context factor we argue to be crucial in the formation and shaping of cross-modal associations (see also Kawai et al., 2020) – and a factor that has not been (systematically) captured by explicit association studies – we will maybe even see differences between red-green and red-white color opposition conditions. For instance, the color red could be more negatively associated in the context of the color green than in the context of the color white, at least for Chinese participants. Such influences would be also a source of differences between explicit and implicit measures and would signal the need to use both complementary methods for a full understanding of the practical implications of color-valence associations.

In the first experiment of the current study, we tested color-valence associations implicitly in an Eastern culture from Mainland China and compared it to a sample "in-between" Eastern and Western cultures from Macau – now a part of China, but a former Western Portuguese/European colony (until 1999) –, and one sample from a Western population from Austria. We used positive and negative words in either red or green color or in either red or white color, and we asked our participants to categorize the words by their valence (as positive or negative).

In Experiment 2, we went one step further and used positive and negative silhouettes instead of words. Since it is impossible to control for equivalence in all characteristics of translated words between different languages (e.g., their lengths, their transparencies, their orthographic neighbors, etc., as used in Experiment 1), we used silhouettes as stimuli in the second experiment, that were genuinely identical and thus comparable between the two different cultures tested: Chinese (Mainland China) and Western (Germany/Austria). Experiment 2 was carried out as an online study, and we used simple silhouettes rather than realistic photographs, as it was easier to manipulate the colors of the silhouettes without corrupting their meaning altogether than it would have been the case with photos. In addition, silhouettes are not that rich in visual detail referring to flora, fauna, objects, buildings, landscape, weather conditions, traffic, clothing, etc., and, thus, they allow more control with respect to specific cultural content than photographs.

Experiment 1

The color red is – according to implicit and explicit association studies – ambiguous. Research demonstrated associations with negative (e.g., anger and danger, see Fetterman et al., 2012; He, 2011; Pravossoudovitch et al., 2014; Sorokowski et al., 2014) as well as positive emotions (e.g., high status, see Wu et al., 2018) across different populations. In contrast, green and white have been shown to be mostly positively associated in Western cultures, while white, in terms of explicit associations, additionally received negative connotations by Chinese participants (e.g., Jonauskaite, Abu Akel, et al., 2020; Jonauskaite et al., 2019; here see also Lakens et al., 2012), making it a more ambivalent color for this culture group. Here, we used an implicit measure and tested which of the resulting potential color-valence associations – that between red and negativity or that between red and positivity – is predominant as a function of two factors: culture (Chinese vs. Western) and color context (red vs. green; red vs. white). For our implicit measures of color-valence associations, we used varying font colors (cf. Jonauskaite, Parraga, et al., 2020) rather than color words (cf. Jonauskaite, Abu Akel, et al., 2020), as the usage of physical color allows more control over what participants actually see and incorporate into their judgments.[1]

Our cross-cultural comparison tested color-valence associations in Austria, Mainland China (from here on simply "China"/"Chinese"), and Macau. Apart from Portuguese, the official (and predominantly used) language in Macau is Cantonese; the writing system most widely used is Traditional Chinese characters. In Xi'an, where we collected our Chinese sample, Mandarin was the language we used in our tests (since it was widely understood) in Simplified Chinese characters. Due to the cultural dissimilarities between Macau and China – and motivated by research demonstrating different color-associations between China and Hong Kong, another special administrative region (Jiang et al., 2014) – we hypothesized that culture-based associations made by participants from Macau should fall in-between those from Austria and China.

In general, based on prior research reviewed above, we expected more similarities between cultures than differences, especially in the red-green condition (cf. Jonauskaite, Abu Akel, et al., 2020), but maybe less cross-cultural similarities and more differences in the red-white condition, where white could be associated with negative valence among Chinese participants only, in turn, "boosting" the opposite positive valence of the color red (cf. He, 2011; Jonauskaite, Abu Akel, et al., 2020). The latter prediction follows from the fact that colors take on particular emotional meaning in the context of specific color systems (cf. He,

[1] When color words are used, it is unclear which exact color participants imagine, as each color label covers a range of physically different colors. In addition, with visual color words instead of real colors, there is also a complicating influence of incongruence between word color and word-color meaning. For example, in Jonauskaite, Abu Akel, et al. (2020) all color words were shown in the same achromatic presentation color (gray/black), meaning that they were cast in incongruent colors relative to their meaning at times. In comparison to color words, the current experiment, thus, at least allowed some control over color appearances (cf. T. Wang et al., 2014).

2011; Kawai et al., 2020) and contexts have generally the potential to change which meaning currently dominates behavior (cf. Elliot & Maier, 2012).

Method

Participants

Data from 281 participants were collected; 104 in Austria (University of Vienna), 90 in Macau (University of Macau), and 87 in Mainland China (Shaanxi Normal University, Xi'an, China). Participants were randomly assigned to one of the conditions that resulted from permuting block order and key location.

We excluded participants with a reported country of origin other than Germany or Austria for the German-speaking group (n = 10), Mainland China for the Chinese group (n = 1), and Macau for the Macanese[2] group (n = 47). We also excluded all participants who did not reach the full score in the color-deficiency test or were self-reportedly color-blind (n = 4) and those who classified less than 40 stimuli per valence category in accordance with our valence assignment (n = 2, see *Procedure*). No participant had an accuracy rate lower than 75%. This left a total of 217 participants: group Austria with 91 subjects (M_{age} = 20.9±2.8; 15 male), group Macau with 41 subjects (M_{age} = 19.6±1.7; 15 male), and group China with 85 subjects (M_{age} = 18.9±1.8; 15 male).

Material

For the experimental groups in Austria, we used the same German words as stimuli as in Kawai et al. (2020). From the Berlin Affective Word List Reloaded (BAWL-R) database

[2] Throughout the paper, we use "Macanese" as an adjective denoting "from Macau" or "pertaining to Macau". We do not use the adjective in reference to the ethnic group (Macanese People). We also do not use it as a proper noun that denotes the Creole language.

(Võ et al., 2009), 60 positive (mean emotion value ≥ 0.6) and 60 negative (mean emotion value ≤ –0.6) German words were selected. The number of nouns, verbs and adjectives was balanced, and values for word arousal, imageability, letter count and word frequency (Cai & Brysbaert, 2010) were kept constant across the two valence categories. For the studies in China, the German word list was translated by a native Mandarin speaker to Mandarin (to be used in Mainland China, written in simplified Chinese characters) and Cantonese (used in Macau, written in traditional Chinese characters). This list was checked and verified by two native Cantonese and two native Mandarin speakers.

To ensure isoluminance for the stimuli presented in red and green, brightness values for red, green, and grey (background color) were measured with a spectrophotometer (X-Rite i1XTreme, Grand Rapids, MI, USA) for each of the five monitors used in the Austrian sample that completed the red-green mixed condition. Color values were selected accordingly. For the samples from China and Macau, balanced RGB values for red (213,0,0) and green (0,213,0) were selected, and a medium gray (128,128,128). Since white has no chroma and is practically ultimate brightness of the monitor, equating values for hue and luminance in the red-white blocks was not feasible. We only had the chance to measure isoluminance of the displayed color values on the monitors in Austria but not at the test sites in China/Macau. However, color values were identical between the experimental sessions in China and Macau and lighting conditions were kept similar between all three countries. Additionally, the visual appearance of the colors on the monitors in the laboratories in China/Macau was judged by the first author of the present study to be sufficiently similar to that in the Austrian sample.

The complete stimulus list as well as a table with all colors and monitor resolutions that were used throughout the experiments are available in the online supplementary material on osf.io/27gf8.

In Austria, the size of the colored word stimuli was set to 50 pixels (angular size 1.45°), with a fixed viewing distance between eye and center display of 60 cm, through the utilization of chin rests. In the labs in China and Macau, viewing distance was approximately 50 cm, but not fixed.

Procedure

After signing the consent forms, participants were asked to provide demographic information (age, gender, country of origin). Instructions about the experimental procedure were then presented on screen. All text (instructions, stimuli, labels, etc.) was written in the participants' native language. The study consisted of two tasks: an initial valence-rating task and a subsequent binary valence-classification task. The experimental session ended with a short test for color deficiency.

Valence Ratings. To ensure that every participant agreed with the valence category of a given stimulus, a rating task *('Please rate the valence of the word.')* preceded the valence classification task. Participants judged each of 120 potential target words on a 10-point Likert scale (from 'very negative' to 'very positive') by moving the mouse cursor to the corresponding tick mark and confirming their selected valence value with a left-click. Words that received a rating below six were classified as 'rated negative' and appeared in a text box on the left side of the screen, below the negative scale label; words with a rating of six or higher were coded as 'rated positive' and appeared in a text box on the right side of the screen, below the positive scale label. Each word stayed on screen (centered above the rating scale) until the judgment was made. There were no time constraints for this task and participants were informed about this. After the participants rated all the words, the 50 most positively and the 50 most negatively rated words were selected for the upcoming valence-categorization task.[3] The maximum of stimuli presented in the categorization task was 100, while there was no minimum of items specified in the experiment. However, as mentioned above, we set per valence category (positive, negative) a lower boundary of at least 40 correctly rated words as participant exclusion criterion.

Valence Categorization. The binary categorization task started after participants had read the instructions on the monitor, which informed them that, per each trial, they would be presented with a single word (which they had previously seen in the valence rating task). Each target word was shown for a maximum of 2 s. For each word, participants judged the valence (*Is this word positive or negative?*) and indicated their choice by pressing either the 'E' (for positive valence) or the 'I' (for negative valence) key (key assignment was balanced across participants). We did not inform participants in advance that words would be presented in different colors. The task started with a 10-trial practice. Participants from China and Macau completed two blocks – a red-green color block and a red-white color block (order of presentation was counterbalanced). For the Austrian group, the data analyzed here came from

[3] If the number of words per category was lower than 50, the shortest word list (list of words rated positive or list of words rated negative by the participant) determined the overall number of stimuli that appeared in the subsequent task. For an equal number of positive and negative items, the word list of the remaining valence category was trimmed accordingly. For example, if a participant rated 43 of the 60 positive words (per BAWL-R database) as positive and 52 of the 60 negative words as negative, then the 43 positive-rated words were selected for the participant's valence-categorization task, as well as the same number of negative-rated words, in which case the total number of items would be 43 + 43 = 86 items. Note that the number of positively rated (e.g., 43) and negatively rated words (e.g., 52) does not necessarily have to add up to 120. In the example, 43 out of 60 positive words were "correctly" judged by the participant as, in fact, positive, while 17 of the 60 positive words were judged as negative. Furthermore, the participant judged 52 out of the 60 negative words as, in fact, negative, but judged eight of the negative words as positive (e.g., 'naïve' is a negative term according to the BAWL-R database, but could be judged as a more positive concept by some participants).

two distinct studies, one with red and green, the other with red and white stimuli.[4] Stimuli were presented in randomized order, with the restrictions that not more than five words of the same valence and/or color were shown in a row. Participants completed the entire experiment in less than 30 min.

Color Vision Test. After participants completed valence rating and valence categorization, we asked them to enter the numbers printed on three color plates, which were displayed on the computer screen (digitalized pictures of the Ishihara color plates, provided in the online supplementary material).[5] Upon entering all three numbers, participants were debriefed in written form and the experiment ended.

Data Analysis

Our analyses focus on the dependent variables reaction time (RT, measured in milliseconds from the onset of the target word) and error rates (calculated as the sums of wrong responses out of the totals of all trials). For RT analyses, individual correct median RTs were averaged across participants, as the median is less sensitive to disproportionately slow responses than the mean. However, across participants, we calculated the mean of these median RTs. For error analyses, mean error rates were arcsine square root transformed, to stabilize and normalize variances of proportional data (Sokal & Rohlf, 1981).

---

[4] The Austrian red-green data was taken from the "mixed blocks" from Kawai et al. (2020). The Austrian red-white data was taken from a study preregistered under osf.io/dfs9e, where participants saw a red-white and a red-black color block. Here, we only included data from the red-green mixed block and the red-white block from those participants that saw this block first.

[5] Note that this is a non-standardized assessment of the Ishihara test for color deficiency and may not give accurate results as for the presence or absence of color-blindness. However, since our experimental setup also made use of computer-display colors, it was important to us to verify if participants were able to discriminate shades of red and green on a computer monitor. So we opted to include this control measure, in addition to self-reported color deficiencies.

Trials with RTs below 150 ms and above 2 s were excluded from analysis. To demonstrate the magnitude of the observed effects, partial eta-squared (η_p^2) values, 90% confidence intervals (CI), and generalised eta-squared (η_G^2) values are reported for *F*-tests (Steiger, 2004). We report Bayes factors (as BF_{10} when supporting difference, and BF_{01} when supporting equivalence) using the default *r*-scale of .707 (Morey & Rouder, 2018). In case of analyses of variance (ANOVAs), we report inclusion BFs based on matched models (Makowski et al., 2019; Mathôt, 2017). Where applicable, we report Welch-corrected *t*-tests (Delacre et al., 2017) with corresponding Cohen's *d* values (Lakens, 2013). We used the conventional alpha level of .05 for all statistical significance tests. All analyses were conducted in R (R Core Team, 2019; via: Kelley, 2019; Lawrence, 2016; Lukács, 2020; Morey & Rouder, 2018).

Results

Valence Ratings

We compared the correctly classified mean ratings per valence group (negative = 1–5, positive = 6–10) between the Austrian and the Chinese sample. The mean ratings were very similar in magnitude regardless of sample (Figure 1) – nonetheless, we found some statistically significant differences. For positive stimuli, there was evidence for a difference in the mean ratings (raw mean difference: 0.22, 95% CI [0.05, 0.39]), $t(169.3) = 2.59$, $p = .010$, $d_{between} = 0.39$, 95% CI [0.09, 0.69], $BF_{10} = 3.60$, with slightly lower (less positive) ratings in China ($M_{Rating} \pm SD = 7.68 \pm 0.59$) than Austria (7.91±0.54).[6] For the negative stimuli, ratings differed significantly (raw mean difference: –0.54, 95% CI [–0.70, –0.37]), $t(173.2) = –6.55$,

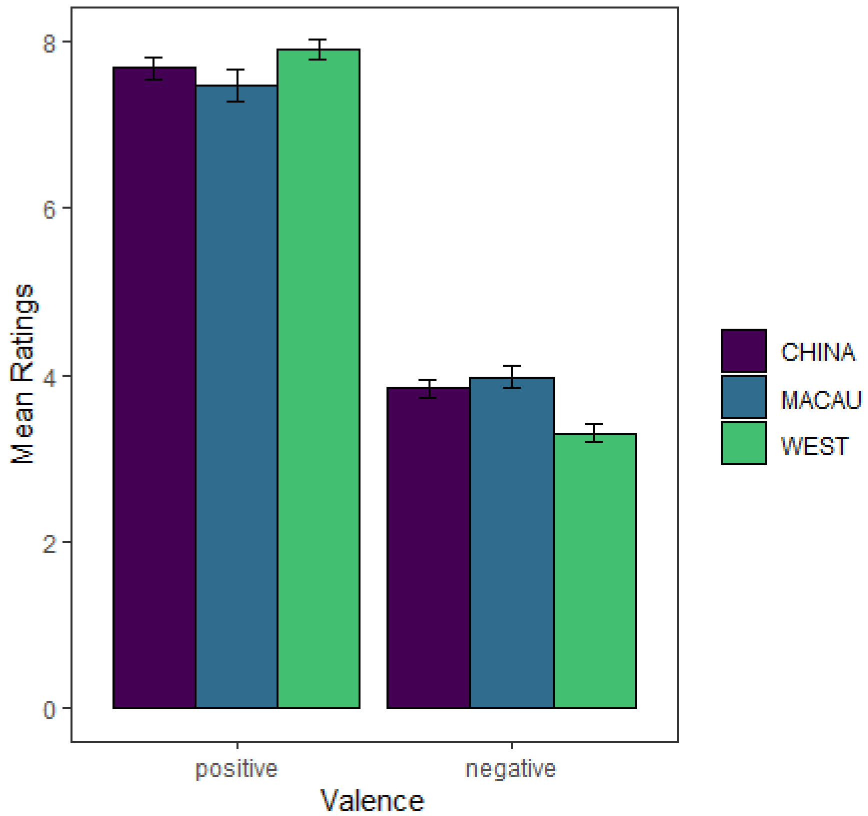
[6] Bonferroni-corrected alpha level for a set of three *t*-tests is .017.

$p < .001$, $d_{between} = –0.99$, 95% CI [–1.30, –0.67], $BF_{10} = 1.42 \times 10^7$, with lower (more negative) ratings by the Austrian (3.31±0.54) than the Chinese sample (3.84±0.54).

Interestingly, mean valence ratings between China and Macau did not differ significantly (positive valence: $M±SD$ = 7.47±0.61 [Macau] vs. 7.68±0.59 [China] and t[76.9] = –1.81, $p = .074$, $d_{between} = –0.35$, 95% CI [–0.72, 0.03], $BF_{01} = 1.11$; negative valence: $M±SD$ = 3.98±0.46 [Macau] vs. 3.84±0.54 [China] and $t(92.4) = 1.48$, $p = .143$, $d_{between}$ = 0.26, 95% CI [–0.11, 0.64], $BF_{01} = 2.09$). Unsurprisingly, mean valence ratings between Austria and Macau also showed significant differences (positive valence: raw mean difference of 0.43, 95% CI [0.21, 0.65]) and $t(68.9) = 3.87$, $p < .001$, $d_{between} = 0.77$, 95% CI [0.38, 1.14], $BF_{10} = 260.88$; negative valence: raw mean difference of –0.67, 95% CI [–0.85, –0.49] and $t(90.7) = –7.37$, $p < .001$, $d_{between} = –1.30$, 95% CI [–1.70, –0.90], $BF_{10} = 3.90 \times 10^7$). Figure 1 illustrates the mean valence ratings for the Austrian, Chinese, and Macanese group.

Figure 1

Mean Ratings for Positive and Negative Words in Experiment 1 by Participants from the Groups Austria, China, and Macau. Error Bars Indicate 95% CIs of Means.



Valence Categorization

Data was analyzed per (first) color block, once with the correct median categorization times (RTs), and once with the arcsine transformed error rates (ERs). Note that in the Macau group, more than half of the participants were of non-Macanese origin (mainly students from China), which – after exclusion – left this participant group distinctly smaller than the two other groups. We will here report the analyses of the complete data set (three-leveled Country factor including Austria, China, Macau)[7] because, in general, analyses showed very similar results (compared to a two-level factor comprising only Austria vs. China). The results of the two-country comparison (Austria vs. China) are available online in the supplementary material, together with two additional sets of preregistered supplementary statistical analyses: One set investigates potential Block Order effects in an analysis of all experimental blocks instead of only first-presented ones; another set constitutes a more conservative analysis of our data, where only color-repetition trials are included (color of the current word is the same as that of the previous word; all color-switch trials are excluded from the data sets). In summary, the analysis of all experimental blocks as well as the color-repetition trial analysis largely confirm the effects (especially Country-modulated Color-Valence interactions, which were a major focus) found in the main analysis, which we will present here.

Below, we first report the results from the analyses of the red-green color block and then from the red-white color block. Aggregated means and *SD*s for the RTs and ERs in the different factor combinations can be found in Table A1.

Red-Green Color Opposition

We analyzed RT and ER data from the 110 participants (44 from Austria, 43 from China, 23 from Macau). For both dependent variables, we ran repeated measures ANOVAs with Country as a three-level between-participants factor (Austria vs. China vs. Macau), and Color (red vs. green) and Valence (positive vs. negative) as within-participant factors.

---

[7] Originally, the analyses with a three-leveled factor Country were preregistered as supplementary. The main analysis was originally planned as focusing only on an Austria versus China comparison. Since we report the three-country comparison here, we report the two-country comparison as secondary analysis in the supplementary online material.

Reaction Times (RTs). We found a significant main effect for Country, $F(2, 107) = 24.01$, $p < .001$, $\eta^2_p = .310$, 90% CI [.186, .406], $\eta^2_G = .287$, $BF_{10} = 2.39 \times 10^6$, with faster responses in Macau (647.53±51.93 ms) and China (619.02±76.74 ms) than Austria (725.50±81.09 ms). The main effect for Valence was significant, $F(1, 107) = 69.37$, $p < .001$, $\eta^2_p = .393$, 90% CI [.274, .489], $\eta^2_G = .029$, $BF_{10} = 9.98 \times 10^{11}$, with faster responses to positive words (654.70±88.53 ms) than negative words (681.46±90.43 ms). Color did not influence RTs significantly, $F(1, 107) = 3.46$, $p = .066$, $\eta^2_p = .031$, 90% CI [0, .102], $\eta^2_G = .001$, $BF_{01} = 4.47$. The Color × Valence interaction was significant, $F(1, 107) = 105.43$, $p < .001$, $\eta^2_p = .496$, 90% CI [.384, .580], $\eta^2_G = .033$, $BF_{10} = 14.82 \times 10^{17}$. Importantly, this interaction was modulated by Country, expressed by a significant three-way interaction with $F(2, 107) = 6.60$, $p = .002$, $\eta^2_p = .110$, 90% CI [.026, .198], $\eta^2_G = .004$, $BF_{10} = 22.39$. Neither the Country × Color, nor the Country × Valence interactions reached significance (all Fs < 2, all ps > .10). Means of the median correct RTs for all three country groups are shown in Figure 2.

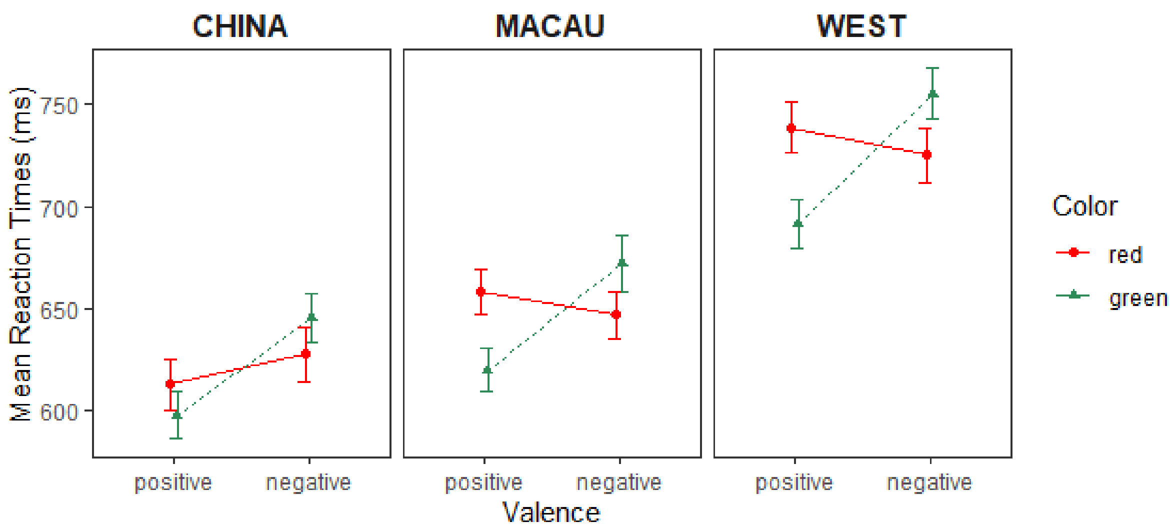
To follow up on the significant three-way interaction, we calculated congruence effect sizes for all country groups. In line with prior findings in Western culture (e.g., Kuhbandner & Pekrun, 2010), we assigned red-negative and green-positive trials as 'congruent pairings' and red-positive and green-negative trials as 'incongruent pairings'.[8] The size of the congruence effect for the response times (CE_{RT}) was then calculated as the sum of the differences between incongruent and congruent trials: RT[red_positive] – RT[red_negative] +

[8] We will use this measure as a quantitative means of comparison to determine in how far the Western congruence assignment is in accordance with the Chinese. If there is a red-positivity advantage in China (RT[red_positive] – RT[red_negative] < 0) or if both positive and negative words are primed by red equally in China (RT[red_positive] – RT[red_negative] = 0), the result would be a reduced CE for China compared to the West (given similar reactions to green between the cultures).

RT[green_negative] – RT[green_positive]. For the Austrian group, *M*±*SD* of the mean CE_{RT} size was 76.84±66.41 ms, for the Chinese group 32.42±49.44 ms, and for the Macanese group 63.23±54.66 ms. A series of Welch's *t*-tests showed that the mean CE_{RT} size in Austria was significantly larger than in China (t[79.4] = 3.54, p = .001, $d_{between}$ = 0.76, 95% CI [0.32, 1.19], BF_{10} = 42.51), but did not differ significantly from Macau (t[52.9] = 0.90, p = .374, $d_{between}$ = 0.22, 95% CI [–0.29, 0.72], BF_{01} = 2.83). With a Bonferroni-corrected level of α = .05/3 = .017, the CE size difference between Macau and China did not reach significance either (t[41.3] = 2.25, p = .030, $d_{between}$ = 0.60, 95% CI [0.08, 1.12], BF_{10} = 2.43).

Figure 2

Means of the Median Correct Reaction Times in Experiment 1 for Positive and Negative Words Presented in Red and Green Font Color per Country Group. Error Bars Indicate SEMs.



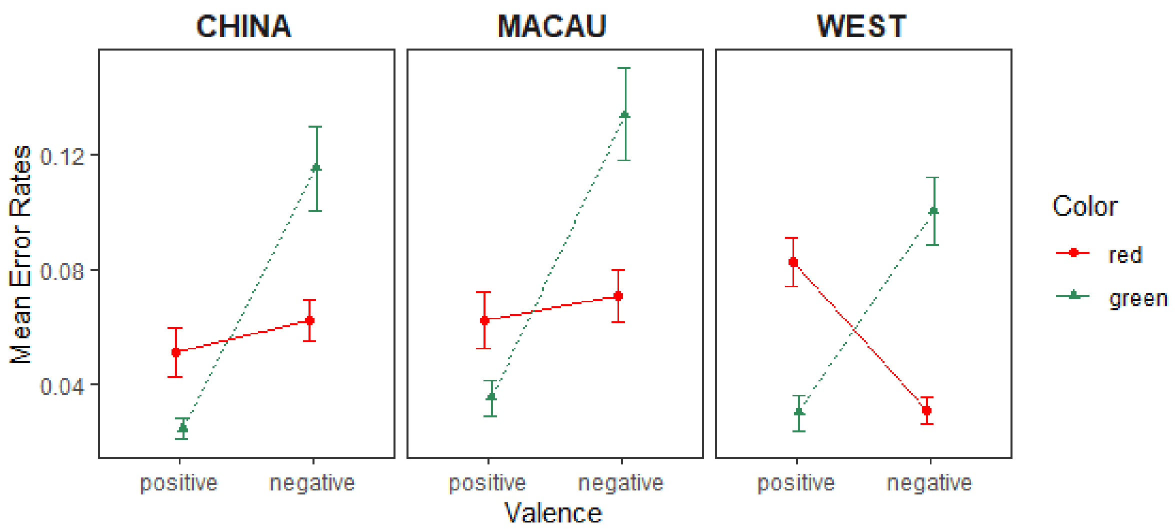
Error Rates (ERs). In the arcsine-transformed ERs, the main effect for Valence was significant, $F(1, 107) = 62.78$, $p < .001$, $\eta_p^2 = .370$, 90% CI [.251, .467], $\eta_G^2 = .100$, $BF_{10} = 6.99 \times 10^8$, with generally more errors for negative (.27±.10) than positive (.20±.09) targets. The interaction between Country and Valence was significant, $F(2, 107) = 12.73$, $p < .001$,

η_p^2 = .192, 90% CI [.084, .289], η_G^2 = . 043, $BF_{10} = 2.89 \times 10^4$, showing that in the Austrian group, the amount of errors was similar between both valence categories (raw mean difference: .02, 95% CI [–.02, .06]), while the Macanese and Chinese group showed more errors for negative than for positive words (raw mean difference in both groups: .11, 95% CI [.07, .15] for Macau, 95% CI [.07, .15] for China). The ERs also confirmed the significant Color × Valence interaction we found in the RTs, with $F(1, 107) = 151.75$, $p < .001$, η_p^2 = .586, 90% CI [.486, .657], η_G^2 = .176, $BF_{10} = 3.64 \times 10^{24}$. Again, this interaction was modulated by Country, resulting in a significant three-way interaction, $F(2, 107) = 6.34$, $p = .003$, η_p^2 = .106, 90% CI [.024, .193], η_G^2 = .018, BF_{10} = 18.35. No other effect was significant (all *F*s < 2.2, all *p*s > .10).

We again followed up on the significant three-way interaction by calculating the congruence effect sizes in the ERs (CE_{ER}; formula as above, but adapted for ERs) for all country groups. For the Austrian group, *M*±*SD* of the mean CE_{ER} size was .28±.20, for the Chinese group .15±.17, and for the Macanese group .18±.12. A series of Welch's *t*-tests showed that the mean CE_{ER} size in Austria was significantly larger than in China (t[83.4] = 3.22, $p = .002$, $d_{between} = 0.69$, 95% CI [0.25, 1.12], BF_{10} = 17.91) and in Macau (t[63.5] = 2.62, $p = .011$, $d_{between} = 0.58$, 95% CI [0.07, 1.09], BF_{10} = 2.16). The CE size difference between Macau and China did not reach significance (t[58.8] = 0.69, $p = .491$, $d_{between} = 0.16$, 95% CI [–0.35, 0.67], BF_{01} = 3.23). Mean ERs for all three country groups are shown in Figure 3.

Figure 3

Untransformed Mean Error Rates in Experiment 1 for Positive and Negative Words Presented in Red and Green Font Color per Country Group. Error Bars Indicate SEMs.



Red-White Color Opposition

We analyzed RT and ER data from the 107 participants (47 from Austria, 42 from China, 18 from Macau)[9]. For both dependent variables, we ran repeated measures ANOVAs, with Country as a three-level between-participants factor (Austria vs. China vs. Macau), and Color (red vs. white) and Valence (positive vs. negative) as within-participant factors.

Reaction Times (RTs). We again found a significant main effect for Country in the red-white block, $F(2, 104) = 10.86$, $p < .001$, $\eta^2_p = .173$, 90% CI [.067, .270], $\eta^2_G = .154$, $BF_{10} = 470.36$, with faster responses from Macau (649.06±59.29 ms) and China (615.14±71.86 ms) than Austria (684.64±70.91 ms). The main effect for Valence was also significant again, $F(1, 104) = 87.90$, $p < .001$, $\eta^2_p = .458$, 90% CI [.340, .547], $\eta^2_G = .060$, $BF_{10} = 3.28 \times 10^{24}$,

[9] In the preregistration, we said that we wanted to test at least 43 participants per country. However, due to exclusions, the number of participants in China was actually only 42, and Macau only 18. As mentioned above, the analyses with Country as a two-level factor (Austria vs. China) showed very similar effects (available in the online supplementary material).

with faster categorization times for positive (634.43±77.04 ms) than negative words (670.97±78.70 ms). Color showed a significant main effect, with $F(1, 104) = 4.65$, $p = .033$, $\eta_p^2 = .043$, 90% CI [.002, .121], $\eta_G^2 = .001$, $BF_{01} = 2.82$, with responses to red being slightly slower than to white stimuli (656.15±75.28 ms vs. 651.29±76.89 ms, respectively). The Color × Valence interaction was significant as well, $F(1, 104) = 49.92$, $p < .001$, $\eta_p^2 = .324$, 90% CI [.204, .427], $\eta_G^2 = .012$, $BF_{10} = 1.26 \times 10^5$. However, in this case, the interaction was *not* modulated by Country, $F(2, 104) = 0.35$, $p = .708$, $\eta_p^2 = .007$, 90% CI [0, .038], $\eta_G^2 < .001$, $BF_{01} = 9.05$. For all other effects *F*s < 2, all *p*s > .20. Means of the median correct RTs for all country groups are illustrated in Figure 4.

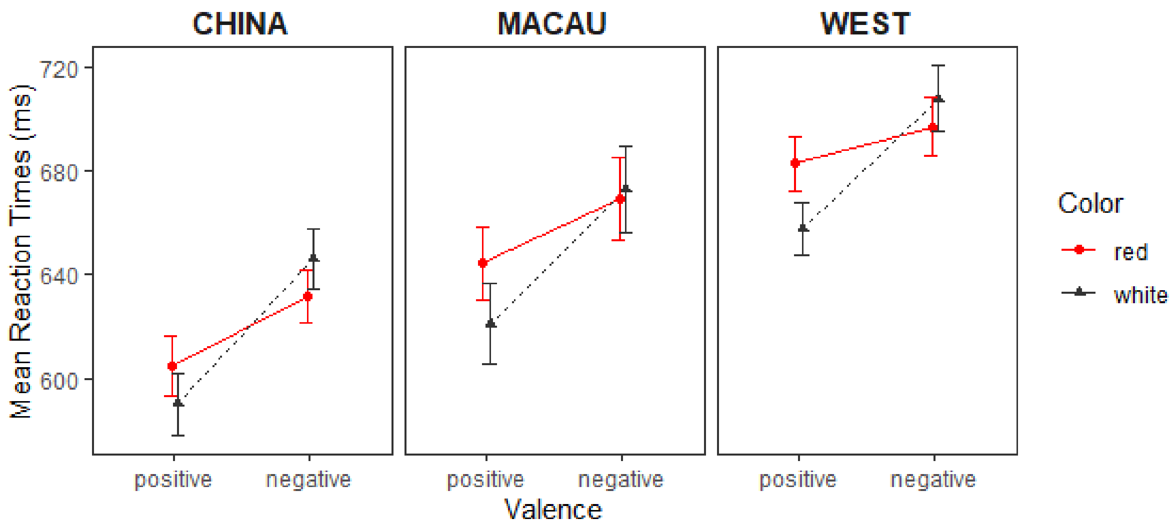
Although not motivated by a significant three-way interaction, for an easier direct comparison of the results from red-green and red-white color blocks, we again calculated the CE_{RT} in the red-white block by coding the red-negative and white-positive combinations as congruent[10], separately for different countries. These effect sizes also entail information about the nature of the significant Color × Valence interaction. The CE_{RT} had a size of 36.05±53.67 ms in the Austrian group, 29.18±39.77 ms in the Chinese, and 27.26±41.44 ms in the Macanese group (since Country was not a modulating factor in the color-valence interaction, the size of the countries' CE does not differ greatly, all Welch's *t*-tests showed *p* > .45). Nominally, the response-time related CEs in the red-white block were smaller than in the red-green block for Austria and Macau (red-green block: 76.84 ms in Austria, 63.23 ms in Macau), but closer in size across different color-opposition blocks for China (red-green block 32.42 ms). For an illustration of the respective mean RT congruence effect sizes per color block and country, see Figure 6 (left panel).

---

[10] The formula for calculating the congruence effect size is similar to above; we only substituted white for green.

Figure 4

Means of the Median Correct Reaction Times in Experimen1 1 for Positive and Negative Words Presented in Red and White Font Color per Country Group. Error Bars Indicate SEMs.



Error Rates (ERs). In the ERs, the main effect for Valence was significant, $F(1, 104) = 44.44$, $p < .001$, $\eta_p^2 = .299$, 90% CI [.181, .403], $\eta_G^2 = .092$, $BF_{10} = 5.08 \times 10^8$, with more errors made to negative than to positive words (.26±.10 vs. .20±.08, respectively). As was the case in the red-green ERs, Country interacted significantly with Valence, $F(2, 104) = 8.49$, $p < .001$, $\eta_p^2 = .140$, 90% CI [.044, .234], $\eta_G^2 = .037$, $BF_{10} = 1240.75$, showing again that, in the Austrian group, the amount of errors was similar between both valence categories (raw mean difference: .02, 95% CI [–.02, .05]), while the Macanese and Chinese groups showed more errors for negative than for positive words (raw mean difference in both groups: .10, 95% CI [.04, .16] for Macau, 95% CI [.06, .14] for China). Similar to the RTs, the Color × Valence interaction was significant, $F(1, 104) = 56.24$, $p < .001$, $\eta_p^2 = .351$, 90% CI [.230, .451], $\eta_G^2 = .081$, $BF_{10} = 1.88 \times 10^9$. Again, this interaction was *not* modulated by Country, *F*(2,

104) = 0.44, p = .648, η^2_p = .008, 90% CI [0, .043], η^2_G = .001, BF_{01} = 7.26. For all other effects Fs < 3.2, all ps > .07. The ERs are illustrated in Figure 5.

We used the same approach to calculate the CE_{ER} in the red-white block as above, separately by Country. The CE_{ER} had a size of .15±.19 in the Austrian group, .12±0.18 in the Chinese, and .12±0.14 in the Macanese group (all Welch's t-tests showed p > .40). Thus, the error-rate related CEs in the red-white block were nominally smaller than in the red-green block – only now for all three countries (red-green block: .28 in Austria, .15 in China, .18 in Macau). For an illustration of the respective mean ER congruence effect sizes per color block and country, see Figure 6 (right panel).

Figure 5

Untransformed Mean Error Rates in Experiment 1 for Positive and Negative Words Presented in Red and White Font Color per Country Group. Error Bars Indicate SEMs.

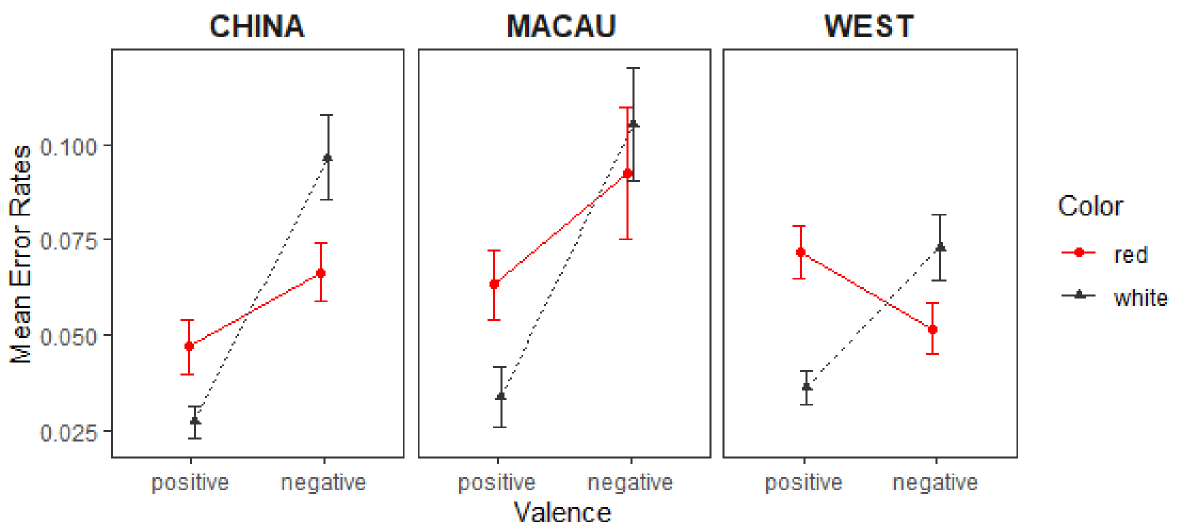
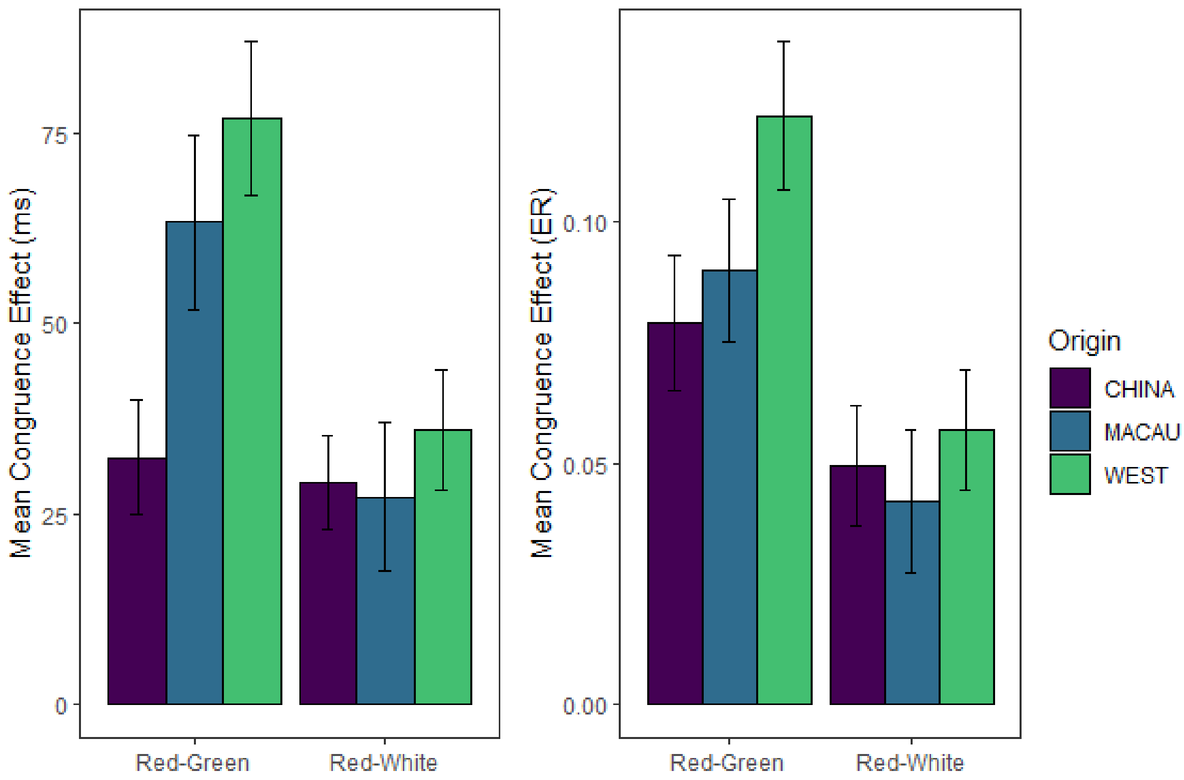


Figure 6

Mean Congruence Effects (Performance in Incongruent Conditions Minus Performance in Congruent Conditions) in Experiment 1 for Reaction Times (ms; Left Panel) and for Untransformed Mean Error Rates (ER; Right Panel) per Color Opposition Block and Country. Error Bars Indicate SEMs.



Discussion

Our results provided some evidence for a cultural origin of color-valence associations, as they show differences in red-valence associations between cultures. In Austria, in red-green color conditions, implicit associations between the color red and the emotional concept of negativity are *stronger* than in Mainland China. The results for Macau lie in-between; in the RTs the Macanese congruence effect in red-green blocks was comparable to Austria, in the ERs not reliably so and, thus, comparable to China. The West-versus-China difference dissipates, however, when the opposing colors are red and white, instead of red and green.

This was not expected on the basis of different cultural origins or the same cultural origin of each of our tested color-valence associations.

One partial explanation for the culture dependencies is that red is perceived as somewhat less negative in China than in Western cultures, due to the fact that red is generally *used* in more positive and less negative contexts, culturally. However, clearly, this influence would also additionally depend on the type of color system employed, as the higher positivity of the color red's valence association in the Chinese than in the Western sample only showed in the condition with red opposed to green but *not* in the condition with red opposed to white. In the latter condition, even in the Western sample, red's association to positivity showed at least in the RTs.[11] A possible explanation is that the red-green opposition fostered the dominance of the red-negativity association among Western participants only, potentially by some type of idiosyncratic connection, such as the German sayings with positive meanings incorporating the color green (e.g., "Grün ist die [Farbe der] Hoffnung", *green is the [color of] hope*; or "Alles im grünen Bereich", *everything is alright, lit. 'in the green range'*) and, in turn, boosting the negative valence of the color red as its contextual opposition. This specific finding would be in line with the color-in-context theory in general (Elliot & Maier, 2012) and with the influence of color systems in particular.

In any case, the results from this implicit measure show a divergence from those achieved with explicit measures in past research: Where explicit research predicted similar results cross-culturally (namely between associations for red and green), our implicit measures showed differences; and where explicit measures could have conceivably predicted

[11] In the Western sample, the RT differences between red and green trials ($M \pm SD$ = 46.99±44.43) was significantly larger than the difference between red and white trials ($M \pm SD$ = 25.15±34.60), raw mean difference: –21.84, 95% CI [–38.52, –5.15]); $t(81.2) = -2.60$, $p = .011$, $d = -0.55$, 95% CI [–0.97, –0.13], $BF_{10} = 4.30$.

differences (between associations for white), our implicit measures showed no difference (cf. Jonauskaite, Abu Akel, et al., 2020). Critically, this difference can most likely be attributed to the particular color systems or color oppositions employed in the present study. Their influence was reflected in the fact that culture-dependencies were present for one and the same color – red – only if it was opposed to green, but not if it was opposed to white.[12] In the present study, this influence of a specific opposing color on the red-valence responses was a necessary consequence of the particular type of implicit measure that we took – a valence-color congruence effect. However, this influence would have been more heterogeneous or even absent in the case of explicit rating studies of colors, in which a number of different alternative colors were used in other trials and in which it is, therefore, not clear which of these alternative colors might have provided a background against which a participant would have judged each actual color (if participants used any such mental color comparisons at all).

One potential confound in the results of Experiment 1 and in particular in the found cultural differences could be the use of language-specific material. This is a complication, as words are not strictly the same stimuli in Chinese and German. In fact, our data shows that Chinese speakers – Mandarin and Cantonese alike – categorized word valence faster than speakers of German (here: Austrians) did. Importantly, general differences of word processing between languages might play a role in our observed group differences. Both Asian groups use the Chinese character writing system.[13] Several studies suggest an

---

[12] Looking only at the red trials, the difference between the (red) negative and the (red) positive response latencies was significantly different between West and China *only* in the red-green opposition condition ($M \pm SD$ in ms = –13.40±48.96 vs. 14.86±40.63, respectively; $t(82.8) = -2.93$, $p = .004$, $d = -0.63$, 95% CI [–1.06, –0.19], $BF_{10} = 8.62$), but not in the red-white opposition condition ($M \pm SD$ in ms = 14.14±49.56 vs. 26.56±41.69, respectively; $t(86.7) = -1.28$, $p = .203$, $d = -0.27$, 95% CI [–0.69, 0.15], $BF_{01} = 2.22$.)

[13] The mean length of all used Austrian words was 6.83 letters or 2.18 syllables. The mean length of the Mandarin words was 2.07 characters or 18.06 strokes.

advantage of the time it takes to access semantic information of a word in Chinese over languages using the Latin alphabet, with evidence from reading times (Lü & Zhang, 1999) and semantic preview benefits in eye-tracking studies, due to the fact that "the Chinese writing system is based on a closer association between graphic form and meaning than is alphabetic script" (Yan et al., 2009, p. 561).[14] Of course, the Country main effect might also reflect a general, stimulus-independent processing advantage of Chinese participants over Austrians. We suspected, however, that it is more likely a linguistic artefact. Importantly, if meaning (i.e., semantic information about the affective valence of a word/concept) is extracted very quickly when reading Chinese characters, the color information carried *with* the linguistic cue might not be as effective in facilitating (in case of congruent color-valence pairings) or inhibiting (in case of incongruent pairings) lexical access, semantic retrieval and response execution. Interestingly, when looking at the results of the Macanese group, in particular in the red-green context, we found a great similarity to Western color-valence association patterns, but less overlap with the Mainland Chinese results. One might argue that this speaks against a merely word-processing based interpretation of the found cultural differences. However, to confirm differences in implicit cross-modal associations between cultures *independently* of word-processing differences between these cultures, we used color-manipulated pictorial images instead of words in Experiment 2.

[14] Chinese words are composed of one or more characters (logograms); and each character, in turn, is composed of one or more constituents that each carry meaning. This results in the sublexical units of the Chinese writing system having "more direct contact with meanings" (Perfetti et al., 2002, p. 36).

Experiment 2

Like words, pictures can reliably signify emotional content. Complex photorealistic pictures as well as black-and-white outline drawings or silhouettes have been shown to elicit emotion responses (e.g., Bradley & Lang, 2007; Giner-Sorolla et al., 1999; Schimmack, 2005). The effect of color, however, was rarely isolated in such studies (for brightness-effects, see Lakens et al., 2013). We do not know of any prior study that investigated implicit color-valence associations carried out by silhouette pictures between cultures. In an online study, we tested native Mandarin-speakers (from China) and native German-speakers (from Austria and Germany) in two color-opposition blocks: red-green and red-white.

Hypotheses were similar to Experiment 1. If color-valence associations are of a cultural origin, we should find differences between red-valence associations between the two cultures tested, at least in the red-green color conditions. Based on Experiment 1, we did not expect much cultural differences in the red-white color conditions. In addition, if cultural differences in color-valence associations in Experiment 1 were due to word-processing differences between the Chinese and the German language, no culture-dependent differences in color-valence congruence effects were to be expected in Experiment 2, with its pictorial stimuli.

Method

In Experiment 1, the data from Macau was useful in confirming a general main effect for Chinese (Mandarin/Cantonese) speakers over German speakers. However, to determine the influence of the participants’ language on the previous results, one Chinese sample was sufficient for Experiment 2.

Participants

Data from a total of 251 participants was collected online. For the Chinese-speaking group, 124 participants were recruited through an advertisement that was posted via WeChat to the open group of the psychology laboratory of Shaanxi Normal University, Xi'an, China. Chinese participants were rewarded with 20 CNY for valid participation.[15] For the German-speaking group, 46 participants from the University of Vienna were recruited in return for (partial) course credit.[16] An additional 81 participants (students with Austrian or German nationality) were recruited via Prolific (www.prolific.co) and paid 2.20 GBP for valid participation, resulting in a total of 127 German-speakers. Participants were randomly assigned to one of the conditions that resulted from permuting block order and key location.

The same exclusion criteria applied as in Experiment 1.[17] From the collected data (n_{GER} = 127, n_{CH} = 124), we excluded participants with an accuracy rate lower than 75% (n_{GER} = 0, n_{CH} = 2), a failed color discrimination test (n_{GER} = 9, n_{CH} = 9), or a reported birth place other than Germany or Austria (n_{GER} = 7) for the German speakers or Mainland China (n_{CH} = 1) for the Mandarin speakers. This left data from a total of 223 participants, 112

[15] The amount of monetary reimbursement we preregistered per participant was lower (10 CNY). We decided to increase the amount to give sufficient incentive for Chinese participants because they saw the advertisement on their mobile device but were not allowed to run the study on it. Instead, they had to use a PC to participate in the study.

[16] We preregistered to open 120 places for online participation for students from the University of Vienna only. After a span of one month, no more students volunteered to take part in the study, so we completed participant collection via Prolific. The results from Austrian students and from German-speaking Prolific participants are highly similar and, in fact, give the same key results in comparison with the Chinese results even when tested separately.

[17] Since the nationality of the German-speaking sample in this experiment was predominantly German (as opposed to Experiment 1), we will henceforth refer to this group as "German" or "German speakers".

subjects (age = 19.8±4.4; 30 male) in the Chinese group and 111 subjects (age = 24.1±4.3; 57 male) in the German group.

Material

Silhouette Selection: Eighty positive and 80 negative silhouettes (300 × 300 px) were taken from the Bicolor Affective Silhouettes & Shapes (BASS) database (https://gasparl.github.io/BASS/; Kawai et al., 2021). The BASS is well suited for comparing a Western culture and China, since it contains representative valence and arousal ratings from both cultural groups, with the Western (in the BASS data base, US) ratings presumed to be comparable with Austrian ratings.[18] We carefully controlled for culturally comparable valence and arousal ratings from the West/US and East/China. This means specifically, that the Western/US valence and Chinese valence ratings were similar on average among the positive silhouettes as well as among the negative silhouettes. At the same time, there was enough within-category heterogeneity between different silhouettes both in the positive and negative categories, both on the side of the Western/US ratings and the Chinese ratings. Regarding arousal levels from the Western/US sample, they were similar for positive and negative silhouettes, but this was impossible to accomplish for Chinese arousal ratings, due to the stronger linear valence-arousal-relationship among Chinese participants – with positive silhouettes being rated as more arousing than negative ones. However, we reduced the difference as much as possible. Lastly, the number of black and white pixels was comparable between positive and negative silhouettes, which means that the amount of color present (red, green, or white, depending on condition, see next paragraph) was similar across both valence

[18] Taking the mean ratings from the Austrian BASS-norming pilot study (n = 78), the correlation between the Austrian and the US mean valence ratings for the 583 pictures was very high with $r(581) = .928$, 95% CI [.915, .938]. The correlation between the groups' mean arousal ratings was somewhat lower than for valence but still high, with $r(581) = .807$, 95% CI [.777, .834].

conditions. A compilation of these mean values is available in Table A2 of the Appendix. The full list of the 160 silhouettes we used is available in the online supplementary material.

Color Manipulation: Like Experiment 1's design for Chinese and Macanese participants, Experiment 2 consisted of two different blocks with color-manipulated silhouettes: a red-green block and a red-white block. Black pixels of the original black-on-white silhouettes were replaced with red, green, and white color. Shades of red and green were taken from Wilms and Oberfeld (2018), with similarly high saturation/lightness: red (L*a*b* [50, 81.29, 82.05] = RGB [245.63, 0, 0]) and green (L*a*b* [50.03, -98.59, 55.35] = RGB [0, 148.83, 0]). White pixels (i.e., the background) of the original silhouettes were replaced with a mid-gray (L*a*b* [50, 0, 0] = RGB [119, 119, 119]). Note that, since the experiment was run online, we were not able to control the monitor settings of the participants' setup. However, at the beginning of the experiment, we asked participants to set their monitor brightness to the highest level, and we included a short color discrimination test as well (see Procedure) to make sure that the colors we manipulated were discernible.

Procedure

The experiment could only be completed using a Google Chrome Browser.

Color Vision Test: Before the experimental task, we showed three pictures of the Ishihara number plates (500 × 500 px) as a first screen-out test. In addition, four rectangles (200 × 100 px) colored in the shades of the red and green we used for the silhouettes, as well as one brown tone ("Saddle brown", RGB[139,69,19]) and one olive-green tone ("Olive Drab", RGB[107,142,35]), were presented and participants were asked to select the color they saw for each rectangle. Only if participants entered all three numbers from the Ishihara plates and the colors of the rectangles correctly, they could proceed with the experiment.

Valence Categorization Task: Participants were then presented with the informed consent and, except for the study on Prolific, were asked to provide demographic data (age, gender, place of origin). Thereafter, they saw the instructions (in German for German speakers, in Simplified Chinese for Mandarin speakers) that informed them about the upcoming task (i.e., that they will see a series of silhouette pictures in different colorations, and need to press key "E" for positive images, and key "I" for negative images, or vice versa). To make the valence category clear to the participants, all 80 positive and 80 negative silhouettes were shown (in black-on-white) before the task started. After successful completion of a practice round, the two experimental blocks, the red-green and red-white block, were presented (in counterbalanced order).

Per block, each silhouette was shown twice – once in each color. This resulted in a total of 80 (positive) + 80 (negative) = 160 * 2 (in Color 1 + in Color 2) = 360 * 2 (Block 1 + Block 2) = 720 experimental trials.

In both experimental blocks, silhouettes were presented in the center of the screen against a gray background (RGB[112,112,112]). When no answer was logged within 2 s after stimulus onset, the message "Too slow!" (in the participant's language) was shown for 500 ms and the stimulus disappeared. If the participant gave an incorrect response within the response window, the message "Incorrect!" (in the participant's language) was shown for 500 ms and the stimulus stayed on screen until the correct response was given. After the correct response was logged, the next trial started (i.e., the next silhouette was displayed). Between the two experimental blocks, participants could take a break of self-determined length and were informed of the altered color of the upcoming stimuli.

Data Analysis

For all analyses, only the first response to each stimulus presented was used, and all practice trials were excluded. Responses below 150 ms and above 2 s were discarded. For RT analysis, only correct responses were used. For ER analysis, ERs were calculated as incorrect responses per correct plus incorrect responses.

For each image, we checked ERs per participant group. One silhouette (*falling.png*) was classified incorrectly over 40% of the times in the German subject group (and over 35% in the Chinese subject group). Error rates for all other images were below 30% in both culture groups. Consequently, we excluded all responses to “falling.png” trials from analysis.[19]

Results

Data from both culture groups was collected in two conditions that differed in the presentation order of the experimental blocks (red-green block first vs. red-white block first). Since analyses showed that Block Order was not a determining factor in the three-way interactions (Color × Valence × Country) we were most interested in (see supplementary analysis in online material), we collapsed data over color blocks, disregarding whether they were presented as the first or second experimental block.

For both dependent variables (mean correct response times [RTs] and mean error rates [ERs]), we ran repeated measures ANOVAs for each color block, with Country (West vs. China) as between-participants factor, and Valence (positive vs. negative) and Color (red vs. 2nd color) as within-participant factors. Again, we first report the results from the analyses of

[19] The exclusion of this stimulus did not affect the analysis results to a meaningful extent.

the red-green color block and then from the red-white color block. Means and *SD*s for RTs and ERs can be found in Table A1, respectively.

Red-Green Color Opposition

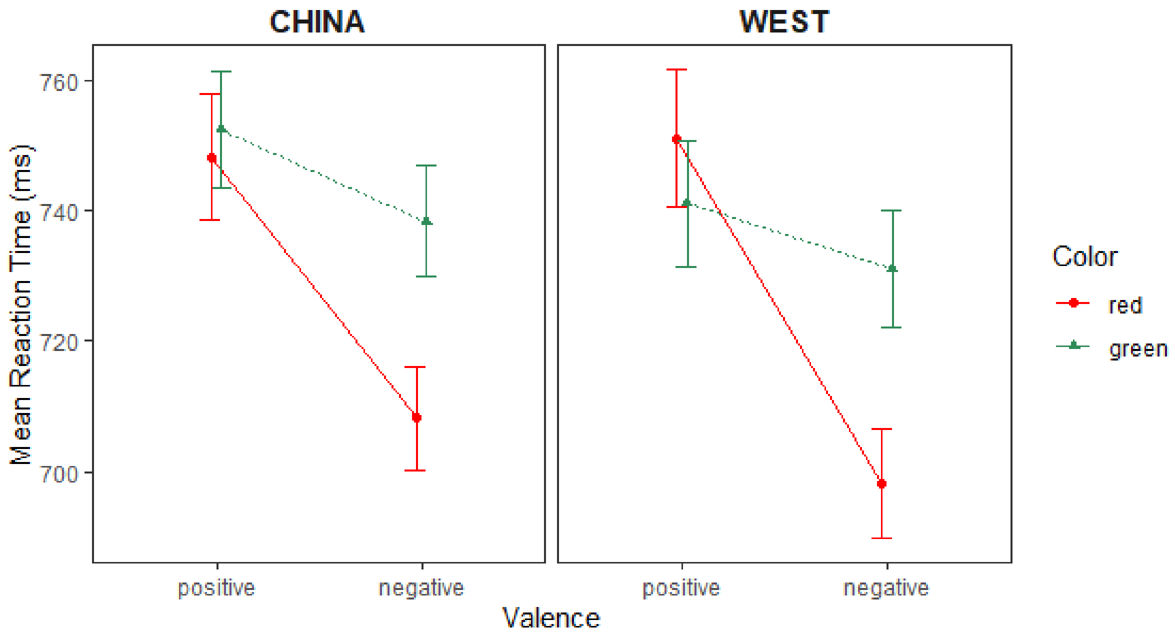
Reaction Times. We found significant main effects for the factors Valence and Color, with $F(1, 221) = 110.66$, $p < .001$, $\eta^2_p = .334$, 90% CI [.252, .406], $\eta^2_G = .023$, $BF_{10} = 3.72 \times 10^{30}$, and $F(1, 221) = 64.68$, $p < .001$, $\eta^2_p = .226$, 90% CI [.150, .301], $\eta^2_G = .006$, $BF_{10} = 1.64 \times 10^7$, respectively, showing faster responses to negative (719.02±87.58 ms) than to positive silhouettes (748.23±100.32 ms) and faster responses to red (726.42±93.93 ms) than to green silhouettes (740.83±91.72 ms). Interestingly, the main effect for Country was not significant in the silhouette categorization (cf. Experiment 1), $F(1, 221) = 0.27$, $p = .605$, $\eta^2_p = .001$, 90% CI [0, .020], ηG2 = .001, $BF_{01} = 2.29$.

Most importantly, the Color × Valence interaction was significant, $F(1, 221) = 94.08$, $p < .001$, $\eta^2_p = .299$, 90% CI [.218, .372], $\eta^2_G = .008$, $BF_{10} = 6.24 \times 10^{11}$. This interaction was modulated by Country, resulting in a significant three-way interaction, with $F(1, 221) = 5.79$, $p = .017$, $\eta^2_p = .026$, 90% CI [.002, .069], $\eta^2_G < .001$, $BF_{10} = 1.04$. All other interactions were not significant (all *F*s < 2.60, *p*s > .10). Figure 7 illustrates the RT data.

To follow up on the significant three-way interaction, we again calculated congruence effect sizes for both country groups (see Valence Categorization Results in Experiment 1). For the German group, *M*±*SD* of the mean CE_{RT} size was 43.08±54.84 ms, for the Chinese group 25.97±51.31 ms (raw mean difference: 17.11 ms, 95% CI [0.43, 27.52]). Welch's *t*-test showed that the distributions between the two groups differed significantly, $t(219.7) = 2.40$, $p = .017$, $d_{between} = 0.32$, 95% CI [0.06, 0.59], $BF_{10} = 2.18$.

Figure 7

Mean Reaction Times for Positive and Negative Silhouettes in Experiment 2 Presented in Red and Green Color per Country Group ("WEST" Denotes the German-Speaking Group). Error Bars Indicate SEMs.

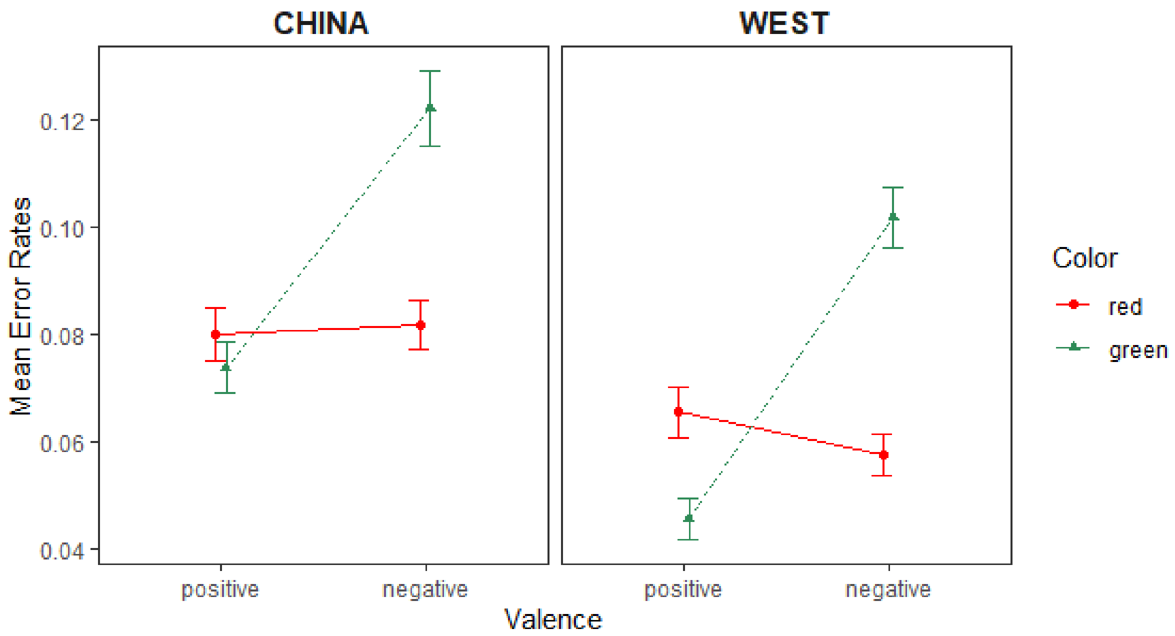


Error Rates. The main effects for all three factors were significant, for Valence $F(1, 221) = 63.29$, $p < .001$, $\eta_p^2 = .223$, 90% CI [.147, .297], $\eta_G^2 = .051$, $BF_{10} = 2.04 \times 10^{14}$, for Color $F(1, 221) = 52.60$, $p < .001$, $\eta_p^2 = .192$, 90% CI [.120, .266], $\eta_G^2 = .019$, $BF_{10} = 3.06 \times 10^4$, and for Country $F(1, 221) = 15.86$, $p < .001$, $\eta_p^2 = .067$, 90% CI [.023, .125], $\eta_G^2 = .041$, $BF_{10} = 204.85$. Negative silhouettes elicited higher ERs than green silhouettes. The Chinese group made more misclassifications than the German group. The Color × Valence interaction was significant, with $F(1, 221) = 112.10$, $p < .001$, $\eta_p^2 = .337$, 90% CI [.255, .409], $\eta_G^2 = .064$, $BF_{10} = 1.29 \times 10^{21}$, but not modulated to a significant degree by the factor Country, $F(1, 221) = 2.90$, $p = .090$, $\eta_p^2 = .013$, 90% CI [0, .048], $\eta_G^2 = .002$, $BF_{01} = 1.55$.

Although not motivated by a significant three-way interaction, for facilitated direct comparison with the RT results, effect sizes of the Color × Valence interaction were calculated separately for color and for country. CE_{ER} sizes for the interaction were .06±.08 for the German group and .05±.08 for the Chinese group (i.e., no difference between the countries, $t(220.2) = 1.70$, $p = .090$, $d_{between} = 0.23$, 95% CI [–0.04, 0.49], $BF_{01} = 1.76$). All other interactions were non-significant as well (all Fs < 1.50, ps > .20). The mean ERs are illustrated in Figure 8.

Figure 8

Mean Error Rates in Experiment 2 for Positive and Negative Silhouettes Presented in Red and Green Color per Country Group. Error Bars Indicate SEMs.



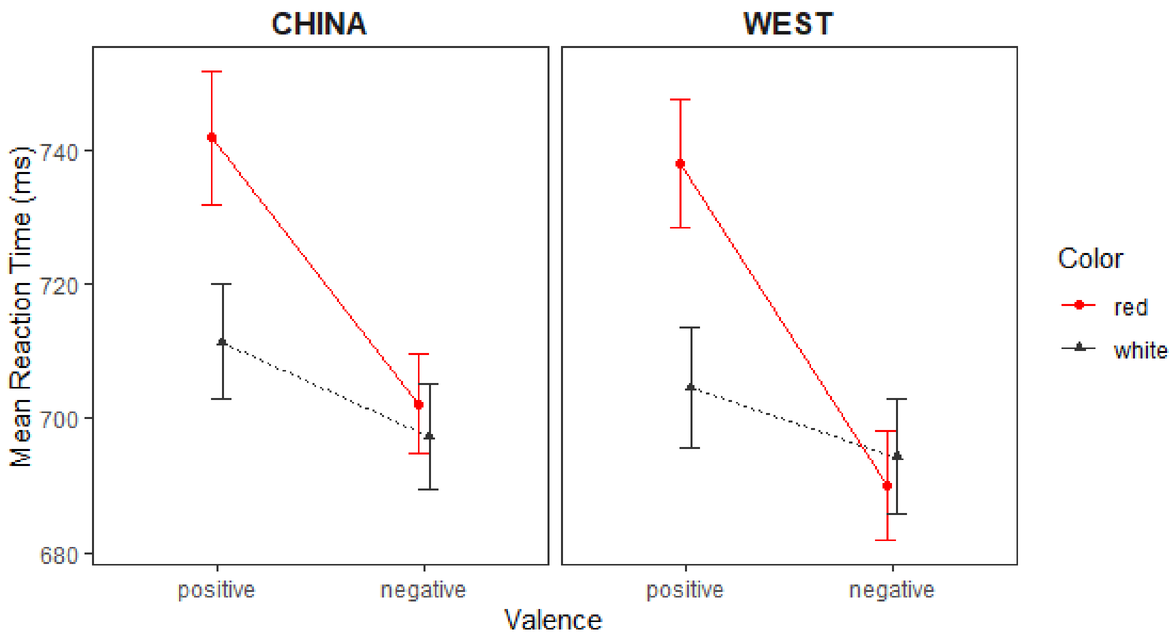
Red-White Color Opposition

Reaction Times. Main effects for Valence and Color were significant, with $F(1, 221) = 93.53$, $p < .001$, $\eta^2_p = .297$, 90% CI [.217, .371], $\eta^2_G = .023$, $BF_{10} = 6.22 \times 10^{26}$, and $F(1, 221) = 75.31$, $p < .001$, $\eta^2_p = .254$, 90% CI [.176, .329], $\eta^2_G = .008$, $BF_{10} = 4.41 \times 10^8$, respectively. While the Valence effect was in the same direction as previously, in this color block, responses to red (717.97±89.15 ms) were slower than to white silhouettes (701.90±86.26 ms). Just as in the red-green block RTs, the main effect for Country was not significant, $F(1, 221) = 0.31$, $p = .579$, $\eta^2_p = .001$, 90% CI [0, .021], $\eta^2_G = .001$, $BF_{01} = 2.61$. The interaction between Color and Valence was significant, $F(1, 221) = 71.88$, $p < .001$, $\eta^2_p = .245$, 90% CI [.167, .320], $\eta^2_G = .007$, $BF_{10} = 1.07 \times 10^9$, but – as opposed to the RTs in the red-green block – not significantly modulated by Country, $F(1, 221) = 2.74$, $p = .099$, $\eta^2_p = .012$, 90% CI [0, .047], $\eta^2_G < .001$, $BF_{01} = 2.17$. All other interactions were non-significant as well (all *F*s < 0.75, *p*s > .25). The mean RTs for the red-white block are illustrated in Figure 9.

CE_{RT} sizes in the red-white block were 37.83±50.55 ms for the German and 25.97±51.31 ms for the Chinese group. Welch's *t*-test showed no evidence for a significant difference, $t(221.0) = 1.74$, $p = .083$, $d_{between} = 0.23$, 95% CI [–0.03, 0.50], $BF_{01} = 1.66$. For an illustration of the respective mean RT congruence effect sizes per color block and country, see Figure 11 (left panel).

Figure 9

Mean Reaction Times in Experiment 2 for Positive and Negative Silhouettes Presented in Red and White Color per Country (West is Germany and Austria collapsed) Group. Error Bars Indicate SEMs.



Error Rates. Main effects for Valence and Country were significant, with $F(1, 221) = 54.53$, $p < .001$, $\eta^2_p = .198$, 90% CI [.125, .272], $\eta^2_G = .051$, $BF_{10} = 4.23 \times 10^{14}$, and $F(1, 221) = 18.16$, $p < .001$, $\eta^2_p = .076$, 90% CI [.029, .136], $\eta^2_G = .046$, $BF_{10} = 390.59$, respectively, both showing the same patterns as in the red-green block ERs. The Color × Valence interaction was significant, with $F(1, 221) = 97.54$, $p < .001$, $\eta^2_p = .306$, 90% CI [.225, .380], $\eta^2_G = .049$, $BF_{10} = 2.56 \times 10^{15}$, but the three-way interaction was not, $F(1, 221) = 0.22$, $p = .643$, $\eta^2_p = .001$, 90% CI [0, .019], $\eta^2_G < .001$, $BF_{01} = 6.08$. All other effects and interactions were non-significant as well (all *F*s < 0.50, *p*s > .45). The mean ERs are illustrated in Figure 10. CE_{ER} sizes were virtually the same for both the German and Chinese group, with .04±.07 and .05±.08, respectively. Welch's *t*-test showed no evidence for a difference in distributions between the two groups, $t(217.1) = –0.35$, $p = .728$, $d_{between} = –0.05$, 95% CI [–0.31, 0.22], $BF_{01} = 6.46$. For an illustration of the respective mean ER congruence effect sizes per color block and country, see Figure 11 (right panel).

Figure 10

Mean Error Rates in Experiment 2 for Positive and Negative Silhouettes Presented in Red and Green Color per Country Group. Error Bars Indicate SEMs.

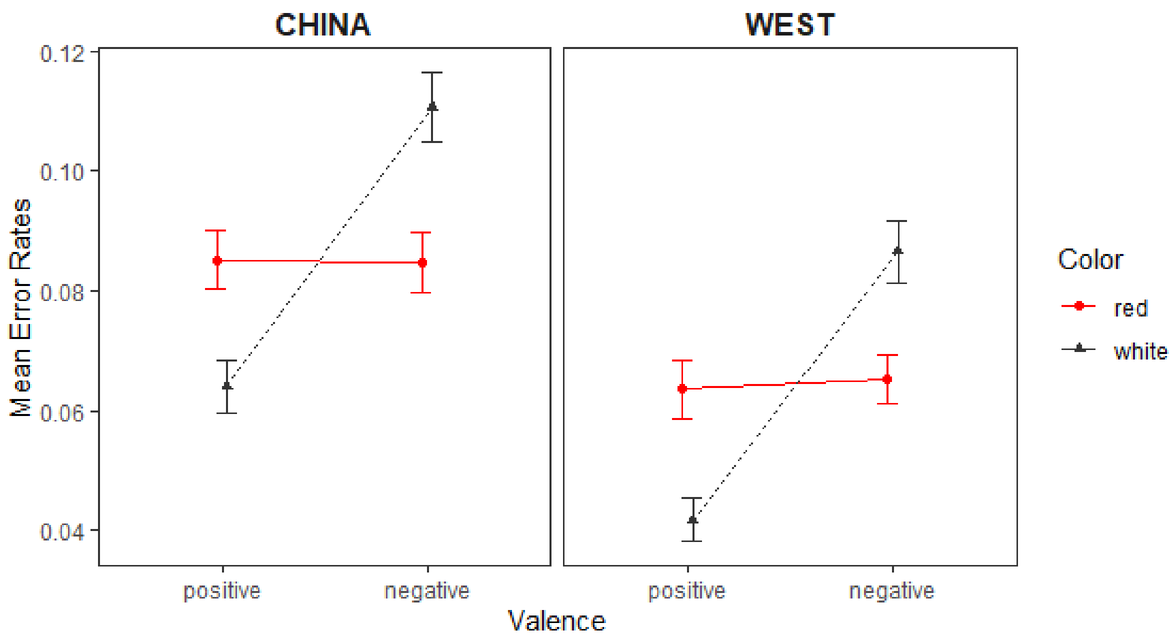
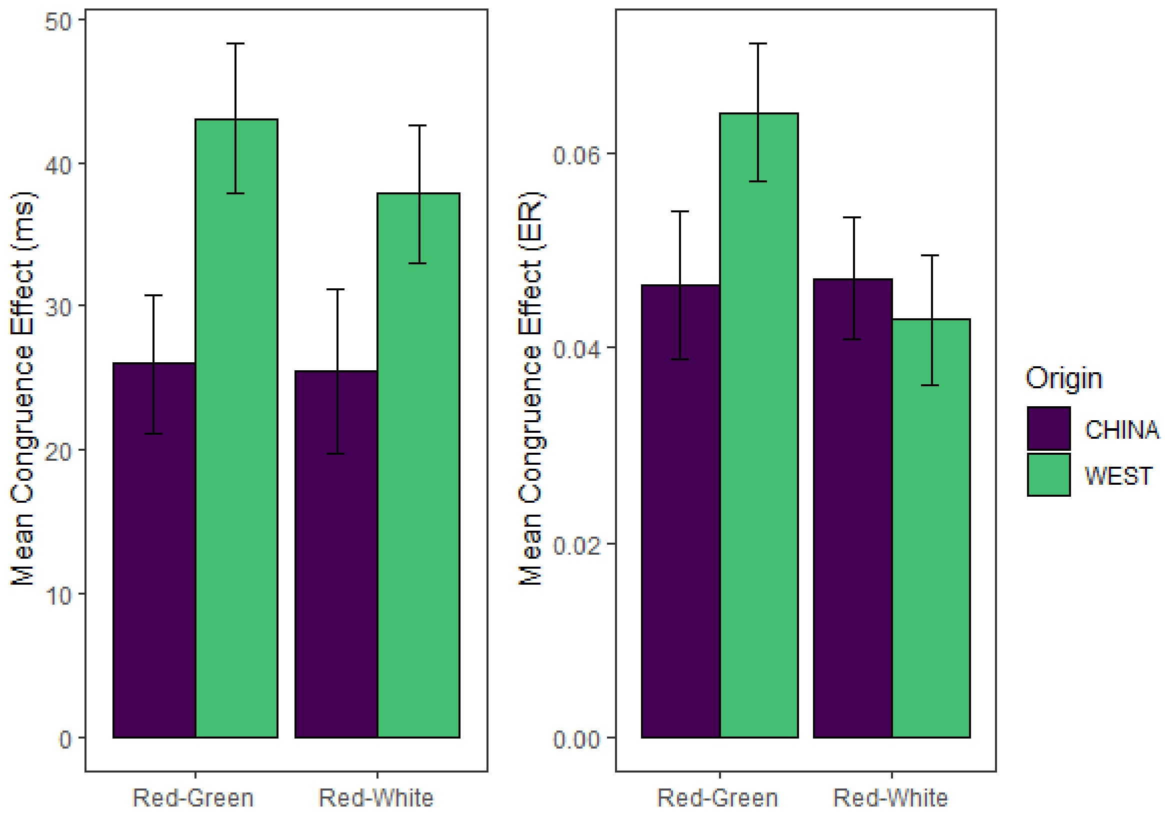


Figure 11

Mean Congruence Effect (Incongruent Minus Congruent Performance) in Experiment 2 for Reaction Times (ms; Left Panel) and for Error Rates (ER; Right Panel) per Color Opposition Block and Country. Error Bars Indicate SEMs.



Discussion

In general, the data still supported a cultural contribution to the red-valence association: In the red-green blocks, again RT congruence effects were weaker among Chinese than among Western (here, mostly German) cultures. Experiment 2, thus, showed that implicit color-valence associations also show with non-linguistic, pictorial stimulus material. The difference between cultures, however, was substantially weaker than in Experiment 1. This could be due to the elimination of word-processing, which might have contributed to the congruence effect in Experiment 1 but not in Experiment 2, or it could be due to, for instance, the higher noise levels in the online Experiment 2 than in the laboratory Experiment 1. In addition, as in Experiment 1, a stronger color-valence association in the Western as compared to the Chinese sample depended on the color-opposition context. The size of these color-valence interactions varied substantially between Western countries and China *only* when red was opposed to green color (see Figure 11). With a red-white color pairing, we did not observe a difference across cultures. Both of these findings are in contrast to anything we would have expected on the basis of existing explicit measures.

At the same time, similar RTs across countries in Experiment 2 showed that the speed advantage for Chinese speakers found in Experiment 1 was likely related to word processing.

General Discussion

In the current study, we investigated the possible origin of color-valence associations as either culturally specific or universal. We compared samples from an Eastern culture from Asia (China) with samples from a Western culture from Europe (Austria and Germany), as well as a culture with traits from both Eastern and Western cultures, only in Experiment 1, namely Macau (China/Asia).

In general, we found that, whether the stimulus is lexical or pictorial, our participants exhibited a great overlap in implicit color-valence associations across cultures. This was particularly true when color-valence associations were tested in red-white color opposition blocks (as compared to red-green color opposition blocks). Here, regardless of culture, our participants showed a trend for faster categorization of white-positive and red-negative pairings, and these associations were stronger (less ambiguous) for white than for red stimuli. This might come as a surprise in the light of reported explicit sadness associations for the color white by Chinese participants (Jonauskaite, Abu Akel, et al., 2020). A second surprise our studies revealed was that – maybe most interestingly – where explicit measures conceivably predicted cross-cultural similarity (due to homogenous explicit emotion associations for the colors red and green across cultures), our implicit measures revealed cultural differences, which were reflected in the strength of red-valence associations in red-green color opposition blocks. In both experiments, we demonstrated that the red-negative/green-positive association was stronger in Western than in Mainland Chinese samples and, in Experiment 1, that the association of Macanese participants lay in-between these groups. Green remains an unanimously positively associated color across cultures, according to our results, but it seems that its presence can tip the scale for the ambiguous red towards predominantly negative associations in the West, but not in the East. One possible reason for this may lie in the fact that red is simply perceived as a threat cue in the West because it is repeatedly used as a danger signal, most often when green is present and takes on the opposite (positive) meaning. In China, red may have predominantly positive associations and might be used (more often) in non-threatening contexts (e.g., red lanterns, street lights and building illumination; red coloring for hits in search results in internet search engines, etc.).

The present results, thus, support the value of implicit measures as an additional source of information besides the usage of explicit measures for understanding color-valence associations in general and how they work in color-opposition systems in particular.

Polar Opposites and Polarity Correspondence. In the current study, it is also possible that, in the Chinese sample, red was not a strong “antipole” to green because the two colors are not used in opposition as much in China as is the case in Western cultures. As mentioned earlier, white is at times cited as the counterpart to red (see *Introduction*). However, we did not find any evidence that would speak for a stronger (“more prototypical”) red-white than red-green opposition in the Chinese group (especially not with white being more negatively associated than red, see above). The fact that there was a trend towards a white-positivity association in both words and silhouettes speaks for the robustness of the effect and does not support a possible culture-specific ambiguous nature of white, but this invites further research. Maybe associations with positive emotions such as relief dominated the color-valence congruence effects in the Western and Chinese sample (cf. Jonauskaite, Abu Akel, et al., 2020). One should bear in mind that we used stimuli which did not semantically fit to only one particular color-system context, such as the Beijing opera or funerals versus weddings.

In this context, we want to mention that the *dimension-specific polarity attribution theory* by Schietecat et al. (2018a, 2018b) offers theoretical reasons for task-specific performance contributions in the presently chosen implicit categorization task. According to polarity attribution theory, the formation of color-emotion associations depends on the context that is semantically most salient (e.g., valence). The better the “fit” between the salient color and emotion dimension, the stronger the size of the association effect. This is explained by the gain of polarity attributions that play out stronger when two colors (e.g., green and red) can be mapped to a plus-pole and a minus-pole of the *same* semantic dimension as the two emotional categories (e.g., positive valence and negative valence). The better the overlap, the stronger the contribution of polarity correspondence effects (see also Lakens, 2012; Proctor & Cho, 2006). According to Schietecat et al. (2018b), red and green vary most strongly on the valence dimension, which allows for a strong mapping of red and negative valence to the minus-pole and simultaneously of green and positive valence to the plus-pole – as was demonstrated for a Western participant sample.

In the framework of the present study, this suggests the following: For the German-speaking sample, red and green seem to form a potent opposition on the valence dimension – stronger than the opposition that red and white form. This is possibly due to the fact that a red-white distinction activates another dimension (e.g., activity or dominance) more prominently than valence (but this remains to be tested). The more effective red-green opposition lead to stronger dimension-specific polarity contributions than the red-white opposition did. In our Chinese sample, in contrast, judging by the similarity in cross-modal association strength (as measured by congruence effect sizes) in both color context conditions, red does not form a stronger opposition to white than to green on the valence dimension. Congruence effect size was remarkably similar for the Chinese participants across both color opposition conditions and stimulus modalities (roughly around 25 ms in Experiments 1 and 2). Neither of the color contexts give rise to the same efficient mapping and, hence, to the same amount of dimension-specific polarity contributions as is the case in the 'prominent' red-green opposition in the Western sample.

Besides this general explanation of how the level or analysis or method chosen could have influenced the present result, this does not give us any more clues as to how to understand the particular origin in cultural experiences of this specific cross-cultural difference. What gives rise to the mappings of colors onto semantic dimensions in the first place? In theory, the answer to this question can be explored by explicit association tasks, but, judging by our findings of an absence of a culture-specific 'ambiguous white effect' as well as a presence of a culture-specific red-green-opposition effect, some effects might not come to light through explicit measures alone. Practically, one would, therefore, have to look beyond explicit measures, as the implications of culture-specific 'color-in-context' findings for applications are important in themselves, as color opposition and, thus, polarity correspondences, conceivably play a role in utility research and everyday tasks, too. As an example, think of the usage of color systems green and red buttons on a control board for go versus stop responses, respectively. In summary, by showing clear differences to explicit measures – more cross-cultural similarities for the color white, less cross-cultural similarities for the color red – the current implicit measure study reinforces the view that explicit and implicit approaches can yield complementary results and that both should be taken into account when planning applications. Without more systematic research on cross-cultural differences and similarities in implicit measures of color-valence and color-emotion associations, one can otherwise not easily predict which of several possible associations dominates in a particular context (cf. Elliot & Maier, 2012).

Words versus Pictures. Interestingly, the general effect of valence on response patterns differed between words and pictures in more than one respect. Not only were Chinese speaking participants faster to respond to words than German speaking participants, a performance difference that could not be replicated with silhouettes and that could be reflective of the faster semantic access to words in Chinese than German scripture: In Experiment 1, negative words also elicited *slower* responses and *more* errors, whereas in Experiment 2, negative silhouettes elicited *faster* responses and *fewer* errors. This finding is not without precedents. Other studies have shown that negative content captures attention efficiently (e.g., ‘Automatic Vigilance Effect’, Pratto & John, 1991) by drawing cognitive resources away from the analysis process and subsequent response execution when conveyed in word form (under similar paradigms as ours, e.g., Ansorge & Bohner, 2013; Meier & Robinson, 2004; Moller et al., 2009), but capturing attention and allocating resources towards faster response selection when conveyed in pictorial form (De Houwer & Hermans, 1994; Mogg et al., 2000; Schimmack, 2005; but see Giner-Sorolla et al., 1999; Ihssen & Keil, 2013).

Limitations. The online experiment (Experiment 2) poses the problem of less reliable RT measurement, but, fortunately, reliability was at least high enough for a replication of the general pattern found under more controlled laboratory conditions, too (i.e., in Experiment 1). In addition, as mentioned earlier, we did not have strict control over the actual colorimetrics throughout both studies (usage of different laboratories in different countries in Experiment 1, online study in Experiment 2). The slower reactions and higher error rates for green silhouettes (color main effect) could reflect problems with the presentation of this color. This might be an artefact of the online setting, despite our efforts to equate red and green in lightness and saturation and testing participants' color discrimination ability. Judging from the response patterns in the ERs, however, it seems that the negative silhouettes in green received disproportionately many misclassifications in comparison to the relatively low ERs among the other color-valence combinations. This suggests generally a strong interference of green color on correct classification of negative content. Since there was a speed advantage for white over red stimuli, too (which was to be expected due to the maximal lightness of white), the color main effects (particularly in the RTs) could be salience-related (red was perceptually more salient than green; white was perceptually more salient than red). In any case, it would be desirable to run a similar study in the future under more rigidly controlled laboratory conditions, to ensure consistent color presentation and viewing conditions. Such a study could also help to tell the influences of the different color dimensions to the currently measured color-valence congruence effects apart (cf. Schloss et al., 2020).

Conclusion

Color-valence associations are important for many applications. The current study showed that color-valence associations form via various modes of delivery – linguistic and pictorial. However, not all color-valence associations apply universally and in each context of color systems, so caution is advised when using them in international contexts and together with alternative colored signals. For Western populations, association strength of negative-red and positive-green is stronger than for Chinese participants. The red-green opposition seems, hence, particularly effective (in terms of polarity attributions) in the West. A red-white opposition seems to allow for relatively weaker red-negativity associations from a Western viewpoint, but neither more nor less efficient ones from a Chinese viewpoint.

Open Practices Statements

All studies were preregistered using the OSF (Foster & Deardorff, 2017), accessible via osf.io/27gf8. The supplementary analyses, source code for the experiments, all data collected and materials used are available via the same link.

References

Ansorge, U., & Bohner, G. (2013). Investigating the association between valence and elevation with an implicit association task that requires upward and downward responding. *Universitas Psychologica*, *12*(5), 1453-1471. https://doi.org/10.11144/Javeriana.UPSY12-5.iave

Bradley, M. M., & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In *Handbook of emotion elicitation and assessment* (pp. 29–46). Oxford University Press.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*(6), Article e10729. https://doi.org/10.1371/journal.pone.0010729

China National Tourist Office (2020, April 17). *Peking Opera Facial Masks*. http://cnto.org.uk/the-culture4-2/beijing-opera.html

Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray.

De Houwer, J., & Hermans, D. (1994). Differences in the affective processing of words and pictures. *Cognition & Emotion*, *8*(1), 1–20. https://doi.org/10.1080/02699939408408925

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's *t*-test instead of Student's *t*-test. *International Review of Social Psychology*, *30*(1), 92. https://doi.org/10.5334/irsp.82

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face*. Prentice Hall.

Elliot, A. J., & Maier, M. A. (2012). Color-in-context theory. In *Advances in experimental social psychology* (Vol. 45, pp. 61-125). Academic Press.

Elliot, A. J., & Niesta, D. (2008). Romantic red: Red enhances men's attraction to women. *Journal of Personality and Social Psychology*, *95*(5), 1150–1164. https://doi.org/10.1037/0022-3514.95.5.1150

Elliot, A. J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., & Liu, H. (2010). Red, rank, and romance in women viewing men. *Journal of Experimental Psychology: General*, *139*(3), 399–417. https://doi.org/10.1037/a0019689

Fetterman, A. K., Robinson, M. D., & Meier, B. P. (2012). Anger as "seeing red": Evidence for a perceptual association. *Cognition & Emotion*, *26*(8), 1445–1458. https://doi.org/10.1080/02699931.2012.673477

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association : JMLA*, *105*(2), 203–206. https://doi.org/10.5195/jmla.2017.88

Garrido, M. V., Prada, M., Simão, C., & Semin, G. R. (2019). The Impact of Stimuli Color in Lexical Decision and Semantic Word Categorization Tasks. *Cognitive Science*, *43*(8), 1–21. https://doi.org/10.1111/cogs.12781

Giner-Sorolla, R., Garcia, M. T., & Bargh, J. A. (1999). The automatic evaluation of pictures. *Social Cognition*, *17*(1), 76–96. http://dx-doi-org.uaccess.univie.ac.at/10.1521/soco.1999.17.1.76

Hochman, S., Henik, A., & Kalanthroff, E. (2018). Stopping at a red light: Recruitment of inhibitory control by environmental cues. *PloS One*, *13*(5), e0196199. http://dx-doi-org.uaccess.univie.ac.at/10.1371/journal.pone.0196199

He, G. (2011). A comparative study of color metaphors in English and Chinese. *Theory and Practice in Language Studies*, *1*(12), 1804–1808. https://doi.org/10.4304/tpls.1.12.1804-1808

Ihssen, N., & Keil, A. (2013). Accelerative and decelerative effects of hedonic valence and emotional arousal during visual scene processing. *Quarterly Journal of Experimental Psychology*, *66*(7), 1276–1301. https://doi.org/10.1080/17470218.2012.737003

Jiang, F., Lu, S., Yao, X., Yue, X., & Au, W. T. (2014). Up or down? How culture and color affect judgments. *Journal of Behavioral Decision Making*, *27*(3), 226–234. https://doi.org/10.1002/bdm.1800

Jonauskaite, D., Abu-Akel, A., Dael, N., Oberfeld, D., Abdel-Khalek, A. M., Al-Rasheed, A. S., Antonietti, J.-P., Bogushevskaya, V., Chamseddine, A., Chkonia, E., Corona, V., Fonseca-Pedrero, E., Griber, Y. A., Grimshaw, G., Hasan, A. A., Havelka, J., Hirnstein, M., Karlsson, B. S. A., Laurent, E., … Mohr, C. (2020). Universal patterns in color-emotion associations are further shaped by linguistic and geographic proximity. *Psychological Science*, *31*(10), 1245-1260. https://doi.org/10.1177/0956797620948810

Jonauskaite, D., Parraga, C. A., Quiblier, M., & Mohr, C. (2020). Feeling blue or seeing red? Similar patterns of emotion associations with colour patches and colour terms. *I-Perception*, *11*(1), Article 204166952090248. https://doi.org/10.1177/2041669520902484

Jonauskaite, D., Wicker, J., Mohr, C., Dael, N., Havelka, J., Papadatou-Pastou, M., Zhang, M., & Oberfeld, D. (2019). A machine learning approach to quantify the specificity of colour–emotion associations and their cultural differences. *Royal Society Open Science*, *6*(9), Article 190741. https://doi.org/10.1098/rsos.190741

Kawai, C., Lukács, G., & Ansorge, U. (2020). Polarities influence implicit associations between colour and emotion. *Acta Psychologica*, *209*, Article 103143. https://doi.org/10.1016/j.actpsy.2020.103143

Kawai, C., Lukács, G., & Ansorge, U. (2021). A new type of pictorial database: The Bicolor Affective Silhouettes and Shapes (BASS). *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01569-7

Kelley, K. (2019). MBESS: The MBESS R Package. R package version 4.5.1. https://CRAN.R-project.org/package=MBESS

Kuhbandner, C., & Pekrun, R. (2013). Joint effects of emotion and color on memory. *Emotion*, *13*(3), 375–379. https://doi.org/10.1037/a0031821

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 726–736. https://doi.org/10.1037/a0024955

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Lakens, D., Fockenberg, D. A., Lemmens, K. P. H., Ham, J., & Midden, C. J. H. (2013). Brightness differences influence the evaluation of affective pictures. *Cognition & Emotion*, *27*(7), 1225–1246. https://doi.org/10.1080/02699931.2013.781501

Lawrence, M. A. (2016). Ez: Easy analysis and visualization of factorial experiments. R package version 4.4-0. https://CRAN.R-project.org/package=ez

Lehmann, G. K., Elliot, A. J., & Calin-Jageman, R. J. (2018). Meta-analysis of the effect of red on perceived attractiveness. *Evolutionary Psychology, 16*(4), 1474704918802412. https://doi.org/10.1177/1474704918802412

Lukács, G. (2020). neatStats: An R Package for neat and painless statistical reporting. R package version 1.5.1. https://github.com/gasparl/neatstats

Lü, X., & Zhang, J. (1999). Reading efficiency: A comparative study of English and Chinese orthographies. *Reading Research and Instruction*, *38*(4), 301–317. https://doi.org/10.1080/19388079909558298

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Mathôt, S. (2017, May 15). Bayes like a baws: Interpreting Bayesian repeated measures in JASP. https://www.cogsci.nl/blog/interpreting-bayesian-repeated-measures-in-jasp

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science*, *15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x

Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white: Automatic inferences about stimulus valence based on brightness. *Psychological Science*, *15*(2), 82–87. https://doi.org/10.1111/j.0963-7214.2004.01502002.x

Mogg, K., McNamara, J., Powys, M., Rawlinson, H., Seiffer, A., & Bradley, B. P. (2000). Selective attention to threat: A test of two cognitive models of anxiety. *Cognition & Emotion*, *14*(3), 375–399. https://doi.org/10.1080/026999300378888

Moller, A. C., Elliot, A. J., & Maier, M. A. (2009). Basic hue-meaning associations. *Emotion*, *9*(6), 898–902. https://doi.org/10.1037/a0017811

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor

Perfetti, C. A., Liu, Y., & Tan, L.-H. (2002). How the mind can meet the brain in reading: A comparative writing systems approach. *Cognitive Neuroscience Studies of the Chinese Language*, 36–60.

Peperkoorn, L. S., Roberts, S. C., & Pollet, T. V. (2016). Revisiting the red effect on attractiveness and sexual receptivity: No effect of the color red on human mate preferences. *Evolutionary Psychology, 14*(4), Article 1474704916673841. https://doi.org/10.1177/1474704916673841

Pflüger, L. S., Valuch, C., Gutleb, D. R., Ansorge, U., & Wallner, B. (2014). Colour and contrast of female faces: Attraction of attention and its dependence on male hormone status in Macaca fuscata. *Animal Behaviour, 94*, 61-71. https://doi.org/10.1016/j.anbehav.2014.05.022

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*(3), 380–391. https://doi.org/10.1037/0022-3514.61.3.380

Pravossoudovitch, K., Cury, F., Young, S. G., & Elliot, A. J. (2014). Is red the colour of danger? Testing an implicit red–danger association. *Ergonomics*, *57*(4), 503–510. https://doi.org/10.1080/00140139.2014.889220

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416–442. https://doi.org/10.1037/0033-2909.132.3.416

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Schietecat, A. C., Lakens, D., IJsselsteijn, W. A., & De Kort, Y. A. W. (2018a). Predicting context-dependent cross-modal associations with dimension-specific polarity attributions. Part 1 – Brightness and aggression. *Collabra: Psychology*, *4*(1), 14. https://doi.org/10.1525/collabra.110

Schietecat, A. C., Lakens, D., IJsselsteijn, W. A., & De Kort, Y. A. W. (2018b). Predicting context-dependent cross-modal associations with dimension-specific polarity attributions. Part 2: Red and valence. *Collabra: Psychology*, *4*(1), 21. https://doi.org/10.1525/collabra.126

Schimmack, U. (2005). Attentional interference effects of emotional pictures: Threat, negativity, or arousal? *Emotion*, *5*(1), 55–66. https://doi.org/10.1037/1528-3542.5.1.55

Schloss, K., Witzel, C., & Lai, L. Y. (2020). Blues don't bring the blues: challenging conventional notions of color-emotion associations. *Journal of the Optical Society of America A*, *37*(5), 813-824. http://eprints.soton.ac.uk/id/eprint/438858

Sokal, R. R., & Rohlf, J. F. (1981). *Biometry: The principles and practice of statistics in biological research* (2nd ed.). Freeman.

Sorokowski, P., Szmajke, A., Hamamura, T., Jiang, F., & Sorokowska, A. (2014). “Red wins”, “black wins” and “blue loses” effects are in the eye of beholder, but they are culturally universal: A cross-cultural analysis of the influence of outfit colours on sports performance. *Polish Psychological Bulletin*, *45*(3), 318–325. https://doi.org/10.2478/ppb-2014-0039

Squire, L. R. (1986). Mechanisms of memory. *Science, 232*(4578), 1612–1619. https://doi.org/10.1126/science.3086978

Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology, 44*(1), 453-495. https://doi.org/10.1146/annurev.ps.44.020193.002321

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, *9*(2), 164–182. https://doi.org/10.1037/1082-989X.9.2.164

Stephen, I. D., & McKeegan, A. M. (2010). Lip colour affects perceived sex typicality and attractiveness of human faces. *Perception, 39*(8), 1104-1110. https://doi.org/10.1068/p6730

Võ, M., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M., & Jacobs, A. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

Wang, T., Shu, S., & Mo, L. (2014). Blue or red? The effects of colour on the emotions of Chinese people. *Asian Journal of Social Psychology*, *17*(2), 152–158. https://doi.org/10.1111/ajsp.12050

Wang, Y. (2013). From funeral to wedding ceremony: Change in the metaphoric nature of the Chinese color term white. *Semiotica*, *2013*(193), 361–380. https://doi.org/10.1515/sem-2012-0034

Wilms, L., & Oberfeld, D. (2018). Color and emotion: Effects of hue, saturation, and brightness. *Psychological Research*, *82*(5), 896–914. https://doi.org/10.1007/s00426-017-0880-8

Wu, Y., Lu, J., van Dijk, E., Li, H., & Schnall, S. (2018). The color red is implicitly associated with social status in the United Kingdom and China. *Frontiers in Psychology*, *9*, Article 1902. https://doi.org/10.3389/fpsyg.2018.01902

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, *16*(3), 561–566. https://doi.org/10.3758/PBR.16.3.561

Appendix

Table A1

Means and SDs of the Mean Response Times (in ms) and Mean Error Rates for Experiments 1 and 2.

			Red-Green Color Block				Red-White Color Block			
			Red		Green		Red		White	
Experiment	Country	DV	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive
1	China	RT	628±86	613±83	645±77	598±76	631±66	605±77	646±75	590±79
		ER	6.2±4.7	5.1±5.7	11.5±9.5	2.5±2.4	6.6±5.0	4.7±4.7	9.6±7.1	2.7±2.8
	Macau	RT	647±56	658±53	672±67	620±50	669±67	644±61	673±70	621±65
		ER	7.1±4.4	6.3±4.7	13.4±7.6	3.5±2.9	9.2±7.2	6.3±3.9	10.5±6.2	3.4±3.3
	West	RT	725±88	738±83	755±83	691±79	697±78	683±72	708±85	658±69
		ER	3.1±3.0	8.3±5.8	10.0±7.7	3.0±4.2	5.2±4.5	7.2±4.8	7.3±5.9	3.6±3.0
2	China	RT	708±84	748±102	738±90	752±95	702±78	742±106	697±83	711±90
		ER	8.2±4.9	8.0±5.2	12.2±7.4	7.4±5.1	8.5±5.2	8.5±5.2	11.1±6.2	6.4±4.6
	West	RT	698±89	751±111	731±94	741±101	690±86	738±101	694±90	705±95
		ER	5.7±4.0	6.5±5.0	10.2±5.9	4.6±4.2	6.5±4.3	6.4±5.0	8.6±5.4	4.2±3.9

Note: In Experiment 1, we show the means of the median correct RTs and the means of the untransformed mean ERs, and only for the first experimental block performed by the participant. RT = Reaction Time; ER = Error Rate

Table A2

Mean Valence and Arousal Rating Scores (US and China) and the Average of Colored (i.e., Not Background-Colored) Pixels for the Positive and Negative Silhouettes used in Experiment 2.

Valence	Count	Arousal US	Arousal China	Valence US	Valence China	Colored pixels (out of 90,000)
Negative	80	5.260	4.671	3.408	3.759	21,906
Positive	80	5.036	5.653	6.949	6.531	22,176
		5.148	**5.162**	**5.178**	**5.145**	**22,041**

3. Discussion

Each of the three studies of this doctoral thesis explored potential modulating factors in the formation of color-valence associations, by leveraging the characteristic traits of the implicit valence-classification protocol (i.e., structural a/symmetry and color opposition). Study I established the green-positivity/red-negativity relationship and demonstrated that the contribution of polarity correspondence greatly affects connotations of red color. Study II replicated the strong oppositional congruence of green-positivity/red-negativity (and additionally for white-positivity/black-negativity) for affective silhouette pictures from the BASS. This database provides a tool to select controlled affective stimulus material for an East-West culture comparison that is language-independent. Study III built on these prior findings by exploring the nature of color systems from a cross-cultural perspective with linguistic and non-linguistic stimuli. We found that, when it comes to implicit associations, color context itself functions as a context factor and, importantly, these color systems are subject to cultural modulation: Whether a color's prominent meaning changes in the presence of another color differs between cultures.

3.1 The Pervasiveness of Color Systems

The question posed by our results is: Why are red-responses selectively influenced by green color opposition context for Westerners and not Chinese participants (cf. Schietecat et al., 2018b)? The answer to this question cannot be found in cross-cultural surveys on explicit color-emotion associations (Adams & Osgood, 1973; Jonauskaite et al., 2020) because, as of yet, the comparison of color systems has not constituted a research area of international interest, despite its ecological validity and evident practical implications (see Introduction).

The impact of color on human affect, behavior, and cognition has been demonstrated time and time again (e.g., Elliot & Maier, 2012), and in a context of globalization, geographic mobility, and interconnectedness (factors that could in themselves act upon color meaning development, see Kawai, Lukács, & Ansorge, 2021; Kawai, Zhang, et al., 2021), emotion processing from a culture-encompassing perspective is an especially important topic of research. An understanding of differences and similarities in conceptualization of affect is vital in human-human interaction, but also extends to human-machine interaction and AI development. Here, the importance of culture-specific intuitive design based on color associations cannot be overstated (recall the machine envisioned in the Introduction). In this thesis, I argue for an awareness of culture-specific message signaling within color systems.

3.2 Open Questions

The question of nature-versus-nurture-driven origins of color meaning cannot be settled with this thesis, but our results generally conform with a view that incorporates both aspects: Culture can (selectively) override any evolutionarily developed dominant associations (Elliot & Maier, 2012; on the importance of individual differences see Hong et al., 2020).

Throughout our studies, we used the same methodology as consistently as possible (across cultures and settings) for maximal comparability of the results. Despite our efforts to control for color presentation and stimulus translation, confounding influences in the data from our multi-lab study (Austria, Macau, and Xi'an) cannot be ruled out. Similarly, all silhouette studies were performed online, making color and environmental control notoriously challenging. While online research has generally proven reliable (e.g., Anwyl-Irvine et al., 2020; Gosling & Mason, 2015; see also our lab vs. online comparison in the supplementary material to Kawai, Lukács, & Ansorge, 2021), we implemented color discrimination tests and performance exclusion criteria in an attempt to minimize volatilities. Still, there were variations in the (Western) red-green silhouette classifications that were both performed online (cf. Studies II and III), but since these two studies did not use the exact same stimulus selection and shades, the origin of this divergence is to be determined through follow-up action. Additionally, in Study III, there were some inevitable differences in what concepts were tested in the word versus the silhouette experiment. While we generally tried to include similar concepts for the linguistic and non-linguistic material (positive and negative actions and objects), some abstract affective terms (e.g., ‘unfair’, ‘magnificent’) could not be conveyed with pictures. Instead, Experiment 2 contained a greater number of silhouettes depicting physical exercise and romantic situations on the positive side, and weapons, pests, and predators on the negative side. At the same time, the fact that a clear green-positivity/red-negativity congruence emerged in all cases underscores the stability of the effect.

Finally, I want to emphasize that we limited our investigation to superordinate valence categories, subsets of colors, and specific culture samples. In general, we pooled measurements taken from many individual nuanced concepts, trading specificity (‘sad, ‘weak’, ‘pain’) for generalizability (‘negative’). However, discrete emotions and even narrower semantic categories like ‘success’, ‘failure’, or ‘romance’ (Moller et al., 2009; Winskel et al., 2021) may have distinct links to colors (see also Briesemeister et al., 2011, 2014). Note also that not all of the concepts we tested via words and silhouettes necessarily have to trigger certain color-valence associations (congruences) reliably. It is possible that characteristic terms associated with a particular color were distributed differently between the examined culture groups. These terms might be the exact ones that would come to the minds of the participants first if they were asked to explicitly rate a color's emotional meaning. Terms like ‘funeral’ (see Section 1.1.2), for example, and other words containing a color term in the Chinese character were explicitly excluded from stimulus selection to avoid confounding influences. At the same time, however, our methodology allowed us to measure repeatedly (approximately 50 measurements were taken per participant per cell), increasing reliability and decreasing noise. An advisable procedure would be to follow up specific explicit associations with implicit methods.

The intra-Chinese differences between Macau and the Mainland, as well as the age-modulated differences in emotion ratings revealed in Study II hint at a multitude of sociodemographic factors that could shape the interpretation of a color's affective meaning. Therefore, we encourage replications and follow-up investigations on all studies presented in this thesis (we preregistered all our studies, data and material are available online, links are provided in the respective manuscripts).

3.3 Implications for the Field

This thesis stresses the expediency of a complementary use of implicit and explicit measures to reach a deep and encompassing understanding of color-emotion associations across cultures. The research community should also take on investigations into how the effects we specified here, in particular color systems, may influence explicit measures, like color-emotion ratings or reports.

3.4 Conclusion

It is astounding how so many different hues can display such a high degree of similarity in associated meaning across a multitude of cultures all over the globe. This certainly lends credence to the statement that "color may represent something of a lingua franca—a common language capable of conveying information across the human life span" (Maier et al., 2008, p. 1538). However, our research has shown that there are some dialectic variations in that language to which we should not turn a blind eye.

4. References

Adams, F. M., & Osgood, C. E. (1973). A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology*, *4*(2), 135–156. https://doi.org/10.1177/002202217300400201

Ansorge, U. (2021). Knowledge in motion: How procedural control of knowledge usage entails selectivity and bias. *Journal of Knowledge Structures and Systems*, *2*(1), 3–28.

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2020). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01501-5

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(04). https://doi.org/10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bianchi, I., & Savardi, U. (2008). *The perception of contraries* (1. ed). Aracne.

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2011). Discrete emotion effects on lexical decision response times. *PLOS ONE*, *6*(8), e23743. https://doi.org/10.1371/journal.pone.0023743

Briesemeister, B. B., Kuchinke, L., & Jacobs, A. M. (2014). Emotion word recognition: Discrete information effects first, continuous later? *Brain Research*, *1564*, 62–71. https://doi.org/10.1016/j.brainres.2014.03.045

Clarke, T., & Costall, A. (2008). The emotional connotations of color: A qualitative investigation. *Color Research & Application*, *33*(5), 406–410. https://doi.org/10.1002/col.20435

Elliot, A. J., & Maier, M. A. (2012). Color-in-Context Theory. In *Advances in Experimental Social Psychology* (Vol. 45, pp. 61–125). Elsevier. https://doi.org/10.1016/B978-0-12-394286-9.00002-0

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*(1), 154–168. https://doi.org/10.1037/0096-3445.136.1.154

Elliot, A. J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., & Liu, H. (2010). Red, rank, and romance in women viewing men. *Journal of Experimental Psychology: General*, *139*(3), 399–417. https://doi.org/10.1037/a0019689

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin.

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology*, *66*(1), 877–902. https://doi.org/10.1146/annurev-psych-010814-015321

Greeno, J. G. (1994). Gibson's affordances. *Psychological Review*, *101*(2), 336–342. https://doi.org/10.1037/0033-295X.101.2.336

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464.

Havas, D. A., & Matheson, J. (2013). The functional role of the periphery in emotional language comprehension. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00294

He, G. (2011). A comparative study of color metaphors in English and Chinese. *Theory and Practice in Language Studies*, *1*(12), 1804–1808. https://doi.org/10.4304/tpls.1.12.1804-1808

Hong, X., He, Z., & Zou, R. (2020). Individual differences in implicit associations between red and green colors and failure and success among Chinese adults. *Social Behavior and Personality*, *48*(8), 1–7. http://dx-doi-org.uaccess.univie.ac.at/10.2224/sbp.9228

Jiang, F., Lu, S., Yao, X., Yue, X., & Au, W. T. (2014). Up or down? How culture and color affect judgments. *Journal of Behavioral Decision Making*, *27*(3), 226–234.

Jonauskaite, D., Abu-Akel, A., Dael, N., Oberfeld, D., Abdel-Khalek, A. M., Al-Rasheed, A. S., Antonietti, J.-P., Bogushevskaya, V., Chamseddine, A., Chkonia, E., Corona, V., Fonseca-Pedrero, E., Griber, Y. A., Grimshaw, G., Hasan, A. A., Havelka, J., Hirnstein, M., Karlsson, B. S. A., Laurent, E., … Mohr, C. (2020). Universal patterns in color-emotion associations are further shaped by linguistic and geographic proximity. *Psychological Science*, 0956797620948810. https://doi.org/10.1177/0956797620948810

Jonauskaite, D., Wicker, J., Mohr, C., Dael, N., Havelka, J., Papadatou-Pastou, M., Zhang, M., & Oberfeld, D. (2019). A machine learning approach to quantify the specificity of colour–emotion associations and their cultural differences. *Royal Society Open Science*, *6*(9), 190741. https://doi.org/10.1098/rsos.190741

Kawai, C., Lukács, G., & Ansorge, U. (2020). Polarities influence implicit associations between colour and emotion. *Acta Psychologica*, *209*, 103143. https://doi.org/10.1016/j.actpsy.2020.103143

Kawai, C., Lukács, G., & Ansorge, U. (2021). A new type of pictorial database: The Bicolor Affective Silhouettes and Shapes (BASS). *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01569-7

Kawai, C., Zhang, Y., Lukács, G., Chu, W., Zheng, C., Gao, C., Gozli, D., Wang, Y., & Ansorge, U. (2021). *The good, the bad, and the red: Implicit color-valence associations across cultures* [Manuscript submitted for publication]. Department of Cognition, Emotion, and Methods in Psychology, University of Vienna.

Kaya, N., & Epps, H. H. (2004). Relationship between color and emotion: A study of college students. *College Student Journal*, *38*(3), 396–405.

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, *49*(2), 457–470. https://doi.org/10.3758/s13428-016-0715-3

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 726–736. https://doi.org/10.1037/a0024955

Lakens, D., Semin, G. R., & Foroni, F. (2012). But for the bad, there would not be good: Grounding valence in brightness through shared relational structures. *Journal of Experimental Psychology: General*, *141*(3), 584–594. https://doi.org/10.1037/a0026468

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, *4*(2), 195–208. https://doi.org/10.1207/s15516709cog0402_4

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*, 39–58.

Maier, M. A., Elliot, A. J., & Lichtenfeld, S. (2008). Mediation of the negative effect of red on intellectual performance. *Personality and Social Psychology Bulletin*, *34*(11), 1530–1540. https://doi.org/10.1177/0146167208323104

Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white: Automatic inferences about stimulus valence based on brightness. *Psychological Science*, *15*(2), 82–87. https://doi.org/10.1111/j.0963-7214.2004.01502002.x

Moller, A. C., Elliot, A. J., & Maier, M. A. (2009). Basic hue-meaning associations. *Emotion*, *9*(6), 898–902. https://doi.org/10.1037/a0017811

Osgood, C. E., & Richards, M. M. (1973). From Yang and Yin to and or but. *Language*, *49*(2), 380–412. JSTOR. https://doi.org/10.2307/412460

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois press.

Palmer, S. E., & Schloss, K. B. (2011). Ecological valence and human color preference. *New Directions in Colour Studies*, *64*, 361–376.

Pflüger, L. S., Valuch, C., Gutleb, D. R., Ansorge, U., & Wallner, B. (2014). Colour and contrast of female faces: Attraction of attention and its dependence on male hormone status in Macaca fuscata. *Animal Behaviour*, *94*, 61–71. https://doi.org/10.1016/j.anbehav.2014.05.022

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416–442. https://doi.org/10.1037/0033-2909.132.3.416

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*(2), 139–165. https://doi.org/10.1037/0096-3445.133.2.139

Schietecat, A. C., Lakens, D., IJsselsteijn, W. A., & De Kort, Y. A. W. (2018a). Predicting context-dependent cross-modal associations with dimension-specific polarity attributions. Part 1 – Brightness and aggression. *Collabra: Psychology*, *4*(1), 14. https://doi.org/10.1525/collabra.110

Schietecat, A. C., Lakens, D., IJsselsteijn, W. A., & De Kort, Y. A. W. (2018b). Predicting context-dependent cross-modal associations with dimension-specific polarity attributions. Part 2: Red and valence. *Collabra: Psychology*, *4*(1), 21. https://doi.org/10.1525/collabra.126

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215.

Specker, E., & Leder, H. (2018). Looking on the bright Side: Replicating the association between brightness and positivity. *Collabra: Psychology*, *4*(1), 34. https://doi.org/10.1525/collabra.168

Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, *123*(4), 394–409.

Winskel, H., Forrester, D., Hong, M., & O'Connor, K. (2021). Seeing red as anger or romance: An emotion categorisation task. *Journal of Cognitive Psychology*, 1–14. https://doi.org/10.1080/20445911.2021.1936538

5. English Abstract

Previous research has identified astounding global similarities in explicit emotion associations with a variety of colors—but, at the same time, language and culture do exert modulating influences, especially between Western and Chinese culture. For example, green seems universally associated with positive emotion, and red is found to be highly ambiguous in meaning (love vs. anger) in many cultures; less cross-cultural agreement prevails for White, which displays significantly more negative associations in China than in the West. In this thesis, we assessed and compared different aspects of color-valence associations across two major culture groups—the West and China—with implicit measures, a method that is rarely applied in this context, but very fruitful. Two context factors that shape color-valence associations (and might potentially also be present in explicit measures) were identified as significant: (1) polarity (presence of a color in isolation vs. in color context), and (2) the color system (which colors are opposing each other in an affective context). Importantly, color systems are themselves subject to modulation by culture. Additionally, we showed that, apart from linguistic stimuli, images can give rise to congruence effects between color and valence, too, in similar directions as words do. This was demonstrated using silhouette pictures from the BASS, a bicolor affective pictorial stimulus database we created for cross-cultural comparison. We argue that both—implicit and explicit measures—should be applied in a complementary fashion to reach a deep and encompassing understanding of color-emotion associations across cultures.

6. German Abstract

Kulturvergleichende Studien zu expliziten emotionalen Farbbedeutungen fanden verblüffende Ähnlichkeiten, zeigten aber auch den modulierenden Einfluss von Kultur. Gün ist beispielsweise durchgängig überwiegend positiv assoziiert und Rot ist in vielen Kulturen eine sehr vieldeutige Farbe (Liebe, Wut). Bei Weiß dagegen zeigt sich weniger Übereinstimmung: in China trägt Weiß deutlich mehr negative Assoziationen als im Westen. In dieser Arbeit vergleiche ich verschiedene Aspekte von Farb-Valenz-Assoziationen zwischen diesen zwei Kulturgruppen (Westen vs. China) mit impliziter Methodik. Zwei Kontextfaktoren, die Farb-Valenz-Assoziationen beeinflussen, wurden identifiziert: Polarität und Farbsysteme (Ausprägung der Farbopposition). Die vorliegende Doktorarbeit zeigen erstmals die Kulturabhängigkeit von Farbsystemen; ein Faktor mit praktischer Relevanz in internationaler Kommunikation, Mensch-Maschine-Interaktion, und Usability. Darüber hinaus konnten wir mit Hilfe unserer genormten Silhouettendatenbank (BASS) darlegen, dass nicht nur sprachliche Reize, sondern auch Bilder Kongruenzeffekte zwischen Farbe und Valenz hervorrufen können. Ich plädiere generell für komplementäre Anwendung impliziter und expliziter Methoden, für ein umfassendes Verständnis von emotionalen Farbbedeutungen in verschiedenen Kulturen.