Acta Psychologica 206 (2020) 103067

Contents lists available at ScienceDirect

Acta Psychologica

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

# Bliss is blue and bleak is grey: Abstract word-colour associations influence objective performance even when not task relevant

Stephanie C. Goodhew[a,*], Evan Kidd[a,b,c]

[a] Research School of Psychology, The Australian National University, Australia
[b] ARC Centre of Excellence for the Dynamics of Language, Australia
[c] Max Planck Institute for Psycholinguistics, Netherlands

ARTICLE INFO

*Keywords:*
Concepts
Abstract words
Colour
Valence
Semantic representations
Stroop

ABSTRACT

Humans associate abstract words with physical stimulus dimensions, such as linking upward locations with positive concepts (e.g., *happy* = up). These associations manifest both via subjective reports of associations and on objective performance metrics. Humans also report subjective associations between colours and abstract words (e.g., *joy* is linked to yellow). Here we tested whether such associations manifest on objective task performance, even when not task-relevant. Across three experiments, participants were presented with abstract words in physical colours that were either congruent with previously-reported subjective word-colour associations (e.g., *victory* in red and *unhappy* in blue), or were incongruent (e.g., *victory* in blue and *unhappy* in red). In Experiment 1, participants' task was to identify the valence of words. This congruency manipulation systematically affected objective task performance. In Experiment 2, participants completed two blocks, a valence-identification and a colour-identification task block. Both tasks produced congruency effects on performance, however, the results of the colour identification block could have reflected learning effects (i.e., associating the more common congruent colour with the word). This issue was rectified in Experiment 3, whereby participants completed the same two tasks as Experiment 2, but now matched congruent and incongruent pairs were used for both tasks. Again, both tasks produced reliable congruency effects. Item analyses in each experiment revealed that these effects demonstrated a degree of item specificity. Overall, there was clear evidence that at least some abstract word-colour pairings can systematically affect behaviour.

## 1. Introduction

Humans systematically associate particular stimulus properties across different sensory modalities. For example, associations between particular sounds and certain visual shapes is demonstrated in the now-famous example in which people overwhelming link the sound of the verbal name “bouba” to a rounded-contour visual shape and the name “kiki” to a sharp-angled shape (Ramachandran & Hubbard, 2001). In a similar vein, it has been shown that auditory stimuli are associated with vertical spatial locations, such that high pitch tones are linked with upward locations, whereas low pitch tones are linked to downward locations (Evans & Treisman, 2010). Furthermore, certain tones are associated with particular flavours (Knöferle & Spence, 2012). Such associations even extend into conceptual-sensory links. That is, humans systematically associate particular concepts with particular sensory dimensions. For instance, humans tend to associate valence and physical space, such that positive valence words (e.g., *happy*) are linked to upward locations, while negative valence words (e.g., *sad*) are linked to downward locations. This manifests not only in participants' explicit reports of associations (Goodhew & Kidd, 2016), but also implicitly on objective task performance measures (Dudschig, De la Vega, & Kaup, 2015; Goodhew, McGaw, & Kidd, 2014; Gozli, Chasteen, & Pratt, 2013; Gozli, Chow, Chasteen, & Pratt, 2013). To summarise, certain seemingly unrelated conceptual and sensory stimuli naturally coalesce in the human mind. Documenting these associations is important from a theoretical perspective, because understanding these associations helps us to understand the cognitive architecture of the human mind.

The purpose of the present study was to examine whether humans exhibit, on objective behavioural metrics, systematic associations between abstract words and physical colours. This is a novel research question; however, there are several phenomena already documented in the literature that indicate it is a plausible possibility. First, it has been established that humans associate colours with concrete (as opposed to abstract) words. This is shown via the *semantic Stroop* effect.

* Corresponding author at: Research School of Psychology (Building 39), The Australian National University, Canberra, ACT, Australia.
*E-mail address:* Stephanie.Goodhew@anu.edu.au (S.C. Goodhew).

https://doi.org/10.1016/j.actpsy.2020.103067
Received 16 July 2019; Received in revised form 9 March 2020; Accepted 27 March 2020

This is similar to the standard Stroop, except that instead of direct colour words, colour-associated words are used. For example, "lemon" is used to denote *yellow*, and "sky" is used to denote *blue*. A congruent trial would then be when the word "lemon" appears in yellow and "sky" in blue, whereas an incongruent trial would be when the word "lemon" appears in blue, and "sky" in yellow. Target-related responses are less efficient on the incongruent compared with the congruent trials (e.g., Kinoshita, Mills, & Norris, 2018). Although not specifically dubbed semantic Stroop, similar links between colours and concrete words were first reported in Klein (1964), and have also been observed elsewhere (Naor-Raz, Tarr, & Kersten, 2003; Yee, Ahmed, & Thompson-Schill, 2012). However, it is possible that such colour-linked effects are limited to concrete words, such as words referring to objects for which there is a prototypical colour, and may not extend to abstract words. Therefore, the association between colour and *abstract* words was tested here. Second, there is evidence that stimulus lightness is related to abstract concepts, namely valence, such that white stimuli are associated with positive-valence (Lakens, Semin, & Foroni, 2012; Meier, Robinson, & Clore, 2004). Do these associations with abstract concepts extend beyond lightness and into *chromatic (colour) variation*, such that they affect behaviour? This was the question addressed here.

The third piece of evidence to suggest that humans may encode abstract word-colour associations which manifest on objective task performance is our previous work, which showed that synaesthetes (i.e., individuals who have sensory associations manifesting as atypical perceptual experiences) and non-synaesthetes alike report explicit subjective associations between abstract words and colours (Goodhew & Kidd, 2017). We also found that these could be explained by language use statistics. That is, consensus associations between words and colours (e.g., joy-yellow and sorrow-blue) were reflected in systematic co-occurrence between these words (e.g., between word *joy* and word *yellow*), as quantified by latent semantic analysis (LSA) (Goodhew & Kidd, 2017). However, Goodhew and Kidd (2017) relied on subjective reports of associations provided by participants. While it has been established that concurrent-inducer pairings influence behaviour for synaesthetes (Dixon, Smilek, Cudahy, & Merikle, 2000; Mattingley, Rich, Yelland, & Bradshaw, 2001), the goal of the present study was to determine whether these subjectively reported associations also influence objective task performance in non-synaesthetes, even when they were not relevant to the task. Here, we tested this possibility across three experiments.

## 2. Experiment 1

In Experiment 1, we tested whether there were implicit associations between words and colours that would influence objective task performance even when they were not task-relevant. To do this, we employed an interference paradigm, in which participants had to respond to target words as quickly and accurately as possible, and the words appeared either in colours congruent or incongruent with reported word-colour associations. For example, the congruent colour for *joy* was *yellow*, and in the incongruent colour *black*, whereas the congruent colour for *doom* was *black*, and the incongruent colour *yellow*. Across the three experiments the task was varied. In this experiment, participants' task was to identify the valence (positive/negative) of the word. The colour of the word was not relevant to the task. However, if the congruency manipulation systematic influences response efficiency, then this would provide evidence that word-colour associations involuntary influence objective task performance.

In Goodhew and Kidd (2017) participants reported associations for both concrete words (e.g., *star* = white) and abstract words (e.g., *positive* = green). Here we focussed on abstract words, because they do not have obvious prototypical object colours (e.g., stars appear white in the night sky, whereas there are not tangible "positives" to be green), and so therefore represent a more stringent test of word-colour associations. Moreover, some concrete word-colour associations affecting performance have already been established (Kinoshita et al., 2018; Klein, 1964). Therefore, testing whether *abstract* words are also subject to such associations was the novel question addressed here.

### 2.1. Method

#### 2.1.1. Participants

To determine the required sample size, a power analysis was conducted in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). To be able to detect an effect size $d_z$ of 0.5 (assuming a medium effect) with 80% power, for a two-tailed repeated-measures *t*-test, a sample size of 34 was required. Therefore, thirty-four participants were recruited with normal or corrected-to-normal vision between the ages and 18 and 40 to participate in the study. All participants in this and the following experiment provided written informed consent prior to participation. Participants' mean age was 20.88 years (SD = 2.29). Twenty-six of the participants identified as female, the remaining eight as male. Twenty-eight reported being right-handed, and six left-handed. Twenty-two of the participants reported English as their first language, however, all were studying in English and therefore considered fluent in English.

#### 2.1.2. Apparatus & stimuli

Stimuli were presented on a 21.5-in. iMac running at a refresh rate of 60 Hz. Fourteen abstract words were chosen from Goodhew and Kidd (2017), and the items in that study were originally sourced from our large-scale rating study on the association between words and space (Goodhew & Kidd, 2016). Of the fourteen words, seven had a positive valence (*bliss*, *cheerful*, *happy*, *victory*, *positive*, *genius*, *joy*), and seven had a negative valence (*underworld*, *sorrow*, *miserable*, *unhappy*, *negative*, *bleak*, *doom*). The most commonly-reported colour for non-synaesthetes in the Goodhew and Kidd (2017) study was chosen to represent congruent trials, and each word was paired with a word of the opposite valence, whose congruent colour represented that word's incongruent colour. For example, the congruent colour for *cheerful* was yellow, and the congruent colour for *sorrow* was blue. These were paired, such that the incongruent colour was cheerful was *blue*, and the incongruent colour for sorrow was *yellow* (see Table 1 for full list). On the congruent trials, the word was presented in its congruent colour, whereas on the incongruent trials, the word was presented in its incongruent colour. This ensured that each physical colour appeared equally often under congruent and incongruent trial conditions. Words were presented in size 36 Arial front on a white background, and positioned in the centre of the screen. The RGB values for the colours were: black [0 0 0]; white [255 255 255]; grey [128 128 128]; red [255 0 0]; green [0 128 0]; blue [0 0 255]; and yellow [230 210 0].

#### 2.1.3. Procedure

Prior to the experimental block, participants completed a fourteen-trial practice block which provided on-screen feedback on the accuracy of their response. Participants were required to make at least 12 correct responses on this block to progress to the experimental block (practice

**Table 1**
Each of the words, and their congruent and incongruent colours in Experiment 1 (and the Valence Identification block of Experiments 2 and 3).

| Positive-valence word | Congruent positive/ incongruent negative | Incongruent positive/congruent negative | Negative-valence word |
|---|---|---|---|
| Bliss | Blue | Black | Underworld |
| Cheerful | Yellow | Blue | Sorrow |
| Happy | Yellow | Grey | Miserable |
| Victory | Red | Blue | Unhappy |
| Positive | Green | Black | Negative |
| Genius | Green | Grey | Bleak |
| Joy | Yellow | Black | Doom |

repeated as required). In both the practice and main experimental block, on each trial, a black fixation cross was presented in the centre of the screen for 500 ms, followed by one of the target words, which was shown until the participant responded. Participants were instructed to press the *Z* key if the word had a positive valence and the */?* key (on a standard keyboard) if the word had a negative valence as quickly and accurately as possible. Participants were shown the list of target words and their categorisation as positive and negative prior to the experiment. After each response, the screen was blank for 1000 ms before the next trial began. There was a total of 112 experimental trials, of which 50% were congruent and 50% were incongruent, randomly intermixed. A participant-paced rest break was provided after the first 56 trials.

As the goal of this study was to identify automatic word-colour pairings in non-synaesthetes, we also sought to identify any potential synaesthetes in the sample. To this end, participants completed a paper-and-pencil questionnaire at the end of the experiment to identify the presence of any synesthetic tendencies. This was a conservative approach – generally the reliability of any participants' responses is a prerequisite to their classification as synesthetic (Eagleman, Kagan, Nelson, Sagaram, & Sarma, 2007). However, here, participants were asked to complete the screening questions at the beginning of the standardized computerized synaesthesia battery (Eagleman et al., 2007), which identifies individuals who subsequently undergo further testing to establish whether they have synaesthesia. Any participants who responded yes to any of these questions were flagged as potential synaesthetes. The questions, and responses to them, can be seen in Table 3.

### 2.2. Results & discussion

Raw data are available via OSF here: https://osf.io/hqsjz/. (Note that accuracy values in the file are represented as proportions (×100 to convert to percentages reported here), and RT values are reported in seconds (×1000 to convert to ms reported here). Accuracy and RT are reported for main analyses, RT only for item analysis). Participants' responses during the experimental block were screened for invalid responses, which were defined as trials on which: (1) a key other than one of the two designated response keys was pressed, (2) response time (RT) was quicker than 100 ms (considered too quick to reflect a valid response to the stimulus), or (3) response time exceeded 2.5 standard deviations above the participant's mean RT (considered too slow to be conforming to the instruction to respond as quickly and accurately as possible). For a discussion of RT outlier cut-offs, see Goodhew, Dawel and Edwards (in press). This led to the exclusion of an average across participants of 2.9% of trials (range = 0.9%–4.5%). The RT variables (Congruent and Incongruent) were screened for outliers (z-score > |3.29|), and none were found.

While correct RT (i.e., RT on trials where the response was correct) was the primary dependent variable, accuracy was also analysed to assess for compliance with task instructions and for any speed-accuracy trade-offs. A repeated-measures ANOVA with the single factor Congruency (Congruent, Incongruent) on accuracy revealed that responses were significantly more accurate for congruent trials (M = 97.7%) than incongruent trials (M = 95.2%), $F(1, 33) = 10.38$, $p = .003$, $\eta_p^2 = 0.239$. Furthermore, another ANOVA on RT revealed that responses were significantly quicker in the congruent (555.9 ms) versus the incongruent (571.8 ms) condition, $F(1, 33) = 13.58$, $p = .001$, $\eta_p^2 = 0.291$ (see Fig. 1). All RT variables were approximately normally distributed[1].

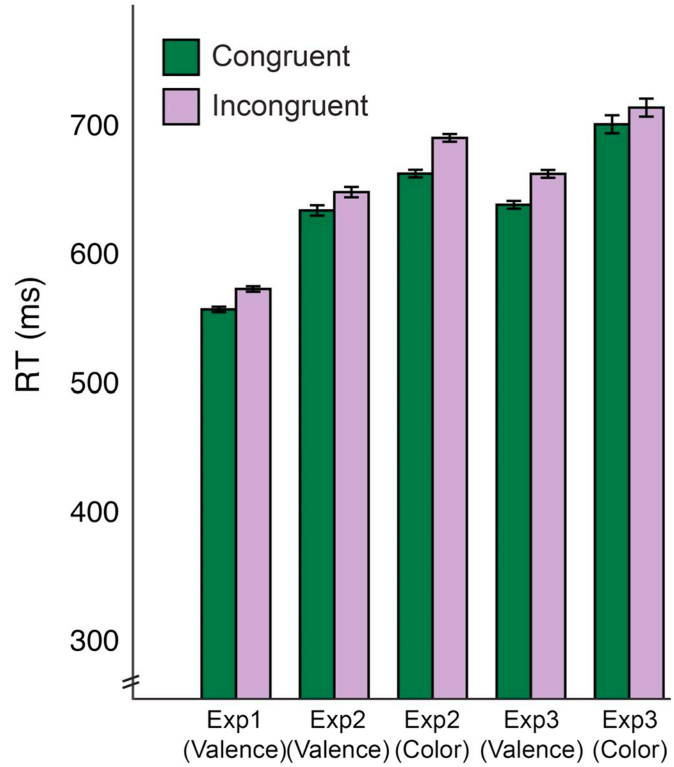


**Fig. 1.** An illustration of the RTs to identify the target (with respect to Valence or Colour) for Experiment 1, and both blocks of Experiment 2 and Experiment 3. Error-bars reflect standard-errors corrected for within-subjects designs (Cousineau, 2005).

Next, we conducted an item analysis to determine whether a subset, or all of the items equally, were driving the overall congruency effect. Two datasets did not have valid values for a given condition (one for *Bliss*, one for *Miserable*), and so these cases were removed for further analysis, leaving 32 cases remaining. We performed the item analysis in two different ways. The first was the standard psycholinguistic approach (Clark, 1973; Raaijmakers, Schrijnemakers, & Gremmen, 1999), whereby the mean for each congruency condition across participants for each item is treated as a random factor. These data were then be submitted to a one-way ANOVA with Congruency as a factor. This approach revealed a significant main effect of Congruency, $F(1, 13) = 6.09$, $p = .028$, $\eta_p^2 = 0.319$.

However, a limitation of this approach is that, while it reveals a robust effect of Congruency, it does not reveal whether there are also interactions between Item and Congruency that qualify this effect. Therefore, in a second analysis we included item as a fixed effect in a 14 (Item) × 2 (Congruency) repeated-measures ANOVA on RT. In this and all following analysis, the Greenhouse-Geisser correction for sphericity was used where Mauchly's test was significant ($p < .05$) for that effect if the variable had more than two levels. The main effect of Item was significant, $F(6.4, 197.7) = 4.15$, $p < .001$, $\eta_p^2 = 0.118$, as was the main effect of Congruency, $F(1, 31) = 16.65$, $p < .001$, $\eta_p^2 = 0.349$. Crucially, there was a significant interaction between Item and Congruency, $F(7.0, 215.7) = 2.93$, $p = .006$, $\eta_p^2 = 0.086$, indicating that the congruency effect changed as a function of Item. To better understand this pattern of results, the congruency effect (i.e., comparing RTs for Congruent versus Incongruent conditions) for all 14 items was computed, and this is displayed in Table 2.

Table 2 shows that, while 10/14 items showed congruency effects in the predicted direction, the size of the effect differed considerably, ranging from negligible to large effects. It is interesting that *Bleak* in particular produced such strong effects, since its congruent colour was grey and incongruent colour was green. Importantly, given that grey

[1] Non-parametric tests for all experiments showed convergent congruency effects as the ANOVA reported in the main text. That is, Wilcoxon-Signed Ranks tests comparing RTs for Congruent and Incongruent trials (averaged across Blocks in Experiments 2 and 3) was significant for all three experiments (ps < .012). This indicates that if any mild non-normality was present, it was not problematic, such that it did not distort the key result.

**Table 2**
Item analyses from Experiment 1. The colour word in brackets beside each item indicates the congruent colour for that item. The absolute congruency effect was computed as: Incongruent RT minus Congruent RT, such that a positive value indicates a speed advantage for the Congruent condition. Note that while repeated-measures *t*-tests in SPSS do not provide partial-eta squared values, a *t*-test is mathematically equivalent to a one-factor, two-level ANOVA. Therefore, these $\eta_p^2$ were obtained via an ANOVA for each item comparing the congruent and incongruent condition. * = $p < .05$, ** = $p < .004$ (correction for 14 comparisons). The final column, "Frequency of Identification" refers to the raw frequency of selection of the congruent colour as the colour associated with this concept word by non-synaesthetes in Goodhew and Kidd (2017). Bliss is 8 if include *pale blue* in addition to *blue* responses (otherwise 5), Victory is 10 if include *dark red* in addition to *red* (otherwise 9), Sorrow is 9 if include *pale blue* and *dark blue* in addition to *blue* responses (otherwise 7), Miserable is 11 if include *dark grey* in addition to *grey* (otherwise 9), and Unhappy is 14 if include *dark blue* in addition to *blue* (otherwise 10). Please see Goodhew and Kidd (2017) for more information. Finally, note that in all Valence Identification item analyses in the raw data file, the items are numbered consecutively according to their order in this table (e.g., Item 1 = Bliss, etc).

| Word | Absolute congruency effect (ms) | *p*-Value | Effect size ($\eta_p^2$) | Frequency of identification |
|---|---|---|---|---|
| Bliss (Blue) | +32.7 | 0.062 | 0.108 | 5/8 |
| Cheerful (Yellow) | −17.7 | 0.220 | 0.048 | 18 |
| Happy (Yellow) | −5.3 | 0.676 | 0.006 | 13 |
| Victory (Red) | −15.2 | 0.359 | 0.027 | 9/10 |
| Positive (Green) | +15.1 | 0.345 | 0.029 | 6 |
| Genius (Green) | +31.5 | 0.012* | 0.185 | 5 |
| Joy (Yellow) | +7.4 | 0.531 | 0.013 | 9 |
| Underworld (Black) | +21.2 | 0.111 | 0.080 | 11 |
| Sorrow (Blue) | +28.2 | 0.015* | 0.176 | 7/9 |
| Miserable (Grey) | +46.6 | 0.027* | 0.148 | 9/11 |
| Unhappy (Blue) | −17.2 | 0.201 | 0.052 | 10/14 |
| Negative (Black) | +54.8 | 0.003** | 0.248 | 11 |
| Bleak (Grey) | +47.0 | < 0.001** | 0.415 | 10 |
| Doom (Black) | +23.3 | 0.151 | 0.065 | 11 |

**Table 3**
Each of the questions that participants were asked in the synaesthesia screening questionnaire for both experiments, and the percentage of participants who responded "yes" to each question in Experiment 1, Experiment 2, and Experiment 3.

| Question | Percentage of participants who said "yes" in Exp 1/Exp 2/Exp 3 |
|---|---|
| Do numbers or letters cause you to have a colour experience? Example: Does the letter J "mean" purple to you? Or does "5" make you perceive purple? | 14.7%/20.6%/20.6% |
| Do weekdays and months have specific colours? Example: Does July always mean Navy blue to you? Is Wednesday always orange? | 20.6%/29.4%/17.6% |
| Do you imagine or visualize weekdays, months, and/or years as having a particular location in space around you? Example: is September always located two metres in front of you to the left? | 0%/20.6%/2.9% |
| Does hearing a sound make you perceive a colour? Example: does a shrill car horn cause you to see the colour green? Does the musical note C# make you see pink? | 8.8%/14.7%/23.5% |
| Do certain words trigger a taste in your mouth? Examples: Does the name 'Derek' taste like earwax? | 17.6%/20.6%/14.7% |
| Do you feel a sense of touch when you smell things? Example: Does the smell of coffee make you feel as though you are touching a cold glass surface | 23.5%/38.2%/17.6% |
| We have described a few types of synaesthesia. Many other unusual blendings of the senses have been reported. Do you suspect that you experience an unusual blending that other people do not have (other than the ones listed above)? These could include automatically hearing a sound when you see movement, or the sense of a shape being triggered by a taste, or experiencing a colour when feeling pain. | 8.8%/11.8%/38.2% |

consisted of RGB values of [128 128 128] and green of [0 128 0], this means that the grey could not be lower luminance than the green. This is notable because previous research has drawn associations between white/light and positive valence stimuli and black/dark and negative valence stimuli (Lakens et al., 2012; Meier et al., 2004). This is clearly not what is driving the effect here, and thus this congruency effect cannot be explained by a simple good-light bad-dark dichotomy.

These results indicate that response efficiency was impaired for incongruent trials for the sample as a whole. However, it is possible that there were some synaesthetes in the sample. The questions that were asked to identify potential synaesthetes, and the percentage of participants who responded "yes" to those questions is shown in Table 3.

For the next analysis, participants were classified as a Potential Synaesthete if they answered "yes" to any of the above questions, and were classified as a Non-synaesthete if they answered "no" to all of the above questions. This dichotomous Synaesthesia Screening variable (Non-synaesthete versus Potential Synaesthete) was then entered as a between-subjects variable into the main (non-item) ANOVA. The main effect of Congruency remained significant ($p = .001$), and the Synaesthesia Screening variable did not significantly interact at all with Congruency, $F(1, 32) = 0.02$, $p = .880$, $\eta_p^2 = 0.001$. Moreover, even when the Congruency (Congruent, Incongruent) ANOVAs on accuracy and RT were conducted exclusively on those 19 participants who were definite non-synaesthetes, accuracy was still significantly greater for the congruent (M = 98.5%) versus the incongruent trials (M = 96.3%), $F(1, 18) = 5.18$, $p = .035$, $\eta_p^2 = 0.224$, and RTs were also significantly faster for congruent (M = 551.3 ms) than for incongruent trials (M = 566.6 ms), $F(1, 18) = 9.96$, $p = .005$, $\eta_p^2 = .356$[2].

Altogether, the evidence from this experiment indicates that synaesthetes and non-synaesthetes alike automatically associate abstract words and colours to the extent that it influences their objective behaviour, even when the colour of the word or the association is not task relevant. However, reliable effects were limited to a small number of the total items. While it has been previously shown that systematic

[2] Item analyses could also be performed as a function of Synaesthetic Status, such that item analyses are provided separately for the two groups. However, since both previous studies (Goodhew & Kidd, 2017), and the results in all three experiments here indicated equivalent patterns of results for both non-synaesthetes and potential synaesthetes, we did not see sufficient justification to perform item analyses as a function of this variable.

subjective reports of associations between exist between abstract words and colours for non-synesthetes (Goodhew & Kidd, 2017), this is the first demonstration of such effects on objective task performance for a subset of items. Converging evidence for this was sought in Experiment 2, where are different task was also used in addition to valence identification.

## 3. Experiment 2

The purpose of Experiment 2 was two-fold. First, we sought to determine whether abstract word-colour associations would influence objective task performance when the task was to identify the physical colour of the word, and this physical colour could be either congruent (such that it was the most commonly reported colour for the word), or incongruent (such that it was not the most commonly reported colour for that word). Note that such a task may not promote the same depth of processing as valence-identification. Second, we sought to replicate the effect observed in Experiment 1 by repeating the same task.

### 3.1. Method

#### 3.1.1. Participants

Thirty-four participants[3] were recruited with normal or corrected-to-normal vision between the ages and 18 and 40 to participate in the study. All provided written informed consent prior to participation. Participants' mean age was 21.71 years (SD = 2.84), and 19 identified as female, the 15 as male, 32 reported being right-handed, and 2 left-handed. Twelve of the participants reported English as their first language, however, all were studying in English and therefore considered fluent in English.

#### 3.1.2. Apparatus, stimuli, & procedure

Participants now completed two experimental blocks (order counterbalanced across participants). The Valence Identification block was identical to Experiment 1. In the Colour Identification block, the stimuli were four positive words: *Bliss*, *Happy*, *Victory*, and *Genius.* The words could appear either in congruent colours (blue, yellow, red, and green, respectively), or in incongruent colours (i.e., would appear equi-probably with each of the three non-congruent colours, e.g., for *Bliss*, the incongruent colours were yellow, red, and green). Participants' task was to identify the physical colour of the word as quickly and accurately as possible, by pressing one of the four designated response keys to indicate *red*, *blue*, *yellow*, or *green*. Response keys were the *Z*, *X*, . > and /. The Colour Identification block had the same trial structure as Experiment 1. There were 120 trials in this block, of which 50% were congruent and 50% were incongruent, with trial types randomly intermixed. As per the Valence Identification block, the Colour Identification block had a rest break after 50% of trials. Each block had its own corresponding practice block which was completed prior to commencing that experimental block. Participants completed the screening questions at the end of the experiment (for responses to the questions, please see Table 3).

### 3.2. Results & discussion

Raw data for this and the following experiment are also available via the above OSF link. (Note that in the raw data files, Block A = Valence Identification Block, and Block B = Colour Identification Block). Participants' responses were screened for invalid responses in each experimental block separately, where invalid responses were defined in the same way as Experiment 1. This led to the exclusion of an average across participants of 3.2% of trials in the Valence Identification block (range = 0.7%–6.3%), and 3.0% of trials in the Colour Identification Block (range = 0.8%–5%). Accuracy was high across all conditions (Ms > 96%). One case had low accuracy, however, the results of the following ANOVA on RT were unchanged by the inclusion of this case, and so this case was retained. The RT variables were screened for outliers (z-score > |3.29|), and none were detected. All RT variables were approximately normally distributed.

The data were submitted to a 2 (Task: Valence versus Colour Identification) × 2 (Congruency: Congruent versus Incongruent) repeated-measures ANOVA on accuracy and RT. For accuracy, there were no reliable effects (*ps* ≥0.191). For RT, this analysis revealed a significant main effect of Task, $F(1, 33) = 4.64$, $p = .039$, $\eta_p^2 = 0.123$, such that responses were on average quicker for Valence Identification than for Colour Identification. There was also a significant main effect of Congruency, $F(1, 33) = 14.54$, $p = .001$, $\eta_p^2 = 0.306$, whereby responses were quicker for Congruent versus Incongruent trials. There was no reliable interaction between Congruency and Task, $F(1, 33) = 2.26$, $p = .142$, $\eta_p^2 = 0.064$. This means that the congruency effect was present both when participants were identifying the valence of the words, and when they were identifying the physical colour in which words appeared (see Fig. 1).

Furthermore, to determine whether the order (counterbalanced across participants) in which the tasks (Valence Identification versus Colour Identification) were completed affected the results, we also ran another analysis where we added Order as a dichotomous between-subjects variable to the ANOVA. This revealed that Order did not interact with either Task ($p = .214$) or Congruency ($p = .477$), and did not produce a three-way interaction among Task, Congruency, and Order ($p = .162$).

Following this, we conducted an item analysis. All cases had valid data in all conditions and so were retained for analysis ($N = 34$). There were different items in each of the two blocks, and therefore the item analyses were done separately for each block. We also used the same two approaches here for the Valence Identification Block as for Experiment 1. However, given the small number of items in the Colour Identification block ($N = 4$), we only performed the second analysis approach on this block.

For the Valence Identification Block, when each item was treated as a random effect, the main effect of Congruency was not significant, $F(1, 13) = 3.30$, $p = .093$, $\eta_p^2 = 0.202$. However, since the effect size suggests that the variable of Congruency explained 20% of the variance, the analysis was likely underpowered. In the second approach both Items and Congruency were included as fixed effects. Here, a 14 (Item) × 2 (Congruency) ANOVA indicated a significant main effect of Item, $F(6.5, 214.3) = 7.90$, $p < .001$, $\eta_p^2 = 0.193$, but not of Congruency, $F(1, 33) = 3.09$, $p = .088$, $\eta_p^2 = 0.086$. Moreover, the interaction between Item and Congruency did not reach significance, $F(6.4, 212.1) = 1.94$, $p = .071$, $\eta_p^2 = 0.055$. This suggests that the effect of Congruency did not differ across items in the Valence Identification Block.

For the Colour Identification Block, one case did not have valid values in one or more conditions and was therefore excluded. The remaining 33 datasets were submitted to a 4 (Item) × 2 (Congruency) repeated-measures ANOVA on RT. This revealed a non-significant main effect of Item, $F(3, 96) = 2.15$, $p = .099$, $\eta_p^2 = 0.063$, but a significant main effect of Congruency, $F(1, 32) = 23.08$, $p < .001$, $\eta_p^2 = 0.419$. Critically, there was also a significant interaction between Item and Congruency, $F(3, 96) = 11.58$, $p < .001$, $\eta_p^2 = 0.266$. This suggests that the effect of Congruency did differ across items in the Colour Identification Block. Therefore, the results for individual items are reported in Table 4.

As in Experiment 1, we see variable sized congruency effects across the four items. While 3/4 were in the predicted direction, the effect

[3] We initially planned to perform separate repeated-measures *t*-tests on the Valence Identification and Colour Identification block, and hence why we recruited the sample size indicated by the power analysis in Experiment 1. However, at the suggestion of a reviewer, we instead analysed Experiments 2 and 3 in a 2 × 2 RM factorial ANOVA.

**Table 4**
Item analyses from the Colour Identification block of Experiment 2. The colour word in brackets beside each item indicates the congruent colour. The absolute congruency effect was computed as: Incongruent RT minus Congruent RT, such that a positive value indicates an advantage for the Congruent condition. Note that while repeated-measures *t*-tests in SPSS do not provide partial-eta squared values, a *t*-test is mathematically equivalent to a one-factor, two-level ANOVA. Therefore, these $\eta_p^2$ were obtained via an ANOVA for each item comparing the congruent and incongruent condition. * = $p < .05$, ** = $p < .013$ (correction for 4 comparisons). The final column, "Frequency of Identification" refers to the raw frequency of selection of the congruent colour as the colour associated with this concept word by non-synaesthetes in Goodhew and Kidd (2017). Bliss is 8 if include *pale blue* in addition to *blue* responses (otherwise 5), and Victory is 10 if include *dark red* in addition to *red* (otherwise 9). Please see Goodhew and Kidd (2017) for more information.

| Word | Absolute congruency effect (ms) | *p*-Value | Effect size ($\eta_p^2$) | Frequency of identification |
|---|---|---|---|---|
| Bliss (Blue) | −34.1 | 0.025* | 0.147 | 5/8 |
| Happy (Yellow) | +28.3 | 0.089 | 0.087 | 13 |
| Victory (Red) | +105.7 | < 0.001** | 0.559 | 9/10 |
| Genius (Green) | +10.4 | 0.528 | 0.013 | 5 |

sizes ranged considerably. Altogether, the results of Experiment 2 demonstrate that subjectively reported word-colour associations can influence objective task performance. This was evidenced by the fact that participants' response efficiency to identify the valence, or the colour of the word, was systematically affected by the relationship between the abstract words and the physical colour in which they appeared. This relationship was such that for some abstract words, responses were facilitated when the colour matched the most commonly selected colour for the words compared with when they mismatched.

Finally, the main results were not qualitatively changed by the inclusion of the Synaesthesia Screening variable, and this variable did not reliably interact with any of these effects. That is, the interaction between Synaesthesia Screening and Task was not reliable ($p = .059$), and neither was the interaction between Synaesthesia Screening and Congruency ($p = .572$), nor the three-way interaction among Synaesthesia Screening, Congruency, and task ($p = .312$). This indicates that the effect of Congruency on RT was shared among participants irrespective of the results of the screening.

In summary, Experiment 2 provided evidence that word-colour pairings influenced objective behaviour when participants' task was to identify the colour of the word, in addition to the valence of the word. However, in the Colour Identification Block in Experiment 2, the incongruent colour was randomly selected on each trial from the three non-congruent colours. Since the congruent and incongruent trials were equi-probable (i.e., both 50%), this means that the congruent word-colour pairings (e.g., *Bliss* = blue, 50%) would appear more frequently during the experiment than any particular incongruent word-colour pairing (e.g., *Bliss* = red = 50%/3 = 17%). It is possible, therefore, that participants may have been sensitive to these contingencies, and the faster responding for the congruent pairings could have reflected their greater frequency, rather than their word-colour congruency per se. Consistent with this interpretation, the congruency effect for the Colour Identification task in Experiment 2 was significant in the final one-third of trials (50.2 ms, $p < .001$), whereas it was not in the first one-third of trial (4.3 ms, $p = .713$), suggestive of learning across the experiment contributing to the effect. However, this cannot explain why the effect was most pronounced for *Victory*.

## 4. Experiment 3

To rectify the potential learning effect for the colour identification task of Experiment 2, in Experiment 3, the frequency of the congruent and incongruent word-colour pairings was matched. To achieve this, only a single incongruent pairing was selected and used for the whole experiment (as per Experiment 1). Furthermore, as per the previous experiments, each physical colour was represented equally often under the congruent and incongruent conditions, and so differences in responsiveness to particular colours could not account for the results. The Valence Identification block was identical to the previous two experiments.

### 4.1. Method

#### 4.1.1. Participants

Thirty-four participants were recruited with normal or corrected-to-normal vision between the ages and 18 and 40 to participate in the study. All provided written informed consent prior to participation. Participants' mean age was 21.29 years (SD = 3.91), and 22 identified as female, the 12 as male, one reported being left-handed, one ambidextrous, and the rest right-handed. Fifteen of the participants reported English as their first language, however, all were studying in English and therefore considered fluent in English.

#### 4.1.2. Apparatus, stimuli, & procedure

The apparatus, stimuli, and procedure were identical to Experiment 2, except for the following changes to the Colour Identification Block. In the Colour Identification block, the stimuli were five words: *Bliss*, *Happy*, *Victory*, *Genius*, and *Unhappy*. The word *Unhappy* was added to determine whether the effect generalized beyond positive words. The words could appear either in congruent colours or in incongruent colours (the congruent-incongruent mappings are shown in Table 5). Responses to the screening questionnaire are shown in Table 3 (alongside those from Experiments 1 and 2).

### 4.2. Results & discussion

The raw data are available via the above OSF link. (Note that in the variable names, VIB = Valence Identification Block, and CIB = Colour Identification Block). Participants' responses during the experimental block were screened for invalid responses, which were defined in the same way as Experiments 1 and 2. This led to the exclusion of an average across participants of 2.8% of trials in the Valence Identification block (range = 0.0%–4.5%), and 2.6% of trials in the Colour Identification Block (range = 0.8%–5%). Mean accuracy was high (≥ 95%). The RT variables were screened for outliers (z-score > |3.29|), and one case was detected. However, the results of the 2 × 2 ANOVA were not qualitatively altered by the inclusion/exclusion of this case, and so they were retained (see below). There was some potentially problematic skew and kurtosis on the full data. However, with the exclusion of the outlier, all RT variables were approximately normally distributed, and importantly, the same pattern of results was obtained regardless.

The data were then submitted to a 2 (Task: Valence Identification versus Colour Identification) × 2 (Congruency: Congruent versus Incongruent) repeated-measures ANOVA. For accuracy, there was a main effect of Congruency, $F(1, 33) = 10.82$, $p = .002$, $\eta_p^2 = 0.247$, such that accuracy was greater in the Congruent than the Incongruent condition. No other effects were reliable ($Fs < 1$). For RT, there was a

**Table 5**
Each of the words, and their congruent and incongruent colours in the Colour Identification Block of Experiment 3.

| Target word | Congruent | Incongruent |
|---|---|---|
| Bliss | Blue | Green |
| Unhappy | Blue | Red |
| Happy | Yellow | Blue |
| Victory | Red | Blue |
| Genius | Green | Yellow |

significant main effect of Task, $F(1, 33) = 5.67$, $p = .023$, $\eta_p^2 = 0.147$, such that responses were quicker for Valence Identification than Colour Identification. There was also a main effect of Congruency, $F(1, 33) = 6.19$, $p = .018$, $\eta_p^2 = 0.158$, whereby responses were faster on Congruent versus Incongruent trials. In contrast, the interaction was not significant, $F(1, 33) = 0.50$, $p = .458$, $\eta_p^2 = 0.015$. This means that there was an equivalent benefit for Congruent trials across both tasks (see Fig. 1). With the RT outlier removed, there still a main effect of Task ($p = .006$), and crucially a main effect of Congruency ($p = .031$), and no interaction ($F < 1$, $p = .340$).

Moreover, to determine whether the order in which participants completed the tasks affected the results, we also ran another analysis where we added Order as a dichotomous between-subjects variable to the ANOVA with all of the cases included. Order did not interact with either the main effect of Task or Congruency, or with the interaction (all $Fs < 1$).

Additionally, we performed an item analysis. For the Valence Identification block, one case was excluded. In the Valence Identification block there was a significant main effect of Congruency when item was treated as a random factor, $F(1, 13) = 5.05$, $p = .043$, $\eta_p^2 = 0.280$. We again ran another analysis where we treated item as a fixed factor in a 2 (Valence) × 7 (Item) × 2 (Congruency) repeated-measures ANOVA on RT, which revealed a significant main effect of Item ($p = .001$), and of Congruency ($p < .001$), but not Valence ($p = .070$). None of the two-way interactions were reliable ($ps \geq 0.181$). However, there was a significant three-way interaction among Valence, Item, and Congruency, $F(3.6, 115.1) = 4.36$, $p = .004$, $\eta_p^2 = 0.120$. The effects for each item from this block can be seen in Table 6.

Table 6 revealed some differences across items in terms of the congruency effect that they generated in the Valence Identification Block. The majority of the items produced congruency effects in the expected direction, with some notable exceptions (e.g., *Happy*). *Miserable* produced the strongest congruency effect, followed by *Bleak* and *Victory*.

**Table 6**
Item analyses from the Valence Identification block of Experiment 3. The absolute congruency effect was computed as: Incongruent RT minus Congruent RT, such that a positive value indicates an advantage for the Congruent condition. Note that while repeated-measures *t*-tests in SPSS do not provide partial-eta squared values, a *t*-test is mathematically equivalent to a one-factor, two-level ANOVA. Therefore, these $\eta_p^2$ were obtained via an ANOVA for each item comparing the congruent and incongruent condition. * = $p < .05$, ** = $p < .004$ (correction for 14 comparisons). The final column, "Frequency of Identification" refers to the raw frequency of selection of the congruent colour as the colour associated with this concept word by non-synaesthetes in Goodhew and Kidd (2017). Bliss is 8 if include *pale blue* in addition to *blue* responses (otherwise 5), Victory is 10 if include *dark red* in addition to *red* (otherwise 9), Sorrow is 9 if include *pale blue* and *dark blue* in addition to *blue* responses (otherwise 7), Miserable is 11 if include *dark grey* in addition to *grey* (otherwise 9), and Unhappy is 14 if include *dark blue* in addition to *blue* (otherwise 10). Please see Goodhew and Kidd (2017) for more information.

| Word | Absolute congruency effect (ms) | *p*-Value | Effect size ($\eta_p^2$) | Frequency of identification |
|---|---|---|---|---|
| Bliss (Blue) | +39.7 | 0.139 | 0.067 | 5/8 |
| Cheerful (Yellow) | +13.2 | 0.617 | 0.008 | 18 |
| Happy (Yellow) | −51.6 | 0.012* | 0.180 | 13 |
| Victory (Red) | +69.2 | 0.014* | 0.174 | 9/10 |
| Positive (Green) | +45.2 | 0.148 | 0.064 | 6 |
| Genius (Green) | +56.3 | 0.152 | 0.063 | 5 |
| Joy (Yellow) | +4.3 | 0.828 | 0.001 | 9 |
| Underworld (Black) | +11.7 | 0.454 | 0.018 | 11 |
| Sorrow (Blue) | −22.4 | 0.465 | 0.017 | 7/9 |
| Miserable (Grey) | +106.1 | 0.006* | 0.210 | 9/11 |
| Unhappy (Blue) | −26.9 | 0.324 | 0.030 | 10/14 |
| Negative (Black) | +8.7 | 0.710 | 0.004 | 11 |
| Bleak (Grey) | +84.2 | 0.001** | 0.314 | 10 |
| Doom (Black) | +25.7 | 0.251 | 0.041 | 11 |

**Table 7**
Item analyses from the Colour Identification block of Experiment 3. The absolute congruency effect was computed as: Incongruent RT minus Congruent RT, such that a positive value indicates an advantage for the Congruent condition. Note that while repeated-measures *t*-tests in SPSS do not provide partial-eta squared values, a *t*-test is mathematically equivalent to a one-factor, two-level ANOVA. Therefore, these $\eta_p^2$ were obtained via an ANOVA for each item comparing the congruent and incongruent condition. * = $p < .05$, ** = $p < .010$ (correction for 5 comparisons). The final column, "Frequency of Identification" refers to the raw frequency of selection of the congruent colour as the colour associated with this concept word by non-synaesthetes in Goodhew and Kidd (2017). Bliss is 8 if include *pale blue* in addition to *blue* responses (otherwise 5), Victory is 10 if include *dark red* in addition to *red* (otherwise 9), and Unhappy is 14 if include *dark blue* in addition to *blue* (otherwise 10). Please see Goodhew and Kidd (2017) for more information.

| Word | Absolute congruency effect (ms) | *p*-Value | Effect size ($\eta_p^2$) | Frequency of identification |
|---|---|---|---|---|
| Bliss (Blue) | +82.5 | 0.003** | 0.238 | 5/8 |
| Happy (Yellow) | −27.6 | 0.252 | 0.040 | 13 |
| Victory (Red) | +60.1 | 0.108 | 0.076 | 9/10 |
| Genius (Green) | −13.4 | 0.675 | 0.005 | 5 |
| Unhappy (Blue) | −33.7 | 0.213 | 0.047 | 10/14 |

As in Experiment 2, we did not run an item analysis that treated item as a random factor for the Colour Identification block, since there were only five items. However, a 5 (Item) × 2 (Congruency) repeated-measures ANOVA indicated no main effects ($ps \geq 0.365$), but revealed a significant interaction between Item and Congruency, $F(4, 132) = 4.13$, $p = .003$, $\eta_p^2 = 0.111$. This suggests that the effect of congruency varied across the individual abstract words. The effects for each item from this block can be seen in Table 7.

Table 7 illustrates the congruency effects for the five items in the Colour Identification block. Once again, there was considerable variability across the items. The strongest congruency effect was for *Bliss* in the expected direction (i.e., Bliss = blue).

Finally, we assessed whether this pattern of performance was influenced by participants' synesthetic status. To do this, we included the between-subjects variable of Synesthetic Screening into the main ANOVA. The significant main effect of Congruency remained, $F(1, 32) = 6.41$, $p = .016$, $\eta_p^2 = 0.167$, and did not interact with the Synesthetic Screening ($F < 1$). Synaesthetic Screening also did not interact with any of the other effects ($Fs < 1$).

Altogether, Experiment 3 yielded a clear replication of the congruency effect on performance when participants' task was to identify the *valence* or the *colour* of the word, in particular for the items *Bleak* (grey) and *Bliss* (blue) respectively. These effects were underscored by variable congruency effects across items.

## 5. General discussion

Previous research indicates that synaesthetes and non-synaesthetes alike report systematic associations between abstract words and colours (Goodhew & Kidd, 2017). In the present paper we reported evidence that these pairings automatically affected behaviour: response speed to respond to the abstract words was impaired when the words appeared in incongruent versus congruent colours. This was true both for the sample as a whole, including for those individuals who did not report any possibility of synesthetic experience. Altogether, this tells us that humans have a general tendency to systematically associate abstract words with colours in a way that manifests in objective behaviour. This could be considered to be a "conceptual Stroop", distinct from the existing *semantic* Stroop because it occurs for abstract words.

It is noteworthy that not all of the words were equal in their ability to produce a congruency effect when the task was to identify the valence of the word. One of the reliably effective items was *Bleak* (grey). Although *Bleak* is technically an abstract word, it is often used to describe weather (e.g. a bleak day, or a bleak sky). It may be that adjectives such as *bleak* are most consistently and reliably associated with colours when they refer to a prototypical object with a canonical colour. That is, abstract words may be optimally linked to colour when firmly and consistently grounded in perceptual experience. That said, however, other words were also able to produce congruency effects, such as *Negative* (black) and *Bliss* (blue), for which there is no such obvious concrete referent, meaning that abstract words can be linked to colours in their own right. This means that at least for some abstract words, humans have systematic word-colour associations which influence task performance.

Our item-analyses suggested that individual words likely vary in the degree to which they are linked with colours. Controlling for item-level effects is standard in psycholinguistic research (beginning with Clark, 1973), acknowledging the fact that individual words vary on a number of properties that affect their representation and use. It is interesting that in related literatures, such as conceptual cueing (i.e., concept-space associations), such item analyses are not typically performed (though see Goodhew et al., 2014). Paying closer attention to item-level nature of effects has the potential to provide more fine-grained explanations of effects like these considered here, since knowing the conditions under which an effect is or is not present narrows the hypothesis space of possible mechanisms underlying the effect. Notably, in our data the result invites the conclusion that concept-colour association are present and automatic in human cognition, but that they cannot be assumed to be a property of all concepts. Rather, they likely emerge from the frequent co-occurrence of concept-colour pairings.

Furthermore, we selected the most *commonly* reported subjective colour for each word reported in Goodhew and Kidd (2017) to use in the present study. This means that the congruent colour that the item appeared in did not necessarily correspond to the colour that every participant would associate with this word. In this light, it is striking that these most-popular associations were able to produce reliable effects for some words at the group level, and this provides a clear 'proof of concept' that concept-colour associations influence objective behaviour. However, it may be that the words that produced the most reliable congruency effects are those for which the largest number of people have a consistent colour association, and concept-colour associations may produce stronger and more consistent effects when matched to individual participants' reported associations.

It is important to note that the results from the Colour Identification block of Experiment 2 are susceptible to the explanation that participants merely learned (we are agnostic regarding whether this learning is implicit or explicit) that the congruent stimuli were more frequent than the incongruent stimuli, and that this therefore facilitated their responses to the congruent stimuli. It would be interesting in future research to test whether participants are more sensitive to such contingencies when the mappings adhere to those considered *congruent* (versus incongruent) in the current framework. That is, if participants more readily learn that *Bliss* in blue is a frequent pairing than they would learn than *Bliss* in red is a frequent pairing, then this would provide converging evidence for behavioural manifestations of abstract word-colour pairings.

We think it would be valuable for future research to consider the role of individual differences in the effects observed here (see Goodhew & Edwards, 2019). This is because there are large and meaningful individual differences in language (Kidd, Donnelly, & Christiansen, 2018), as well as individual differences in the extent to which individuals attend to or perceive colour (Hsu, Kraemer, Oliver, Schlichting, & Thompson-Schill, 2011), which likely extends to word-associations. Consistent with interpretation, prior work has found reliable individual differences in the extent to which colour information affects performance in standard Stroop as well as semantic concrete word colour-priming tasks (e.g., "emerald" priming "cucumber") (Yee et al., 2012). We did not see evidence that the synesthetic screening tool reliably differentiated individuals who produced high versus low magnitude effects in response to the experimental manipulation of congruency. However, it is entirely possible that other, as yet unmeasured individual-differences variables could do so. For example, the magnitude of this effect could depend on an individual's vividness of mental imagery (see Cui, Jeter, Yang, Montague, & Eagleman, 2007), such that individuals with more vivid mental imagery produce larger, more robust effects. Conversely, since processing the meaning of the word is task-irrelevant in this block, it could be that individuals higher in attentional control produce lower congruency effects when the task is merely to name the physical colour of the word (see Derryberry & Reed, 2002; Ólafsson et al., 2011). Furthermore, there may be interesting individual difference as a function of a person's cultural or linguistic background. For instance, there is evidence that Mandarin speakers associate the word *red* with the spatial direction *up*, which is the direction with which positive concepts are typically associated (Wu, Kidd, & Goodhew, 2019). Thus, *red* may be a more positive concept for Mandarin than English speakers, which suggests that Mandarin speakers may be more likely to associate the colour red with positive-valence words than English speakers. Ultimately, the role of individual differences is an interesting avenue for future research.

In conclusion, like synaesthetes, non-synesthetes also display automatic associations between some abstract words and colours, which systematically influence objective performance in laboratory tasks. This demonstrates the existence of a *conceptual* Stroop effect.

## CRediT authorship contribution statement

**Stephanie C. Goodhew**:Conceptualization, Methodology, Software, Data curation, Formal analysis, Writing - original draft.**Evan Kidd**:Methodology, Writing - review & editing.

## Declaration of competing interest

The authors have no conflicts of interest to declare. The first author has full control of all primary data, which is also available in the Open Science Framework (see link above). The Australian National University's Human Research Ethics Committee approved all aspects of the current research protocol, and all participants provided written informed consent prior to participation. All aspects were in compliance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## Acknowledgements

This research was supported by an Australian Research Council (ARC) Future Fellowship (FT170100021) awarded to S.C.G. We thank Rani Gupta and Nicholas Wyche for assistance with the data collection.

## References

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology, 1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042.

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research, 47*(4), 474–478. https://doi.org/10.1016/j.visres.2006.11.013.

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology, 111*(2), 225–236. https://doi.org/10.1037/0021-843X.111.2.225.

Dixon, M. J., Smilek, D., Cudahy, C., & Merikle, P. M. (2000). Five plus two equals yellow.

*Nature, 406*(6794), 365. https://doi.org/10.1038/35019148.

Dudschig, C., De la Vega, I., & Kaup, B. (2015). What's up? Emotion-specific activation of vertical space during language processing. *Acta Psychologica, 156*, 143–155. https://doi.org/10.1016/j.actpsy.2014.09.015.

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., & Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *Journal of Neuroscience Methods, 159*(1), 139–145. https://doi.org/10.1016/j.jneumeth.2006.07.012.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1), 1–12. https://doi.org/10.1167/10.1.6.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146.

Goodhew, S. C., Dawel, A., & Edwards, M. (2020). Standardizing measurement in psychological studies: On why one second has different value in a sprint versus a marathon. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01383-7 (in press).

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition, 69*, 14–25. https://doi.org/10.1016/j.concog.2019.01.008.

Goodhew, S. C., & Kidd, E. (2016). The conceptual cueing database: Rated items for the study of the interaction between language and attention. *Behavior Research Methods, 48*, 1004–1007. https://doi.org/10.3758/s13428-015-0625-9.

Goodhew, S. C., & Kidd, E. (2017). Language use statistics and prototypical grapheme colours predict synaesthetes' and non-synaesthetes' word-colour associations. *Acta Psychologica, 173*, 73–86. https://doi.org/10.1016/j.actpsy.2016.12.008.

Goodhew, S. C., McGaw, B., & Kidd, E. (2014). Why *is* the sunny side always up? Explaining the spatial mapping of concepts by language use. *Psychonomic Bulletin & Review, 21*(5), 1287–1293. https://doi.org/10.3758/s13423-014-0593-6.

Gozli, D. G., Chasteen, A. L., & Pratt, J. (2013). The cost and benefit of implicit conceptual cues for visual attention. *Journal of Experimental Psychology: General, 142*(4), 1028–1046. https://doi.org/10.1037/a0030362.

Gozli, D. G., Chow, A., Chasteen, A. L., & Pratt, J. (2013). Valence and vertical space: Saccade trajectory deviations reveal metaphorical spatial activation. *Visual Cognition, 21*(5), 628–646. https://doi.org/10.1080/13506285.2013.815680.

Hsu, N. S., Kraemer, D. J., Oliver, R. T., Schlichting, M. L., & Thompson-Schill, S. L. (2011). Color, context, and cognitive style: Variations in color knowledge retrieval as a function of task and subject variables. *Journal of Cognitive Neuroscience, 23*(9), 2544–2557. https://doi.org/10.1162/jocn.2011.21619.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences, 22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006.

Kinoshita, S., Mills, L., & Norris, D. (2018). The semantic Stroop effect is controlled by endogenous attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(11), 1730–1742. https://doi.org/10.1037/xlm0000552.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *The American Journal of Psychology, 77*(4), 576–588. https://doi.org/10.2307/1420768.

Knöferle, K., & Spence, C. (2012). Crossmodal correspondences between sounds and tastes. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-012-0321-z.

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.

Mattingley, J. B., Rich, A. N., Yelland, G., & Bradshaw, J. L. (2001). Unconscious priming eliminates automatic binding of colour and alphanumeric form in synaesthesia. *Nature, 410*(6828), 580–582. https://doi.org/10.1038/35069062.

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.

Naor-Raz, G., Tarr, M. J., & Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception, 32*(6), 667–680. https://doi.org/10.1068/p5050.

Ólafsson, R. P., Smári, J., Guðmundsdóttir, F., Ólafsdóttir, G., Harðardóttir, H. L., & Einarsson, S. M. (2011). Self reported attentional control with the Attentional Control Scale: Factor structure and relationship with symptoms of anxiety and depression. *Journal of Anxiety Disorders, 25*(6), 777–782. https://doi.org/10.1016/j.janxdis.2011.03.013.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*(3), 416–426. https://doi.org/10.1006/jmla.1999.2650.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia–a window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3–34.

Wu, Q., Kidd, E., & Goodhew, S. C. (2019). The spatial mapping of concepts in English and Mandarin. *Journal of Cognitive Psychology, 31*(7), 703–724. https://doi.org/10.1080/20445911.2019.1663354.

Yee, E., Ahmed, S. Z., & Thompson-Schill, S. L. (2012). Colorless green ideas (can) prime furiously. *Psychological Science, 23*(4), 364–369. https://doi.org/10.1177/0956797611430691.