Running Head: SHARED FEATURES ACCOUNT

Feature Transformation: A Shared Feature Account

Sean Hughes, Jan De Houwer, Simone Mattevelli, & Ian Hussey

*Ghent University*

**Authors Note**

SH, JDH, SM, and IH, Department of Experimental Clinical and Health Psychology, Ghent University. This research was conducted with the support of Grant BOF16/MET\_V/002 to JDH. Correspondence concerning this article should be sent to sean.hughes@ugent.be.

**Abstract**

*Keywords*:

Feature Transformation: A Shared Feature Account

Humans possess an unparalleled ability to relate one object to another. Throughout our lives we constantly forge connections between stimuli based on our experiences (e.g., a painful shock from a faulty household appliance will lead people to a subsequently avoid that item in the future), observations (e.g., watching a person happily consume an unknown food will increase our own desire to try it), and the verbal information we receive (e.g., being told that ‘drinking bleach is dangerous’ allows us to connect bleach to a whole new set of consequences). *Relational learning* – the ability to recognize and respond to relationships among objects – is not only central to conditioning, but to many other phenomena such as reasoning, language, perspective-taking, empathy, and problem-solving.

In this paper we introduce a relatively simple but powerful new perspective on relational learning known as the *shared features account*. At its core, this account refers to the fact that, when two stimuli share a feature with one another, people often act as if those stimuli are related to one another, and that they share others features as well. This assumption - that objects which share one feature also share other features - may help us better understand different phenomena throughout psychological science.

For instance, take social psychology. Research on stereotypes and prejudice indicates that when two individuals share one feature, such as their skin color or gender, people often assume that those individuals also share other features (e.g., personality attributes or characteristics) (REF). The very same is true for person perception: research on spontaneous trait transference shows that when two individuals (Bob and Mike) share one feature such as their group affiliation (e.g., Democrats) or an attribute (e.g., generous), people will often assume that they share other features as well (e.g., if Bob is empathetic then so too is Mike) (REF). Now consider matching effects in persuasion. Matching refers to the finding that when there is some shared feature or ‘match’ (e.g., between the source and the message) this commonality influences the subsequent assumptions that the recipient makes (e.g., the message makes more sense or that the source is more trustworthy). These assumptions increase the chance that the message will persuade the recipient, relative to situations where a mismatch exists (REF).

Now take marketing psychology: if brand products share one feature such as their identity (e.g., a telephone and television are both Apple or Sony products) they are often assumed to share others features (e.g., be high in quality, status, or value) (REF). Likewise, fake or counterfeit brands intentionally imitate (share features such as) the physical properties of genuine brands in the hope this will influence assumptions about the fake brand itself (e.g., that it is also high in quality, worth purchasing) (REF). In moral psychology, when one person (John) is accountable for his past actions (e.g., membership of the Nazi party) and shares a feature with a second person (Tom) (e.g., John is the father of Tom), this shared feature influences the assumptions people make about the latter’s moral accountability (e.g., they assume that Tom is also morally responsible for his father’s actions) (REF).

Finally, shared features may play a role in learning psychology. In evaluative conditioning (EC) the fact that a neutral stimulus shares spatio-temporal properties with a valenced stimulus often leads the former to acquire properties of the latter (REF). This is also true for attribute conditioning: pairing an unknown person with a known athlete causes the former to acquire properties of the latter (e.g., be rated as more athletic or healthy: REF). Still other learning phenomena can also be conceived of in this way. For instance, when two operant contingencies share a common stimulus or response, this shared feature causes people to make assumptions about the properties of other stimuli and responses in those same contingencies (i.e., learning via ‘intersecting regularities’; see Hughes, Perugini, & De Houwer, 2016).

Thus, when we strip away the specific types of stimuli and responses involved, we see that many phenomena actually involve a similar situation. One where people make assumptions about the features of one object (e.g., whether a person, group, or brand product is positive or negative) based on the fact that it shares some other feature (e.g., physical characteristics, brand identity, or spatio-temporal properties) with a second object (another person, group, or brand product). Although stereotyping, prejudice, impression formation, conditioning, brand identity, and moral responsibility effects are all unique in the types of stimuli and responses under investigation, they often involve changes in behavior that are functionally rooted in the fact that two or more stimuli share features.

This realization suggests that there are remarkable similarities between different domains in psychological science, and that developments in one area may be informative for work in another. Yet even a cursory glance at the literature will show that the aforementioned phenomena are typically studied in isolation, with different researchers busy documenting the moderators and mediators of the behaviors they are interested in, and rarely interacting with their colleagues (or drawing on findings) located in other intellectual domains. We argue that this does not have to be the case. In what follows we unpack a new conceptual account that helps address this state of intellectual fragmentation and open up new avenues of study.

**The Shared Features Account**

The *shared features account* consists of three main components: (a) it refers to situations which involve ‘source objects’ and ‘target objects’, (b) these source and target objects each possess a range of ‘object features’, and (c) when a source and target object share a common feature with one another, this shared feature influences the types of assumptions people make about the target object’s features.

Let us briefly unpack these three components. First, we adopt a broad definition of ‘objects’ as any potential stimulus in the environment. Objects can refer to people, animals, inanimate items, and even responses. For our purposes, we are interested in two types of objects: *source* and *target objects*. In-line with De Houwer, Richetin, Hughes, and Perugini (*in prep*) we define a ‘source object’ as the object which possess the source feature whereas the ‘target object’ refers to the object which possesses the target feature. Second, we define ‘features’ as any assumed state of an object. These states can have multiple values and can relate to many different properties, from physical (e.g., height), and psychological (e.g., intelligence, valence), to behavioral (e.g., the way in which an object responds to its environment) (see De Houwer et al., *in prep*). In this paper we are interested in two types of features: *source features* and *target features*. [[1]](#footnote-1)

Target features are those features of an object about which assumptions are being made. Source features are those features of an object which give rise to assumptions about target features. The former is typically the dependent variable whereas the latter is the independent variable under investigation. The value of a source feature can be varied in order to investigate if this influences the corresponding value of the target feature. To illustrate, take the well-known halo effect (REF), wherein features of a source object (e.g., how attractive a person is) leads people to make assumptions about a target object feature (e.g., how intelligent that person is) (note: in this case the source and target objects are the same person). Here the value of a source feature (how attractive a person is) can be varied to investigate if this influences the value of a target feature (how intelligent the person is said to be).

Finally, when source features influence the assumptions made about target features, *feature transformation* is said to take place. [[2]](#footnote-2) The shared feature account is concerned with one particular type of feature transformation - namely - those situations where assumptions about target object features arise because the target object shares some other feature with a source object (see Figure 1).

**Source Object**

**Target Object**

Source Feature (*Valence*)

Target Feature (*Valence*)

**Shared Feature** (*Shape*)

*Figure 1*. Visual overview of the shared feature effect. Here assumptions about the target object’s feature (e.g., valence) are influenced by a source object’s features (e.g., valence). These assumptions are based on the fact that the former shares a different feature with the latter (e.g., the source and target object are of a similar shape).

**The Shared Features Account has Heuristic Value**

When we examine the social, marketing, moral, and learning psychology literatures there appears to be no common framework or language that allows researchers to conceptualize and describe the similarities or differences between the shared feature effects studied in these areas. Instead a wide variety of terms are deployed and this multiplicity of concepts can undermine our ability to identify what is genuinely similar or different within and between intellectual domains. For instance, researchers interested in evaluative and attribute conditioning refer to conditioned (CS) and unconditioned stimuli (US), whereas those interested in operant conditioning speak about discriminative stimuli (Sd) and reinforcers (Sr). Likewise, research on marketing, halo, moral accountability, and person perception employs few technical terms that allow us to identify commonalities or differences in the basic procedures or effects involved.

The shared features account provides a solution to this problem. It offers a parsimonious way to conceptualize and describe a wide variety of phenomena using a limited set of terms. Concepts such as source/target and object/feature can be used to describe shared feature effects that are typically studied (under different names) in different domains. For instance, instead of relying on different terms and concepts in impression formation research, we can refer to the fact that a source object (Bob) possesses a certain feature (e.g., is violent). When people then learn that a target object (Mike) shares some other feature with the source (e.g., they have physical characteristics in common), we can say that this shared feature leads people to make assumptions about the target object’s features (e.g., that Mike is also violent). The same concepts can be used in conditioning research as well. Here too people learn that a source object (unconditioned stimulus) possesses a certain feature (e.g., is valenced as in EC or athletic as in attribute conditioning). They then learn that a target object (CS) shares some other feature with the source (e.g., both are presented together in space and time). As a result of this shared feature, people make assumptions about the target object’s features (that the CS is also valenced or athletic). The very same concepts can be used when dealing with different types of learning, stereotyping, prejudice, branding, and so on.

Simply put, the shared features account has heuristic value. It allows researchers to conceptualize and speak about effects in ways that (a) enhance communication within and between intellectual domains, (b) prevent fragmentation, confusion, or conflict resulting from the use of multiple terms to describe the same underlying phenomenon, and (c) reveals similarities and differences between phenomena. While acknowledging important differences between domains, it argues that when many effects are distilled down to their essence, they can be conceptualized as involving four basic elements (source object, target object, source feature, and target feature), a situation wherein the source and target share one feature, and as a result, new assumptions are made about other target object features.

The shared features account also causes us to look at existing phenomena in entirely new ways. Take EC, for instance, which is typically defined as a change in liking due to stimulus pairings. Most researchers would argue that EC effects are driven by the fact that CS and US are presented in spatio-temporal contiguity. Yet our account takes a different perspective. It argues that EC effects may actually be due to the fact that the CS and US *share* a common feature with one another, and in most EC studies, this shared feature just so happens to be contiguity. If correct, then the crucial element in EC is the fact that stimuli share a feature and not the mere fact that they are paired in space and time. Note that this new way of thinking does not draw EC effects themselves into question – simply our prior explanation of those effects. Furthermore, it also highlights that pairings are just one way in which a shared feature can be installed. It also highlights that EC is just one instance of a much broader class of share features effects.

**The Current Research**

Before we can apply our account to existing phenomena we first need to ensure that the shared features effect can be experimentally established and manipulated. Towards this end, we carried out five studies that first generated the effect (Experiment 1), replicated it while testing for potential boundary conditions (Experiments 2-3), and then generalized it across different types of physical (Experiment 4) and symbolic features (Experiment 5). Although our account could be tested using a variety of features or procedures, we decided to initially focus on the domain of attitudes – that is – on one type of feature (valence) in one type of procedure (evaluative conditioning). Doing so provided a simple yet effective context in which to test our ideas. Recall that EC effects are traditionally assumed to be driven by mere contiguity. Our account suggests that it is the fact that stimuli share a feature, rather than contiguity *per se*, which is important. This idea can be tested by keeping contiguity constant while varying the extent to which stimuli overlap with respect to some other feature. If this latter feature moderates EC effects, then it would support the idea that shared features rather than mere contiguity drive those effects and offer initial support for our account. It would also have implications for those domains to which EC effects are relevant (e.g., social, brand, health, and clinical attitude formation and change, impression formation, implicit cognition). [[3]](#footnote-3)

With this in mind we carried out five studies. Each employed a broadly similar format which we will preview here. We first asked participants to completed an EC phase. During this phase a series of trials were presented wherein three stimuli simultaneously appeared onscreen: a positive US (source object), a negative US (source object), and a neutral CS (target object). We then manipulated the extent to which the CS shared a feature with a certain US. In Experiments 1-3 the shared feature was the color in which stimuli were presented: half of the trials presented CS1 in the same color as a positive US whereas the other half presented CS2 in the same color as negative images. In Experiment 4 the shared feature was the size of the stimuli: half of the trials presented CS1 in the same size as a positive US whereas the other half presented CS2 in the same size as negative images. Finally, in Experiment 5 the shared feature was symbolic in nature. We first established a symbolic relationship between two sets of colors (*Blue-Same-Yellow* and *Green-Same-Purple*) and then, during the EC phase, presented a CS in blue or green, along with a positive US in yellow and a negative US in purple. Following the EC phase, we assessed for attitude formation (CS evaluations) using self-report ratings and an Implicit Association Test (IAT). We added an IAT as it is assumed to reflect more automatic attitudes that can influence subsequent behavior in unique ways (REF). If EC effects are driven by simple spatio-temporal contiguity then we would expect to see similar and ambivalent evaluative responses towards both CSs (given that they were both repeatedly paired with positive *and* negative USs). Yet if those same effects are driven by the fact that the CS and US share a feature then we would expect to observe positive evaluations of one and negative of the other. If our account is correct EC effects should be moderated by a range of different physical and symbolic features that are shared by stimuli. [[4]](#footnote-4)

**Experiment 1**

**Method**

**Participants and design**. A total of 114 English-speaking volunteers (62 females) (*Mage* = 33.12, SD = 8.39) participated online via the Prolific Academic website (https://prolific.ac) in exchange for a monetary reward (€1.50). The experiment was programmed in Inquisit 4.0 and hosted via Inquisit Web (Millisecond Software, Seattle, WA). The experiment involved a single-factor between-subjects design (*Color Matching*: positive-color match vs. negative-color match), with self-reported ratings and IAT effects as the main dependent variables. Three method variables were manipulated between participants: *evaluative task order* (self-reports vs. IAT first), *IAT block order* (learning [EC] phase consistent vs. inconsistent first) and *stimulus assignment* (which stimulus appeared in the same color as positive or negative words). The sample size was determined prior to data collection. We stopped data-collection whenever 114 participants had completed all measures of the experiment to ensure that we would have sufficient statistical power to detect medium effects (necessary sample size = 110 to have power = 0.80 to find a medium effect of *d* = 0.50 at alpha = 0.95). Note that a similar analytic strategy was used in Experiments 1-5 and thus studies were powered accordingly. The study designs were pre-registered, and are available, along with the raw data, and analytic plans for this and all other experiments on the Open Science Framework website (https://osf.io/pqm9v/). We report all manipulations and measures used in the study. All data were collected without intermittent data analysis.

**Materials**

**Stimuli**. Two nonsense words (Morag and Struan) served as CS1 and CS2. Six positive (rainbow, pleasure, smile, love, paradise, joy) and six negative adjectives (war, cancer, hate, hell, misery, vomit) served as the USs.

**IAT**. Automatic evaluative responding was measured using an IAT. The two nonsense words served as one set of target stimuli and the words “Good” and “Bad” as another. Eight positively valenced and eight negatively valenced adjectives served as one set of attribute stimuli (fantastic, great, nice, good, pleasant, wonderful, amazing, happy versus terrible, disgusting, nasty, horrible, sick, awful, sad, unpleasant) and the two nonsense words served as the second set.

**Procedure**

Participants were first provided with a general overview of the experiment and then asked for their informed consent. The study consisted of three phases: EC, evaluative measures, and exploratory questions.

**EC**. Evaluative conditioning consisted of three blocks of 16 trials (48 total), with each block containing two types of trials: one trial in which CS1 was eventually presented in the same color as positive words, and another trial in which CS2 was eventually presented in the same color as negative words. Specifically, three stimuli simultaneously occurred onscreen during each trial: a neutral CS (either Morag [CS1] or Struan [CS2]) along with a positive and negatively valenced adjective. All three stimuli were initially presented in white against a black background for 3000ms. On certain trials, CS1 and positive USs both changed to the same color (e.g., blue) whereas the negative US changed to a different color (e.g., green). On other trials, CS2 and negative USs both changed to one color (e.g., yellow) whereas the positive US changed to another (e.g., purple). All stimuli remained onscreen for another 3000ms and were then removed, followed by an inter-trial interval of 1250ms, and the next trial. Stimulus color (i.e., blue, green, yellow and purple) was varied across trials, so that none of the colors could assume any specific positive or negative value (see Figure 1).

**Self-reported ratings**. Self-reported ratings were assessed using four different semantic differential scales. On each trial, one of the two CSs was presented and participants were asked to indicate their general impression of that stimulus using a scale ranging from -5 to +5 with 0 as a neutral point. The four end-points of the scales were as follows: *Negative-Positive*, *Pleasant-Unpleasant*, *Good-Bad*, *I Like It-I Don’t Like It*. A mean evaluative rating was calculated for each CS by averaging scores from these four scales.

**IAT**. We assessed whether the combination of color and stimulus pairings led to a change in automatic evaluative responding using the IAT. Participants were informed that a series of words would appear one-by-one in the middle of the screen and that their task was to categorize those items with their respective target (CS1 or CS2) or attribute categories (‘Good’ and ‘Bad’) as quickly and accurately as possible. They were told that the two words they had previously encountered (targets) as well as the words “Good” and “Bad” (attributes) would appear on the upper left and right sides of the screen and that stimuli could be assigned to these categories using either the left (‘E’) or right keys (‘I’). Each trial began with the presentation of a fixation cross for 200ms in the middle of the screen, followed immediately by a target or attribute stimulus. If the participant categorized the word correctly – by selecting the appropriate key for that block of trials – the stimulus disappeared from the screen and the next trial began. In contrast, an incorrect response resulted in the presentation of a red “X” which remained on-screen until the correct key was pressed. Overall, each participant completed seven blocks of trials. The first block of 20 practice trials required them to sort CS1 and CS2 into their respective categories, with CS1 assigned to the left (‘E’) key and CS2 to the right (‘I’) key. On the second block of 20 practice trials, participants assigned positive words to the “Good” category using the left key and negative words to the “Bad” category using the right key. Blocks 3 and 4 (20 and 40 trials, respectively) involved a combined assignment of target and attribute stimuli to their respective categories. Specifically, participants categorized CS1 and positive words using the left key and CS2 and negative words using the right key. The fifth block of 20 trials reversed the key assignments, with CS2 now assigned to the left key and CS1 to the right key. Finally, the sixth and seventh blocks (20 and 40 trials respectively) required participants to categorize CS1 with negative words and CS2 with positive words.

**Behavioral Intentions.** Participants are presented with two brand products labeled with either CS1 or CS2. They are asked to indicate which of these products they would be willing to try in a supermarket and given the following options: *I would try CS1*, *I would try CS2*, *I would try CS1 and CS2*, *I would try neither*, *I don’t know*.

**Exploratory Questions**. Participants were asked a series of exploratory questions about their *contiguity memory* (i.e., the relationship between the CS-USs), *color memory* (i.e., the color match between the CSs and USs), a *manipulation check* to ensure they did not write down the contingencies during the task, *demand compliance* and *reactance* questions about the rationale for their explicit and implicit evaluations, *hypothesis* and *influence awareness*. Details about each of these questions can be found in the Supplementary Materials section.

**Results**

**Analytic Strategy**

To determine whether self-reported ratings and IAT effects for the two nonsense words (dependent variables) differed as a function of color (match vs. mismatch) established between the nonsense and valenced words (independent variable), a series of ANOVAs and post-hoc *t*-tests were carried out on the rating, behavioral intentions, and IAT data. A similar analytic strategy was used in Experiment 2-5.

**Data Preparation**

**Exclusions**. We excluded data from eight participants who did not complete the entire session. Participants who had IAT error rates above 30% across the entire task or above 40% for any one of the four critical blocks, or who responded faster than 400ms on more than 10% of trials were excluded (*n* = 3). This led to a final sample of 103 participants.

**Self-reported ratings**. Self-reported ratings for CS1 and CS2 were first averaged and then a difference scored was calculated by subtracting scores for CS2 from CS1. Positive values indicate a relative preference for the stimulus that eventually shared a color with positive USs over the CS that shared a color with negative USs. Negative values indicate the opposite.

**IAT**. Following the recommendations of Greenwald, Nosek, and Banaji (2003), response latency data were prepared using the D4 scoring algorithm. The resulting D4 IAT scores reflect the difference in mean response latency between the critical blocks divided by the overall variation in those latencies. The IAT score was calculated so that positive values reflected a response bias for the nonsense words that eventually shared a color with positive stimuli (i.e., CS1) relative to the nonsense word which eventually shared a color with negative stimuli (i.e., CS2). Negative values indicated the opposite pattern of responding.

**Hypothesis Testing**

**IAT.** Submitting IAT scores to a one-way ANOVA with CS-US color matching as a between-subject factor revealed a significant effect for Color on IAT scores, *F*(1, 102) = 44.19, *p* < .001, η2partial **=** 0.30, 95% CI [0.16; 0.43], BF10 > 106. Specifically, when CS1 eventually shared a color with positive words and CS2 shared a color with negative words, participants showed a relative preference for CS1 over CS2 (*M* = 0.37, *SD* = 0.46), *t*(47) = 5.56, *p* < .001, *d* = 0.8, 95% CI [0.47; 1.13], BF10 > 104. A reversed preference emerged when CS1 eventually shared a color with negative words and CS2 shared a color with positive words, such that participants showed a relative preference for CS2 over CS1 (*M* = -0.23, *SD* = 0.45), *t*(55) = -3.73, *p* < .001, *d* = 0.49, 95% CI [-0.77; -0.22], BF10 = 55.99. Both scores significantly differed from zero.

Both frequentist and Bayes Factors (using a default prior with scaling factor *r* = .707) *t* tests demonstrated that when CS1 eventually shared a color with positive words and CS2 shared a color with negative words, participants showed a relative preference for CS1 over CS2 (*M* = 0.37, *SD* = 0.46), whereas when CS1 eventually shared a color with negative words and CS2 shared a color with positive words, they demonstrated a relative preference for CS2 over CS1 (*M* = -0.23, *SD* = 0.45), *t*(55) = -3.73, *p* < .001, *d* = 0.49, 95% CI [-0.77; -0.22], BF10 = 55.99.

**Self-reported ratings**. Submitting difference scores to a similar ANOVA as above revealed a main effect of Color on self-reported ratings, *F*(1, 102) = 69.72, *p* < .001, η2partial **=** 0.41, 95% CI [0.26; 0.52], BF10 > 106. When CS1 shared a color with positive words and CS2 shared a color with negative words, participants showed a relative preference for CS1 over CS2 (*M* = 3.32, *SD* = 4.60), *t*(47) = 5.01, *p* < .001, *d* = 0.72, 95% CI [0.40; 1.04], BF10 = 2358. A reversed preference emerged when CS1 shared a color with negative words and CS2 shared a color with positive words, such that participants showed a relative preference for CS2 over CS1 (*M* = -4.15, *SD* = 4.48), *t*(54) = -6.87, *p* < .001, *d* = 0.93, 95% CI [-1.24; -0.61], BF10 > 105. Both scores significantly differed from zero.

**Behavioral intentions**. We also examined if CS-US color matching influenced intentions to consume the two CSs. We calculated the difference in the proportion of participants who intended to purchase CS1 and CS2 across the two CS-US color conditions. Analyses revealed that the proportion of choices in favor of CS1 was higher when CS1 matched the color of the positive USs than when it matched the color of the negative USs (0.40 vs. 0.15, *z* = 2.88, *p* = 0.002). The opposite pattern emerged for the proportion of choices in favor of CS2 when CS2 matched the color of positive USs than when it matched the color of negative USs (0.42 vs. 0.08, *z* = 3.86, *p* < 0.001)[[5]](#footnote-5).

**Discussion**

Result provide initial support for the shared feature account in the attitude domain. Although a neutral stimulus (CS) was repeatedly paired with both positive and negative USs, it acquired the valence of the US that shared a feature (color) with it. Specifically, CSs that appeared in the same color as positive USs were rated positively whereas CSs that appeared in the same color as negative USs were rated negatively. We also obtained evidence for this moderation by shared feature effect on implicit evaluations and behavioral intention measures.

**Experiment 2**

In our second experiment we set out to replicate and extend our initial findings. In Experiment 1, participants completed an EC phase where all stimuli were initially presented in white and only later changed to the same or a different color. This time we presented all stimuli in the same color during the first half of the trial. During the second half, we switched the color of one the USs, while keeping the color of the other US the same as the CS. This modified design allowed us to test two alternative explanations of our effects. The first (*shared feature hypothesis*) argues that an overlap in some stimulus feature (in this case color) will lead people to indicate that those same stimuli share other properties (valence). If so, then we should expect a similar pattern of findings as obtained in Experiment 1. A second possibility (*salience hypothesis*) which argues that people’s attention is fixated on any salient change in the context. In the case of Experiment 1 the CS and one of the USs changed color (and thus their salience increased). The resulting change in liking for the CS could have resulted from this increase in salience rather than the fact that the CS and US shared a feature. If correct, then people may interpret the change (rather than overlap) in color as a cue highlighting that the US which changes color is actually related to the CS. If so, then the CS should acquire the valence of the US which switches color within the trial, leading to the opposite effect predicted by the shared feature account.

**Method**

**Participants and design.** 118 participants (67 females) (*Mage =* 32.3*, SD =* 8.6) took part in the study via the Prolific Academic website.

**Procedure**

A similar procedure was used as in Experiment 1 with the exception of the EC phase.

**EC**. Training once again consisted of three blocks of 16 trials (48 total), with each block containing two different types of trials: one in which CS1 stayed the same color as positive words, and another in which CS2 stayed the same color as negative words. Unlike Experiment 1, the CS and two USs were initially presented in the same color for 3000ms. During one type of trial, CS1 and the positive US remained in the same color (e.g., blue) whereas the negative US changed color (e.g., purple). During the second type of trial, CS2 and the negative US remained in the same color (e.g., yellow) while the positive US changed color (e.g., green). All stimuli remained onscreen for a further 3000ms before being removed, followed by an inter-trial interval, and the next trial (see Figure 2).

**Results**

**Data Preparation**

We excluded data from twelve participants who did not complete the entire session, and a further three who failed to maintain IAT performance criteria. This led to a final sample of 103 participants.

**Hypothesis Testing**

**IAT**. Submitting IAT scores to a one-way ANOVA with Color as a between-subjects factor did not reveal a main effect for Color, *F*(1, 102) = 1.45, *p* = .23, η2partial **=** 0.01, 95% CI [0.00; 0.09], BF01 = 2.5. When CS1 remained the same color as positive words and the color of negative words changed, and when CS2 remained the same color as negative words and the color of positive words changed, participants showed an IAT effect favoring CS2 over CS1 (*M* = -0.26, *SD* = 0.54), *t*(48) = -3.35, *p* = .002, *d* = 0.48, 95% CI [-0.78; -0.18], BF10 = 19.15. When CS1 remained the same color as negative words and the color of positive words changed, and when CS2 remained the same color as positive words and the color of negative words changed, participants did not show an effect favoring either stimulus, (*M* = -0.12, *SD* = 0.61), *t*(53) = -1.47, *p* = .15, *d* = 0.20, 95% CI [-0.47; 0.07], BF01 = 2.43.

**Self-reports.** Similar analyses on self-reported ratings did not reveal a main effect for Color, *F*(1, 102) = 1.05, *p* = .31, η2partial **=** 0.01, 95% CI [0.00; 0.08], BF01 = 3.00. When CS1 remained the same color as positive words and the color of negative words changed, and when CS2 remained the same color as negative words and the color of positive words changed, participants preferred CS2 over CS1 (*M* = -2.76, *SD* = 5.28), *t*(48) = -3.65, *p* = .001, *d* = 0.52, 95% CI [-0.83; -0.22]. Yet when CS1 remained the same color as negative words and the color of positive words changed, and when CS2 remained the same color as positive words and the color of negative words changed, participants did not show a preference for either stimulus (*M* = -1.59, *SD* = 6.09), *t*(53) = -1.93, *p* = .06, *d* = 0.26, 95% CI [-0.58; 0.01], BF01 = 1.2.

**Behavioral intentions**. Analyses revealed that the proportion of choices in favor of CS1 was lower when CS1 stayed the same color as positive USs and negative USs changed color than when it stayed the same color as negative USs and positive USs changed color (0.15 vs. 0.33, *z* = -2.14, *p* = 0.02). The proportion of choices in favor of CS2 was also lower when CS2 stayed the same color as positive USs and negative USs changed color than when it stayed the same color as negative USs and positive USs changed color (0.24 vs. 0.48, *z* = -2.58, *p* < 0.05).

**Discussion**

The findings of Experiment 2 contradicted those obtained in Experiment 1. During the EC phase a CS and two USs were first presented in the same color, and then one of the USs changed to a different color. Evidence suggests that participants preferred a CS when it shared a color with negative USs (but there was a change in the positive US color) more than they preferred a CS when it shared a color with positive USs (but there was a change in the negative US color). Put another way, rather than base their evaluations on a shared feature (i.e., the fact that a US stayed in the same color as the CS) participants seemed to have based their evaluations on the US that changed color during the trial.

**Experiment 3**

The findings reported in Experiment 2 contradict the shared features account – at least at the aggregated group level. Yet the high degree of variability in evaluative responses combined with a closer inspection of the individual level data (see Supplementary Materials section), as well as responses on the influence awareness question, suggest a more nuanced story. It seems that there may have been different groups of participants in our sample: those that did not show any attitudes towards either CS1 or CS2 (for potentially different reasons), those that showed attitudes in-line with the shared features account (e.g., “*the CS and US share a color, therefore they are related*”), and a third that showed attitudes in-line with a salience hypothesis (e.g., “*the color of one of the stimuli [USs] changed…this [US] must be important or related to the CS in some way*”). It appears that this latter group exerted more of an impact on the (overall) group level responses reported in Experiment 2 than the other groups. Responses on the influence awareness question also broadly map onto these different patterns of evaluation.

In retrospect, we believe there may have been a relative simple explanation for the aforementioned findings: the change in task instructions from Experiment 1 to 2. In *Experiment 2* participants were told that “you will see two new words: Morag and Struan. These words will appear onscreen together with two other words. The new word (Morag or Struan) and other words will initially appear in one color. *Then the color of one of the words will change…Please pay close attention to the color of each word and how they change*”. These instructions may have encouraged people to focus greater attention to the change, rather than the overlap, in color, and thus treat changes in color as more diagnostic about CS valence than the shared feature. If so, then modifying task instructions in a way that directs attention to the shared feature may lead to similar effects as seen in Experiment 1. With this in mind, we replicated Experiment 2 once again while modifying the instructions to emphasize that the CS and US remained in the same color.

**Method**

**Participants and design.** 118 participants (70 females, *Mage* = 28.19, *SD* = 6.08) took part in the study via the Prolific Academic website.

**Procedure**

An identical procedure was used as in Experiment 2 with the exception of the instructions provided prior to the EC phase.

**EC**. Prior to the EC phase participants were told the following: “You are going to see a new word appear on the screen (i.e., Morag or Struan). Morag or Struan will appear on the left of the screen. Two other words will appear on the right. Morag or Struan and other words will first appear in the same color. Morag or Struan will stay the same color as one of the words on the right. Please pay close attention to the colors of the words. You will be asked some questions about this later on.

**Results**

**Data Preparation**

Nine participants did not complete the entire session while an additional twelve did not meet the IAT criteria. This led to a final sample of 97 participants.

**Hypothesis Testing**

**IAT**. Submitting IAT scores to a one-way ANOVA with Color as a between-subjects factor revealed a main effect for Color, *F*(1, 96) = 10.59, *p* = .002, η2partial **=** 0.10, 95% CI [0.02; 0.22], BF10 = 20.09. Specifically, when CS1 remained the same color as positive words and negative words changed color, and when CS2 remained the same color as negative words and positive words changed color, then participants showed an IAT effect favoring CS1 over CS2 (*M* = 0.21, *SD* = 0.46), *t*(46) = 3.02, *p* = .004, *d* = 0.45, 95% CI [0.14; 0.76], BF10 = 8.33. Yet when CS1 remained the same color as negative words and positive words changed color, and when CS2 remained the same color as positive words and negative words changed color, participants showed weak evidence for an IAT effect favoring CS2 over CS1, (*M* = -0.15, *SD* = 0.59), *t*(50) = -1.77, *p* = .08, *d* = 0.25, 95% CI [-0.53; 0.03], BF01 = 1.54.

**Self-reported ratings.** Conducting similar analyses as reported above on the self-reported ratings also revealed a main effect for Color, *F*(1, 96) = 14.51, *p* < .001, η2partial **=** 0.13, 95% CI [0.03; 0.26], BF10 = 99.59. Specifically, when CS1 remained the same color as positive words and negative words changed color, and when CS2 remained the same color as negative words and positive words changed color, then participants preferred CS1 over CS2 (*M* = 2.03, *SD* = 5.53), *t*(45) = 2.49, *p* = .02, *d* = 0.37, 95% CI [0.07; 0.67], BF10 = 2.55. Yet when CS1 remained the same color as negative words and positive words changed color, and when CS2 remained the same color as positive words and negative words changed color, participants preferred CS2 over CS1, (*M* = -2.22, *SD* = 5.46), *t*(50) = -2.91, *p* = .005, *d* = 0.41, 95% CI [-0.69; -0.12], BF10 = 6.31.

**Behavioral intentions**. Analyses revealed that the proportion of choices in favor of CS1 was higher when CS1 stayed the same color as positive USs (.44) than when it stayed the same color as negative USs (.14, *z* = 3.26, *p* < 0.001). The opposite pattern emerged for the proportion of choices in favor of CS2 when CS-US color matching was reversed (i.e., the proportion of CS2 selections was higher when it stayed the same color as positive USs than when it stayed the same color as negative USs (0.33 vs. 0.15, *z* = 2.06, *p* < 0.02).

**Discussion**

When task instructions directed attention towards (rather than away from) the shared feature a shared features effect emerged. Although a neutral stimulus (CS) was repeatedly paired with both positive and negative USs, it acquired the valence of the US that shared a common feature (color) with it. Specifically, CSs that shared a color with positive USs were liked more than CSs which shared a color with negative USs. We obtained evidence for this moderation by shared feature effect on explicit, implicit, and behavioral intention measures.

**Experiment 4**

Until now we have seen how one shared feature (color) comes to moderate implicit and explicit attitudes. Yet our account suggests that other shared features should function in a similar way. Indeed, a common size, direction, location, smell, or taste shared by two stimuli should lead people to act as if those stimuli share other features as well (valence). Therefore, in order to extend and generalize our findings, we swapped one feature (color) for another (size), and set out to demonstrate that this second feature can also moderate likes and dislikes whenever two stimuli share it. In Experiment 4 participants once again encountered an EC phase in which three stimuli (CS, USpos, USneg) were presented onscreen. This time CS1 and positive USs were presented in the same sized font whereas negative USs were presented in a differently sized font. Likewise, CS2 and negative USs shared a common sized font whereas positive USs were always presented in a different sized font. If we are correct, then CSs should acquire the same valence as the USs with which they share a common size.

**Method**

**Participants and design**. 212 participants (103 females, *Mage* = 30.33, *SD* = 6.18) took part in the study via the Prolific Academic website.

**Procedure**

A similar procedure was used as in Experiments 1-2 with the exception of the EC phase.

**EC**. The EC phase consisted of three blocks of 16 trials (48 total) consisting of two different types of trials. During one type of trial CS1 was presented in the same sized font (e.g., Arial 8%) as positive words and a different sized font as negative words (e.g., Arial 4%). In another type of trial CS2 is presented in the same sized font as negative words and a different sized font as positive words. Stimuli were always presented in the same color (white) and the sizes of the fonts was randomly counterbalanced across trials (e.g., sometimes a CS and US share a small [Arial 4%] font and at other times they shared a large [Arial 8%] font) (see Figure 3).

**Results**

**Data Preparation**

Fifteen participants did not complete the entire session while an additional nine did not meet the IAT criteria. This led to a final sample of 187 participants.

**Hypothesis Testing**

**IAT.** Submitting IAT scores to a one-way ANOVA with CS-US size matching as a between-subject factor found a significant effect for Size, *F*(1, 187) = 24.85, *p* < .001, η2partial **=** 0.12, 95% CI [0.04; 0.21], BF10 > 104. Specifically, when CS1 shared a common size with positive words and CS2 shared a common size with negative words, participants showed a relative preference for CS1 over CS2 (*M* = 0.16, *SD* = 0.49), *t*(92) = 3.22, *p* = .002, *d* = 0.33, 95% CI [0.13; 0.55]. A reversed preference emerged when CS1 shared a common size with negative words and CS2 shared a color with positive words, such that participants showed a relative preference for CS2 over CS1 (*M* = -0.18, *SD* = 0.45), *t*(94) = -3.85, *p* < .001, *d* = -0.39, 95% CI [-0.60; -0.19]. Both scores significantly differed from zero.

**Self-reported ratings**. Conducting a similar set of analyses as reported above revealed a main effect of Size, *F*(1, 186) = 73.19, *p* < .001, η2partial **=** 0.28, 95% CI [0.18; 0.38], BF10 > 105. When CS1 shared a common size with positive words and CS2 shared a common size with negative words, participants showed a relative preference for CS1 over CS2 (*M* = 3.57, *SD* = 4.99), *t*(92) = 6.89, *p* < .001, *d* = 0.71, 95% CI [0.95; 0.49], BF10 > 105. A reversed preference emerged when CS1 shared a common size with negative words and CS2 shared a common size with positive words, such that participants showed a relative preference for CS2 over CS1 (*M* = -2.18, *SD* = 4.19), *t*(94) = -5.07, *p* < .001, *d* = -0.52, 95% CI [-0.74; -0.31], BF10 = 7382. Both scores significantly differed from zero.

**Behavioral intentions**. Analyses revealed that the proportion of choices in favor of CS1 was higher when CS1 shared a common size with positive USs than when it shared a size with negative USs (0.45 vs. 0.15, *z* = 4.56, *p* < .001). The opposite pattern emerged for the proportion of choices in favor of CS2 when size matching was reversed (i.e., higher proportion of CS2 selections when CS2 shared a common size as positive USs than when it shared a common size with negative USs (0.32 vs. 0.14, *z* = 2.87, *p* = 0.002).

**Discussion**

Results indicate that size can also function as a shared feature and moderate implicit and explicit attitudes as well as behavioral intentions. During the EC phase a CS was presented with two USs – one positive and another negative. When a CS was presented in the same size as positive USs it was liked more than a CS that was presented in the same size as negative USs. These findings replicate those obtained in Experiments 1-3 and demonstrate that different types of shared features can bias attitudes and intentions.

**Experiment 5**

In Experiments 1-4, we exclusively focused on how perceptual features shared by stimuli (e.g., their color or size) moderate evaluative responses established via stimulus pairings. We interpreted the resulting effects as evidence for the utility of a novel concept: shared features. Yet one could argue that other concepts already exist which are capable of explaining our findings. Take, for instance, Gestalt theory which has identified a number of principles describing how humans organize individual elements into groups, and as a result, perceive and recognize patterns. One principle (similarity) seems particularly relevant to our work. The principle of similarity indicates that items which share visual features - such as their shape, size, color, or orientation - will be perceived as being more related to one another than those that do not (see Gallace, & Spence, 2011, Goldstein, 1999).

If our account was limited to the idea that stimuli share *physical features* with one another, and these physical overlaps led to a transfer of features from one stimulus to another, then Gestalt principles may indeed be sufficient (i.e., people used the overlap in color and shape between CS and US as a cue indicating that those stimuli were related to one another, and this relatedness leads to a transfer of valence from one to the other). However, our shared features account extends far beyond simple perceptual relatedness. It argues that stimuli can also share *symbolic features* with one another, and when stimuli share a symbolic feature, this may cause other features to be shared by those same stimuli as well (e.g., valence). Thus our shared features account speaks to both physical and symbolic feature overlap. [[6]](#footnote-6)

Now, unlike the physical features that stimuli can share, which are - by definition - dependent on some common physical property such as size, color, or shape, symbolic features are free from any such constraints. Shared symbolic features allow stimuli to be connected in a near infinite number of ways and for the properties of one stimulus to influence how people behave towards other stimuli. Consider, for example, research on person perception: learning that a symbolic property of one person (e.g., a personality trait such as ‘extroversion’ or ‘conscientiousness’) is shared by another person can bias evaluations of the latter even when they share no perceptual similarity with the former (e.g., Skowronski, Carlston, Mae, & Crawford, 1998). The same is true in marketing. The symbolic properties of one product (e.g., brand name or identity) can bias evaluations of other products (e.g., Apple laptops and phones) (Ratliff et al., 2012), despite the fact that the products are physically dissimilar (for more on this see Hughes et al., in press).

In Experiment 5 we set out to determine if a symbolic feature shared by two stimuli can moderate evaluations in the same way that physical features did in Experiments 1-4. To achieve this, we first established a symbolic relation between two sets of colors (i.e., *Blue-Same-Yellow* and *Green-Same-Purple*) followed bya similar EC phase to that used in Experiments 1-3. However, this time, we presented a CS in either blue or green along with a positive and negative US that were presented in either yellow or purple. If a CS is presented in blue and a positive US is presented in yellow (along with a negative US in purple) then participants should evaluate that stimulus positively (given that blue and yellow were trained to be symbolically similar to one another in the first phase of the experiment). In contrast, if participants encounter a CS in green along with a negative US in purple (and a positive US in yellow) then they should evaluate that CS negatively (given that green and purple were trained to be symbolically similar to one another in the first phase of the experiment). Such a finding is noteworthy for several reasons. It would further replicate our findings and expand the remit of the shared features concept by demonstrating that the shared feature moderating attitude formation can be symbolic rather than purely physical. It would also demonstrate that shared features can also account for outcomes that extend beyond the remit of previous concepts (such as the Gestalt principle of similarity).

**Method**

**Participants and design.** 214 participants (108 females, *Mage =* 30.65*, SD =* 6.08) took part in the study via the Prolific Academic website.

**Procedure**

The study consisted of four phases: color training, EC, evaluative measures, and exploratory questions.

**Color training.** Color training consisted of three blocks of 16 training trials followed by one block of 16 test trials. A Matching to Sample (MTS) task was used to establish symbolic relations between two sets of colors (e.g., Yellow-Blue and Green-Purple). On each trial, one color was presented at the top of the screen, and two at the bottom. Participants had to select the color at the bottom that went with the color at the top and were told that corrective feedback provided by the computer would help them do so. When a correct response was emitted then all stimuli were removed from the screen, a feedback message (‘Correct’) presented, followed by a 500ms ITI. If an error was made, stimuli were once again removed, corrective feedback provided (‘Wrong’), an ITI followed by the next trial. Test trials were identical to training trials with the exception that corrective feedback was no longer provided (see Figure 4).

**EC**. The EC phase consisted of three blocks of 16 trials (48 total), with each block containing two types of trials: one type of trial where CS1 was presented in blue, a positive word in yellow, and a negative word in purple. In another type of trial CS2 was presented in green, and the valenced words were presented in the aforementioned colors. All three stimuli were presented against a black background for 5000ms. Thereafter, all stimuli were removed, followed by an inter-trial interval of 750ms, and the next trial (see Figure 5).

**Evaluative measures**. Evaluative measures were similar to Experiments 1-4.

**Exploratory questions**. Participants were asked a similar set of exploratory questions as in Experiments 1-4. We also probed for *color contingency awareness* (i.e., what the relationship was between the various colors), and *CS-US color contingency awareness* (i.e., if they could recall what color the CSs and USs were presented in).

**Results**

**Data Preparation**

Fifteen participants failed to provide complete data. A further twenty failed to meet the IAT criteria. This led to a final sample of 179 participants.

**Hypothesis Testing**

**IAT**. A main effect of symbolic color match was obtained, *F*(1, 178) = 14.39, *p* < .001, η2partial= 0.08, 95% CI [0.02; 0.16], BF10 = 109.05. When CS1 was presented in a color that was symbolically similar to the color that positive words were presented in, and CS2 was presented in a color that was symbolically similar to the color that negative words were presented in, then participants preferred CS1 over CS2 (*M* = 0.14, *SD* = 0.46), *t*(98) = 3.09, *p* = .003, *d* = 0.31, 95% CI [0.11; 0.51], BF10 = 9.31. Yet when CS1 was presented in a color that was symbolically similar to the color that negative words were presented in, and CS2 was presented in a color that was symbolically similar to the color that positive words were presented in, then participants preferred CS2 over CS1 (*M* = -0.12, *SD* = 0.46), *t*(79) = -2.33, *p* = .02, *d* = 0.26, 95% CI [-0.49; -0.04], BF10 = 1.55.

**Self-reported ratings.** A one-way ANOVA with symbolic color match as a between-subjects factor revealed a main effect for Color, *F*(1, 178) = 58.61, *p* < .001, η2partial **=** 0.25, 95% CI [0.15; 0.35], BF10 > 106. Once again, participants preferred the stimulus (CS1) that was presented in a color that was symbolically similar to the color that positive words were presented in more than the stimulus (CS2) that was presented in a color that symbolically similar to the color that negative words were presented in (*M* = 2.34, *SD* = 4.12), *t*(98) = 5.66, *p* < .001, *d* = 0.57, 95% CI [0.36; 0.78], BF10 > 104. When the color contingencies were reversed, participants preferred CS2 over CS1 (*M* = -2.38, *SD* = 4.09), *t*(79) = -5.22, *p* < .001, *d* = 0.58, 95% CI [-0.82; -0.35], BF10 > 104.

**Behavioral intentions**. Analyses showed that the proportion of responses in favor of CS1 was higher when the color of CS1 symbolically matched the color of the positive USs compared to when it symbolically matched negative USs (0.29 vs. 0.13, *z* = 2.71, *p* < 0.05). The opposite pattern emerged for the proportion of choices in favor of CS2 (i.e., the proportion of selections of CS2 was larger when its color symbolically matched that of positive than negative USs (0.29 vs. 0.14, *z* = 2.39, *p* < 0.003).

**Discussion**

Experiment 5 extends our account further and shows that shared symbolic features moderate implicit and explicit attitudes in a similar way to shared perceptual features. Prior to an EC phase, two symbolic relations between colors were trained (i.e., *Blue-Similar-Yellow*, and *Green-Similar-Purple*). Then, during an EC phase, a CS was simultaneously presented with two USs. Critically, a neutral stimulus was presented in either blue (CS1) or green (CS2), whereas positive USs were presented in yellow and negative USs in purple. Self-reported ratings, IAT effects, and behavioral intention measures all indicated that the CS presented in blue was preferred relative to the CS presented in green, supporting the idea that shared symbolic (color) features can led to a transfer of other stimulus properties (valence).

**Meta-analysis of Experiments 1-5**

In order to provide a more robust estimate as to whether EC effects are moderated by shared features rather than mere contiguity, we conducted a mini meta-analysis based on the data from Experiments 1-5 following the practice proposed by Goh, Hall, and Rosenthal (2016). Meta-analyses revealed that across studies shared features influenced self-reported ratings as illustrated by a mean weighted effect size of *d* = xx, Z = xx, *p* = xx. A Bayesian one-way ANOVA with Shared Feature as a fixed factor (and study ID as a random factor) further supported this conclusion, such that self-reported ratings of the CS that shared a feature with positive USs (*M* = xx, *SD* = xx) significantly differed from the ratings of the CS that shared a feature with negative USs (*M* = xx, *SD* = xx), BF01 = xx. Meta-analyses also revealed that across studies shared features influenced IAT effects as illustrated by a mean weighted effect size of *d* = xx, Z = xx, *p* = xx. IAT effects for the CS that shared a feature with positive USs (*M* = , *SD* = ) were significantly different to those in which the CS shared a feature with negative USs (*M* = , *SD* = ), BF01 = xx.

**General Discussion**

In this paper we introduce the shared features account. This account draws attention to the fact that when two stimuli share a feature, people will act as if those stimuli are related in some way, and also assume that they share other features as well. This latter idea - that a feature shared by source and target objects will lead people to make assumptions about other target object features - is not only interesting in its own right, but may underpin many phenomena in psychological science. In this first paper we sought to experimentally establish and manipulate the shared features effect in the domain of attitudes, test potential boundary conditions, and demonstrate that it holds across different physical and symbolic features. Across five studies we exposed participants to an EC phase containing three stimuli: a neutral CS, a positive US, and a negative US. We then manipulated each trial so that a CS would either share a color (Experiments 1-3) or size (Experiment 4) with one of the USs. To demonstrate that our account speaks to both physical and symbolic features, Experiment 5 created a symbolic relation between two sets of colors (*Blue-Similar-Yellow*, and *Green-Similar-Purple*) and then, during the EC phase, presented a CS in either blue or yellow and the USs in green or purple. If EC effects were primarily driven by contiguity, as is often assumed, then we should have observed ambivalent responses to the CSs (given that they were repeatedly paired with positive and negative USs). Yet if our shared features account is correct, then CS evaluations should be driven by the feature they shared with a given US.

Our results provide strong and repeated support for the shared features account. EC effects were not driven by mere contiguity but instead by the features CSs and USs shared with one another. These shared feature effects were evident in the form of self-reported ratings, behavioral intentions, and IAT effects which consistently favored one CS over the other. They also emerged regardless of the type (color or size) and nature (physical or symbolic) of feature manipulated. These conclusions were further reinforced by our mini meta-analysis where shared features moderated implicit and explicit attitudes across four of the five studies. Taken together, we can say that the shared features effect is general, reliable, and replicable across a range of measures in the attitude domain.

**Theoretical Implications**

Until now we focused on the shared features effect itself and said little about why this effect actually emerges. There are two different levels at which to explain shared features effects (De Houwer, 2011; Hughes et al., 2016): (1) a mental level that aims to uncover the mental mechanisms that *mediate* the impact of the environment on behavior and (2) a functional level that aims to describe those elements of the environment that *moderate* behavior. We consider both in turn.

**The functional level of explanation (relational responding).** Without going into too much detail, functional explanations are not concerned with identifying the mental representations and processes. Rather they seek to identify which aspects of the environment give rise to changes in behavior. They then use this functional knowledge to create ‘abstract’ principles that can explain how the environment influences behavior across many different situations and settings (for a detailed treatment see Hughes & Barnes-Holmes, 2016a; De Houwer & Hughes, 2019). At this level shared features effects could be conceptualized as one instance of a much larger phenomenon known as relational responding. At its most basic, relational responding is a type of behavior that involves ‘responding to the relationship between stimuli’. This behavior is typically emitted in the presence of a stimulus called a *relational contextual cue*. This stimulus is a *contextual cue* in the sense that it signals (cues) how one should respond and it is *relational* because it signals that a relational response should be emitted in that context. To illustrate, consider a simple cue such as a red traffic light at a busy intersection. The light signals how one should respond in that context (i.e., that walking across the street when the light is red will be dangerous for that person). Relational contextual cues require us to take this idea one step further. They also signal how one should respond. But instead of responding to just one stimulus they indicate that we should respond based to how stimuli are *related* to one another in that context. For instance, color may be a cue signaling that two stimuli are similar (strawberries and tomatoes) or different (strawberries and oranges) whereas size could be a cue indicating that one stimulus is bigger (tomatoes relative to strawberries) or smaller than the other (strawberries relative to oranges). The take home message here is that organisms can treat different properties of a stimulus as a signal indicating how one stimulus is related to another.

Several decades of work reveals that when stimuli are related to one another (in the presence of a relational contextual cue), the psychological properties of one stimulus typically transfer to other related stimuli (see Hughes & Barnes-Holmes, 2016a). To illustrate, imagine that you learn that a green gas is similar to a green liquid and that this liquid is in turn similar to a green solid. Thereafter you come into physical contact with the gas and immediately feel sick. As a result of this experience you may avoid and fear the liquid and solid even though you have no prior experiences of either causing you harm. In this example, color (green) functions as a relational contextual cue indicating that three stimuli (gas, liquid, solid) are similar to one another, and once this relationship has been established, the psychological properties of one stimulus (gas) transfer to the other related stimuli (liquid, gas) (see REF).

These ideas bear remarkable similarity to the shared features effects observe in our studies (and the idea of function transformation more generally: see De Houwer et al., *in prep*). One could conceptualize shared features such as color (Experiments 1-3) and size (Experiment 4) as relational contextual cues which signaled that two of the three stimuli present in an EC trial (a CS and either a positive or negative US) were similar to one another. Once such a relationship was formed still other functions (features) of one stimulus (US) were transferred to the other (CS), thus leading to a change in evaluative responding. The fact that symbolic features also moderated EC effects (Experiment 5) is consistent with past work on relational responding (REF). Thus our shared feature effects are in-line with modern (functional) conceptualizations of learning and behavior (e.g., Hayes, Barnes-Holmes, & Roche, 2001), and particularly with the idea of relational responding and relational contextual cues.

**The functional level of explanation (relating analogically).** One could not only conceptualize shared feature effects as a relational response but go one step further and view them as a relatively complex response known as ‘relating analogically’. Analogies are typically defined as the relating of relations between two pairs of concepts. This takes place whenever there is a source relation, a target relation, and one of the objects in the target relation is similar to one of the objects of the source relation (e.g., ‘*good is to bad*’ [target relation] as ‘*night is to day*’ [source relation]). When these conditions are met, objects from the source and target relations are related to one another, and how people respond to the target relation subsequently changes (Gentner & Smith, 2013; Hussey & De Houwer, 2013). This idea also bears similarity to the shared feature effects reported here. During the EC phase in Experiments 1-3, participants encountered a source relation consisting of stimuli of identical color (‘Green is to Green’) and a target relation (‘CS is to US’). It may be that participants related the source to the target relation based on the fact that elements of those relations shared a feature (i.e., their color) (‘CS is to US’ as ‘Green is to Green’). Once this occurred properties of objects in the target relation changed (“*if green is the same as green then the CS is the same as the US…therefore the CS is good or bad*”). It is also possible that the fact that the CS and the second US did not share a color served to signal that objects in the source relation (‘Red is to Green’) were opposite or different to objects in the target relation (‘CS is to US’), and this may have influenced properties of objects in the target relation as well (“*if green is different to red then the CS is different to the US…therefore the CS is good or bad*”). A similar explanation would hold for Experiments 4-5 as well. If correct, then this would suggest that people may be responding to stimuli in relatively complex ways even in seemingly simple learning tasks. Or put differently, that EC effects, which are often viewed as ‘simple’ changes in behavior that are driven by simple ‘associative’ mental mechanisms are actually more complex than is often believed (for more on this idea see De Houwer and Hughes 2016; Hughes, De Houwer, & Barnes-Holmes, 2016).

**The mental level of explanation (inferential reasoning).** At themental level the goal is to identify the mental representations and processes that mediate between environment and behavior. We believe that shared feature effects fit well with an inferential perspective on human learning and attitudes (e.g., De Houwer, 2018; Van Dessel, Hughes, & De Houwer, 2018). The core conceptual unit of this perspective is a proposition – namely – an informational unit “that contains information about the nature of the relation between stimuli (e.g., A predicts B, A causes B, A co-occurs with B, …)” (De Houwer, 2018, p.3). *Inferences* represent a sub-type of such propositions – namely – those generated from other momentarily available propositions. “The construction process that leads to the inference can be seen as an information generation procedure that involves the application of information generation (i.e., inference) rules to information that is currently entertained” (Van Dessel et al., 2018, p.4). [[7]](#footnote-7)

In our studies the fact that the source and target object shared a feature may have caused participants to form a series of inferences about (a) how those stimuli were related and (b) the properties of the target object. It is these inferences which mediated subsequent evaluative responding. For instance, during the EC phase in Experiments 1-5, participants may have formed a number of simple propositions concerning the source and target objects (e.g., ‘the CS is presented in green’, the positive US is presented green’, and ‘the negative US is presented in purple’). They may have also generated a number of propositions about the source and target object features (e.g., ‘this word [CS] is neutral’, ‘that word [US] is positive’ and ‘that word [US] is negative’). These basic propositions may have served as the ‘raw ingredients’ for a relational inference (‘the CS and US are similar in that they are both green’), and thus an inference about the target objects features (‘the CS is also good’). Note that these inferences were generated on the basis of both physical and symbolic features that objects shared. Thus, from an inferential perspective, the ‘assumptions’ at the core of shared feature effects are actually inferences that are constructed on the basis of past and present propositions about the source and target objects, their features, and how they are related.

If correct, then this raises a number of interesting questions about shared feature effects themselves. The first has to do with rational vs. irrational inferences. On the one hand, the inference that ‘stimuli which share a feature are related to one another’ certainly seems rational given that these stimuli are indeed related (they are both presented in the same color or size). On the other hand, the inference that ‘stimuli which share one feature also share other features’ is potentially irrational given that the individual has no experience or information available to ensure that such a conclusion is correct. The fact that the CS and US are both presented in green does not guarantee that they share other properties with one another. Or in a more applied scenario, just because a source (Mike) and target object (Bob) share one feature (are both bald) does not necessarily mean that they also share other features (e.g., if Mike is dominant this does not mean that Bob is also dominant). Yet this is precisely how participants behaved in our studies.

One possible reason is that we used novel objects that people had no prior knowledge or experience with. In these cases, there is no further information available that can inhibit an irrational inference from occurring. Indeed, under these conditions, it may be rational to assume that objects which share one feature also share others. Yet when a pre-existing network of information does exist for both source and target objects as well as their features (as is the case for celebrities and sports stars, stereotyped groups, known and counterfeit brand products) then this existing network of information may qualify the types of inferences that are drawn on the basis of shared features, and thus the impact that shared features ultimately have on behavior.

The above perspective counters the common idea that inferential reasoning is inherently a ‘cold’, rational, and error-free process (REF), and instead argues that it can be irrational, especially when it is (a) biased by the amount and type of information that is entertained or (b) inferential rules are incorrectly applied. For instance, when people lack additional qualifying information (as was the case in our studies) they may generate inferences that are not logically correct in nature, leading to the types of outcomes observed here. Alternatively, irrational inferences could also emerge through the misapplication of heuristic reasoning rules based on similarity (e.g., if stimuli are similar in one way they are similar in others) or those which lead to probabilistic reasoning errors (if stimuli are similar in one way the probability is high that they are similar in others). In short, the shared features effect may rest on an assumption (or inference) that is irrational and which stems from inferential rules that are incorrect, or correct, but which are misapplied in that context (for more see Van Dessel et al., 2018). This idea seems worthy of future investigation.

A second question concerns the operating conditions under which inferences are generated and applied on the basis of shared features. It is often assumed that the formation of propositions and inferences depends not only on awareness of the to-be-related stimuli but also the cognitive resources, time, and motivation to relate those stimuli (De Houwer, 2009). Although we did not manipulate these factors it could be argued that all necessary preconditions were fulfilled during the EC phase of our experiments. Our findings are also consistent with a growing body of work showing that (once formed) propositions and inferences may be activated automatically from memory and guide the evaluation of stimuli on self-report and automatic measures alike (see Hughes et al., in press; De Houwer, 2014). For example, once the inferred proposition “CS is good” was generated it may have been stored in memory and retrieved automatically during testing.

**Open Questions and Future Directions**

**Empirical implications**. It is important to realize that our account is not limited to attitudes but speaks to human behavior more generally. It argues that shared features have the potential to influence non-evaluative stimulus properties in ways that are relevant to clinical (e.g., fear, anxiety, disgust), social (accessibility), consumer (brand quality, identification), health (avoidance, escape), and cognitive psychology (attention). Future work could test this claim using the procedures outlined here. If we are correct, then the shared features effect may be a ubiquitous phenomenon relevant to many different domains in psychological science.

We also limited our initial efforts to attitude formation. Future work could investigate whether shared features can also be used to strengthen or change attitudes as well. For instance, researchers could establish a novel attitude towards an unknown object, or take a pre-existing attitude towards a known object (e.g., towards a celebrity, brand product, phobic or clinically relevant stimulus such as spiders or alcohol). Those attitudes could then be modified using a similar task as used in Experiments 1-5. For instance, imagine that participants first complete the same EC phase as we used here to generate an attitude and were then exposed to similar task designed to change those attitudes. Researchers could vary this second task so that the CS no longer shares a color with either type of US (extinction), swap the shared feature contingencies so that the CS now shares a feature with the opposite US used in the first phase (counter conditioning) or is exposed to the exact same contingencies as before, but between the formation and change phases, the US the CS shares a feature with is subjected to a US-revaluation procedure. In each case, they could examine if attitudes towards the CS change as a result of such manipulations.

When carrying out this work, researchers should also investigate the factors that moderate shared feature effects. Such work could examine if the type and nature of the shared features, or the properties of the source and target objects, or the types of features being transformed from one object to another matters across different domains. How shared features are established and changed may also matter: it may be easier to form and modify these effects via experience relative to observation or instruction. Our work identified one such boundary condition: attention. In Experiments 2-3 directing attention towards the shared feature led to implicit and explicit attitude formation whereas directing attention away from that feature eliminated the effect. One possibility is that people treat shared features as a cue that is ‘diagnostic’ for how they should respond to the target object (i.e., how they should evaluate the CS). It may be that the impact of shared features on behavior is moderated by other cues in the environment which signal to what extent the shared feature is diagnostic or not when making a judgement. This is worthy of further study. There may be still other conditions necessary for these effects to emerge and change that should also be examined. The current studies utilized just one type of procedure to document these effects and readers should be careful not to conflate the former with the latter. Many other procedures could be devised to study this class of effects.

If we are right, and many well-known phenomena involve source and target objects that share features, and these shared feature lead people to make assumptions about other target object properties, then moderators of shared feature effects have already been discovered. They are simply being labelled using the idiosyncratic terms of different intellectual traditions. Therefore, rather than ‘reinvent the wheel’, future work could first test to see if shared features play a role in stereotyping, impression formation, persuasion, branding, and other areas. If so, then the moderators of those effects should also moderate shared features effects as well.

**Theoretical implications**. We previously argued that shared features could be conceptualized as relational contextual cues. If so, then it should be possible to change their relational meaning, and thus the assumptions people make about target object features on the basis of those cues. Although a shared feature will typically signal that the source and target objects are similar to one another (and thus give rise to *feature transfer*) there is no reason why a shared feature cannot instead signal that the source and target are related in other ways (and thus give rise to *feature transformation*). For instance, it may be that the feature shared by a source and target object signals that those objects are opposite, hierarchical, different, or related in any number of other ways. If subsequent work were to supports this hypothesis, then it may be more accurate to label the shared feature effect as the ‘related features effect’ (i.e., a change in the assumptions about target object features based on the fact that source and target have a feature that indicates how they are related to one another). This also seems like a promising future research direction.

Yet another interesting possibility is that the absence of shared features may influence behavior as well. In a recent set of studies De Houwer & Hussey (*in prep*) exposed participants to a simple learning task where participants had to assign valenced words to the left-side of the screen and unknown nonsense words to the right-side of the screen. Following training the nonsense words acquired an opposite valence to the valence items themselves. In this case, a source (valenced) and target (nonsense word) object did not share a feature with one another. The fact that they did not (i.e., that they were assigned to opposite sides of the screen) may have led people to make certain assumptions about the target object based on the source object features (i.e., that the nonsense words had an opposite valence to the source). If so then there may be a ‘non-shared features effect’ waiting to be systematically explored.

**Conclusion**

The main aim of this paper was to introduce the shared features account and provide an initial test of its heuristic and predictive value. We found that when a valenced source and neutral target object shared one feature with one another (color or size), this was enough to influence assumptions about other features of the target (valence). This was true for both implicit and explicit attitudes and when the type and nature of the shared feature was varied. Although this paper focused on just one domain (attitudes) our conceptual account applies to many more, and offers a unified way to describe and analyze shared feature effects throughout psychological science. In all likelihood, there are many other domains and phenomena that could be conceptualized as instances of shared feature effects than covered here. We hope that our account will stimulate a new wave of research on this topic and contribute to a broader and deeper understanding of the way in which people arrive at assumptions about the features of objects in their environment.

References

Appendix A

*Figure 1*. Schematic illustration of the two types of trials during the evaluative conditioning phase of Experiment 1. During the first half of the trial (*left*) the CS and USs were presented in white. During the second half of the trial (*right*) the CS and one of the USs changed to the same color whereas the second US changed to a different color.

**HELL**

**STRUAN**

**SMILE**

**CANCER**

**MORAG**

**JOY**

**CANCER**

**MORAG**

**JOY**

**HELL**

**STRUAN**

**SMILE**

*Figure 2*. Schematic illustration of the two types of trials during the evaluative conditioning phase of Experiments 2-3. During the first half of the trial (*left*) the CS and USs were presented in the same color. During the second half of the trial (*right*) the CS and one of the USs remained the same color whereas the second US changed to a different color.

**HELL**

**STRUAN**

**SMILE**

**CANCER**

**MORAG**

**JOY**

**CANCER**

**MORAG**

**JOY**

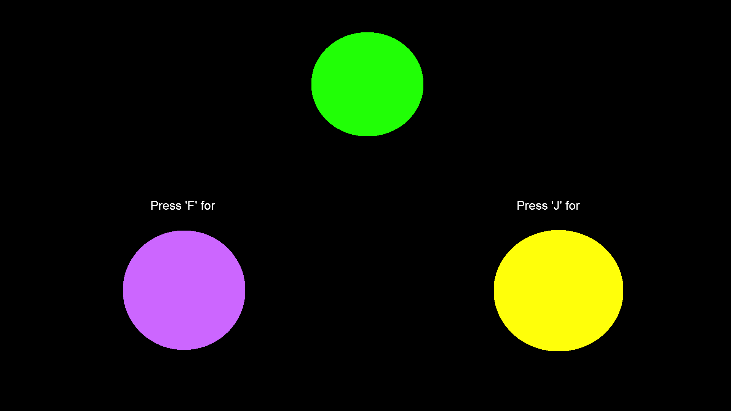
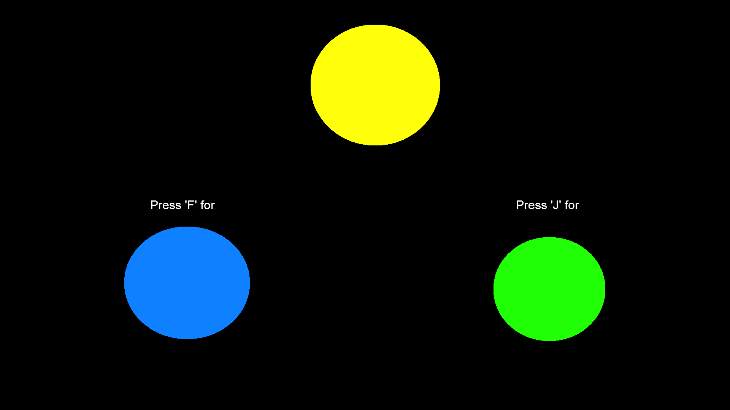
**HELL**

**STRUAN**

**SMILE**

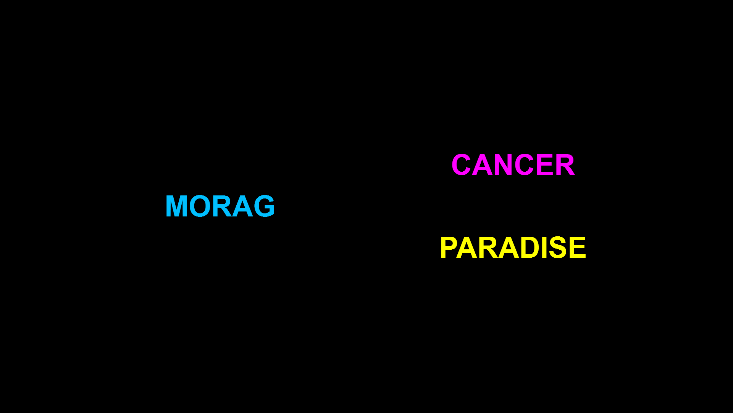
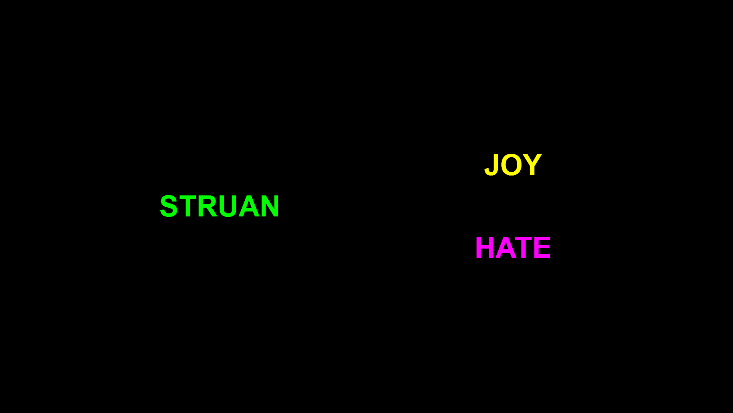


*Figure 3*. Schematic illustration of the two types of trials during the evaluative conditioning phase of Experiment 4. During one type of trial (*left*) one CS and negative USs were presented in the same size (sometimes in large and other times in small font). During a second type of trial (*right*) another CS and positive USs were presented in the same size.



*Figure 4*. Schematic illustration of the two types of trials during the color training phase of Experiment 5. Selecting blue in the presence of yellow, or purple in the presence of green (and vice-versa) was reinforced and any other color relation punished.

*Figure 5*. Schematic illustration of the two types of trials during the evaluative conditioning phase of Experiment 5. During one type of trial (*left*) CS1 and positive USs were presented in colors that had previously been symbolically matched. During a second type of trial (*right*) CS2 and negative USs were presented in colors that had also been symbolically matched during the color training phase.



**Supplementary Materials**

**Experiment 1**

**Exploratory questions**

*Color Matching memory*. A color memory score was calculated that ranged from 0 to 2. A 0 score indicated that participants responded incorrectly on both the CS1 and CS2 color memory question (*N* = 23); a 1 score indicated at least one correct response (*N* = 6) and a 2 score indicated that both questions were answered correctly (*N* = 74). We conducted a moderation analysis to determine if this color memory score qualified the impact of our manipulation on IAT scores and self-reported ratings. We found that color memory moderated the impact of the manipulation on both implicit and explicit attitude change, *b* = .27, *p* < .001 and *b* = .34, *p* < .001. In both cases, the effect of the manipulation increased as a function of correct color memory.

*Contiguity awareness*. 65% of participants (*N* = 67) responded correctly to the contiguity questions about CS1 and CS2 (i.e., they indicated that “*One word always had a positive meaning and the other one a negative meaning*” when asked about the USs that were simultaneously presented with a CS). 23 participants had contiguity responses in which at least one of the two answers (CS1- and CS2-US contiguity) indicated the opposite contiguity (e.g., “*Both words always had a negative meaning*” when the CS was actually paired with positive and negative stimuli). Six participants did not remember at least one of the two contiguities. We noticed however that another type of incorrect contiguity response emerged for a considerable proportion of participants (23%): 24 participants remembered that both the CSs were paired with two US stimuli of the same valence (i.e., “*Both words always had a negative meaning*” for CS matched in color with negative stimuli vs. “*Both words always had a positive meaning*” for CS matched in color with positive stimuli), and 1 participant did so for one of the CS while they responded correct to the other one. Comparing the effect of the manipulation across participants with either correct contiguity or color-matching-driven contiguity memory resulted in a significant interaction only on self-reported ratings, *F*(2, 100) = 8.91, *p* < .001, η2partial **=** .15, such that the EC effect was stronger for participants with color-matching-driven contingency. Therefore, participants memory concerning the valence of the two USs presented on screen was influenced by the valence of the US that matched the color of the CS on that trial. This suggests that, at least for some participants, color matching influences evaluative scores by influencing how stimuli-pairings are processed.

*Hypothesis (color) and influence Awareness.* We looked at participants’ response to the color awareness question (i.e., “*During the first part of the study, did you notice that the color of MORAG and STRUAN switched to the same color as either positive or negative words*?”). We found that 28% of the participants did not notice the fact that the CS and US shared a color in the second half of the trial). We re-ran the analyses considering hypothesis awareness as an additional factor in a 2x2 ANOVA. We found no significant interaction for IAT scores, *F*(1, 101) < 1, *p* = .32, but a significant interaction for self-reported ratings, *F*(1, 101) = 9.36, *p* = .003, η2partial **=** .09. Among participants who noticed the shared CS-US color, we found that 66% reported that the color switching influenced the way that they evaluated the CSs, whereas for the remaining 34%, shared color did not. Results confirmed that for those who explicitly indicated they have been influenced by the color switching (*N =* 49), the impact of our manipulation was stronger on both IAT scores, *F*(1, 48) = 40.46, *p* < .001, η2partial **=** .46, and self-reported ratings, *F*(1, 48) = 189.80, *p* < .001, η2partial **=** .80.

*Demand*.

*Reactance.* XX participants (XX%) self-reported that they were reactant on the self-reported ratings and XX (XX%) on the IAT. The exclusion of reactant participants did not influence the magnitude of the EC effects on either implicit or explicit measures.

**Experiment 2**

**Exploratory questions.**

*Color Switch contingency memory*. We assessed if participants were aware of the valence of the US that switched color during each trial. We calculated a color switch contingency memory score, ranging from 0 to 2, based on responses to the following question: “In the first part of the experiment, when CS1[CS2] appeared on the screen, which words switched to a different color”. A 0 score indicated that people responded incorrectly to both CS1 and CS2 (*N* = 28), 1 indicated at least one correct response (*N* = 11) and 2 indicated that both the questions were answered correctly (*N* = 64). If we code individuals as having passed or failed that test, and add this factor to a one-way ANOVA, we see a significant effect for explicit,  *F*(1, 102) = 9.42, *p* = .003, and implicit evaluations,  *F*(1, 102) = 17.36, *p* < .001, such that both evaluations increase in magnitude, in a negative direction when people can correctly indicate the valence of the US that changed color during a given trial.

*CS-US* *contiguity awareness*. We found that 65% of participants (*N* = 67) responded correctly to the questions about the CS-US contingencies (i.e., they selected “both positive and negative words” when presented with the following question: “In the first part of the experiment was CS1[CS2] presented together with… positive words only, negative words only, both positive and negative words, neither positive or negative words, or I don’t remember”). Of the remaining 35% (n = 36), six (6%) responded to one of the CSs correctly and the other incorrectly. The other 30 participants responded incorrectly on both questions, either indicating that they could not remember (n = 10) or answering both questions incorrectly. Coding participants as either having passed or failed the memory test, and adding this as a factor in a one-way ANOVA, revealed a descriptively larger effect for explicit,  *F*(1, 102) = 3.89, *p* = .05, but not implicit scores,  *F*(1, 102) = 1.10, *p* = .30, when participants were aware of the contingency between CSs and USs.

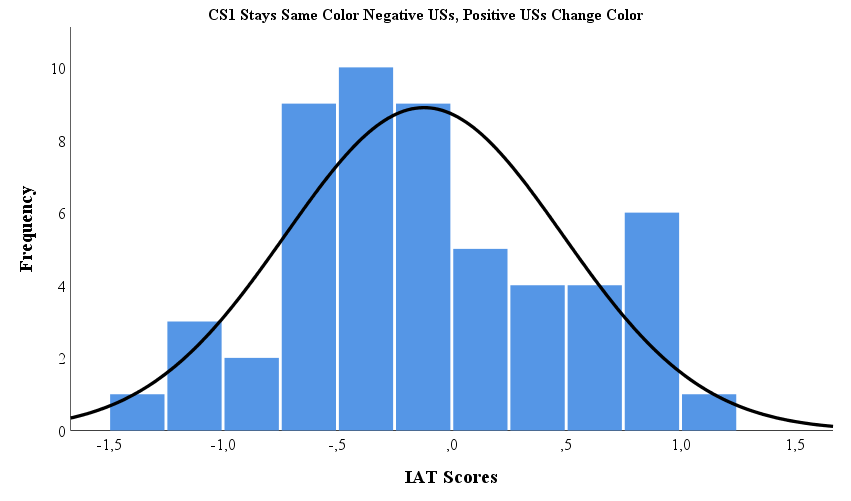
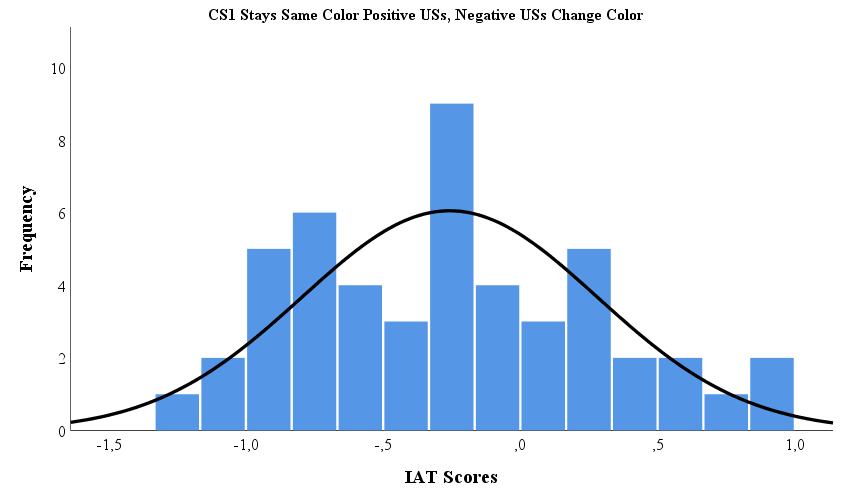
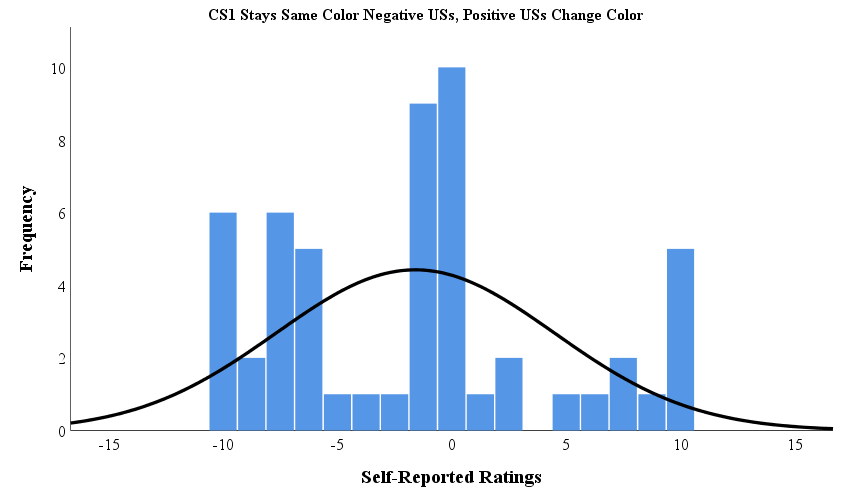
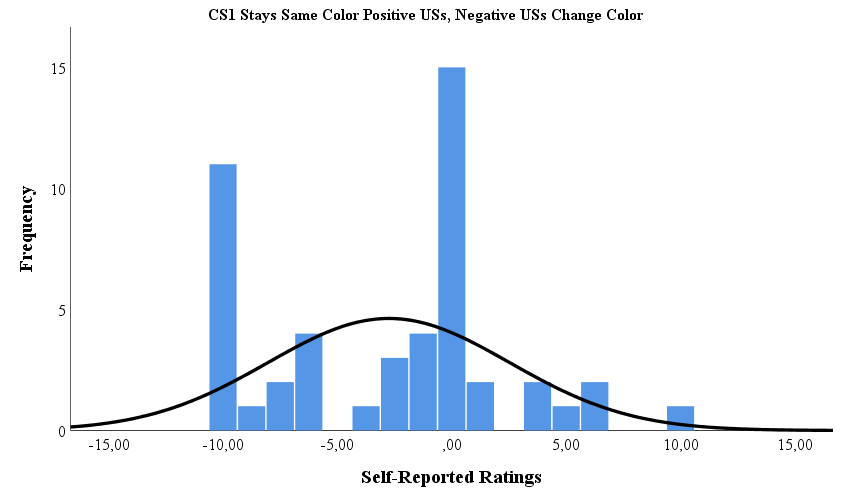
*Hypothesis (color) and influence Awareness.* We looked at participants’ response to the color awareness question (i.e., *During the first part of the study, with the colored words, did you notice that the color of one of the two words presented on the right side of the screen switched, while the word on the left side of the screen (MORAG/STRUAN) stayed the same?*). We found that 87% of participants did notice this whereas 13% did not. We re-analyzed the data considering hypothesis awareness as a factor in a one-way ANOVA. We found no interaction on either implicit,  *F*(1, 102) = 0.65, *p* = .42, or explicit evaluations, *F*(1, 102) = 2.79, *p* = .10. Among participants who noticed the color switch, we found that only 53% of them reported that the color switching influenced the way they evaluated the CSs, whereas the remaining 47% said it did not. Fifty participants explicitly reported being influenced by the color switching. The impact of our manipulation was stronger on both implicit,  *F*(1, 102) = 4.67, *p* = .03, and explicit attitude change,  *F*(1, 102) = 7.21, *p* = .008, again in the negative direction (i.e., a preference for CS2 over CS1).

*Demand*. We had 12% of demand compliant participants for the explicit measures and 15% for the IAT. The exclusion of these participants did not affect the magnitude of implicit or explicit attitude change. If anything both effects become stronger in the negative direction.

*Reactance.* We had 16% of reactant participants for the explicit measures and 10% for the IAT. The exclusion of reactant participants did not affect the magnitude of implicit or explicit attitude change. If anything both effects become stronger in the negative direction.

*Table 1*. Percentage of participant who produced an IAT score or whose self-reported ratings indicated a relative preference for CS1 over CS2, CS2 over CS1, or no preference for either stimulus, as a function of Color Matching in Experiment 2.

|  |  |  |  |
| --- | --- | --- | --- |
| **Color Match** | **Relative Stimulus Preference** | | **No Preference** |
| *Self-Reported Ratings* | CS1 over CS2 | CS2 over CS1 | Neither CS1 or CS2 |
| CS1 Same Positive US, Negative US Changes |  |  |  |
| CS1 Same Negative US, Positive US Changes |  |  |  |
| *IAT Scores* |  |  |  |
| CS1 Same Positive US, Negative US Changes |  |  |  |
| CS1 Same Negative US, Positive US Changes |  |  |  |



**Experiment 3**

**Exploratory questions.**

*Color Switch contingency memory*. 30 people responded incorrectly to both CS1 and CS2, 1 indicated at least one correct response, and 66 indicated that both the questions were answered correctly. When coding these individuals as having passed or failed that test, and adding this factor to a one-way ANOVA, we did not find an effect for implicit, *F*(96) = 0.01, *p* = .93, nor explicit evaluations, *F*(96) = 1.14, *p* = .29.

*CS-US* *contiguity awareness*. We found that 72% of participants (*N* = 70) responded correctly to the questions about the CS-US contingencies (i.e., they selected “both positive and negative words” when presented with the following question: “In the first part of the experiment was CS1[CS2] presented together with… positive words only, negative words only, both positive and negative words, neither positive or negative words, or I don’t remember”). Of the remaining 28% (n = 27), two (7%) responded to one of the CSs correctly and the other incorrectly. The other 25 participants responded incorrectly on both questions, either indicating that they could not remember (n = 5) or answering both questions incorrectly. Coding participants as either having passed or failed the memory test, and adding this as a factor in a one-way ANOVA, revealed no effect for explicit, *F*(96) = 0.63, *p* = .43, or implicit scores, *F*(96) = 1.32, *p* = .25.

*Hypothesis (color) and influence Awareness.* We looked at participants’ response to the color awareness question (i.e., *During the first part of the study, with the colored words, did you notice that the color of one of the two words presented on the right side of the screen switched, while the word on the left side of the screen (MORAG/STRUAN) stayed the same?*). We found that 91% of participants did notice this whereas 9% did not. We re-analyzed the data considering hypothesis awareness as a factor in a one-way ANOVA. We found no interaction on either implicit, *F*(96) = 0.04, *p* = .84, or explicit evaluations, *F*(96) = 0.96, *p* = .33. Among participants who noticed the color switch, we found that only 44% of them reported that the color switching influenced the way they evaluated the CSs, whereas the remaining 56% said it did not. 41 participants explicitly reported being influenced by the color switching. The impact of our manipulation was strong on implicit attitude change, *F*(96) = 6.32, *p* = .02, but not on explicit attitude change, *F*(96) = .27, *p* = .61, as participants who were aware about the influence showed a stronger implicit preference for CS1 over CS2.

*Demand*. We had 23% of demand compliant participants for the explicit measures and 19% for the IAT. The exclusion of these participants did not affect the magnitude of implicit or explicit attitude change.

*Reactance.* We had 14% of reactant participants for the explicit measures and 11% for the IAT. The exclusion of reactant participants did not affect the magnitude of implicit or explicit attitude change.

**Experiment 4**

**Exploratory questions.**

**Experiment 5**

**Exploratory questions.**

*CS-US contingency awareness.* We found that 75% of participants (*N* = 134) responded correctly to the two questions about the CS-US contingencies (i.e., they selected “One word always had a positive meaning and the other a negative meaning” when presented with the following question: “In the second part of the experiment we presented CS1 [CS2] along with two other words. Did those two other words always have a positive meaning, have a negative meaning, or did one word always have a positive meaning and the other a negative meaning?”). Of the remaining 25% (n = 45), nine (5%) responded to one of the CSs correctly and the other incorrectly. The other 36 participants responded incorrectly on both questions, either indicating that they could not remember (n = 3) or answering both questions incorrectly. Coding participants as either having passed or failed the contingency awareness test, and adding this as a factor in a one-way ANOVA together with Color Matching did not reveal any interactions for explicit, *F*(1, 177) = 0.79, *p* = .38 or implicit scores, *F*(1, 177) = 1.34, *p* = .25.

*CS-US Color contingency awareness*. Participants were asked which colors were associated with the CSs and USs (i.e “In the second part of the experiment, when CS1 [CS2] appeared on the screen with two other words, what color was CS1 [CS2] presented in:” or “In the second part of the experiment, when CS1 or CS2 appeared on the screen, what color were the POSITIVE WORDS presented in:”). 82 participants (46%) correctly matched all the colors with their associated CSs and USs. Of the remaining 97 participants, 50% responded correctly to at least two questions and 34 of them had no correct responses. 11% of all participants indicated not remembering any of the associations. If we code individuals as having passed or failed this test, and add this factor to a one-way ANOVA together with Color Matching, we did not see a significant interaction effect for explicit, *F*(1, 177) = 0.20, *p* = .65, and implicit evaluations, *F*(1, 177) = 0.002, *p* = 0.96.

*Color contingency awareness*. Participants were asked which colors were associated with one another (i.e. “Think back to the first part of the experiment where you learned about the relationship between colors. What color was BLUE related to:”). A color contingency score that ranged from 0 to 4 was calculated. The majority of participants (N = 142) was able to link all four colors (blue, yellow, green and purple) correctly to one another (79%). Only 13 participants scored 0 and were therefore unable to correctly identify the relationship between colors.

*Hypothesis (color) and influence Awareness.* We looked at participants’ response to the color awareness question (i.e., “*Think back to the first part of the experiment. During that part of the study, we trained you to relate Blue to Yellow and Green to Purple. In the second part, we presented one of the words on the left in blue (or green) and the words on the right in yellow or purple. Did you notice this during the study?*”). We found that 84% of participants did notice this whereas 16% did not. We re-analyzed the data considering hypothesis awareness as a factor in a one-way ANOVA together with Color Matching. We found a significant interaction effect on explicit evaluations, *F*(1,177) = 7.51, *p* = 0.007, η2partial **=**0.04**,** such that explicit attitudes towards CS1 [CS2] were stronger for participants that noticed how the color relationship from the first part related to the color arrangement in the second part (*M* = 2.73, *SD* = 4.20), compared to participants that did not notice this procedure (*M* = 0.44, *SD* = 2.89). In case of the implicit measure, we did not find such an interaction effect, *F*(1,177) = 0.17, *p* = 0.68. Among participants who noticed this particular procedure, we found that only 51% of them reported that the color matching influenced the way they evaluated the CSs, whereas the remaining 49% said it did not. Eighty-two participants explicitly reported being influenced by the color matching. As was the case with hypothesis awareness, influence awareness shaped the impact of our manipulation on explicit attitudes, as demonstrated by the significant interaction effect with color matching, *F*(1, 177) = 26.63, *p* < 0.001, η2partial **=**0.13. This suggests that participants who believed this color matching influenced their subsequent attitude towards CS1 [CS2], in fact showed the strongest explicit EC effect (*M* = 4.00, *SD* = 4.58) compared to participants who did not believe it had an influence (*M* = 0.98, *SD* = 3.03). In case of the implicit attitudes, no such interaction effect was found, *F*(1, 177) = 0.68, *p* = 0.41.

*Demand.* We had 19% of demand compliant participants for the explicit measures and 14% for the IAT. The exclusion of these participants did not affect the magnitude of implicit or explicit attitude.

*Reactance.* We had 15% of reactant participants for the explicit measures and 8% for the IAT. The exclusion of reactant participants did not affect the magnitude of implicit or explicit attitude.

*Writing down manipulation check.* We had 4% of participants that indicated writing down what happened in order to help them figure out what was going on.

1. We distinguish target and source feature from target and source objects because the target and source features can either be part of the same object or they can be part of different objects (for a more detailed treatment see De Houwer et al., *in prep*). In the case of shared features effects the source and target objects are typically different. [↑](#footnote-ref-1)
2. The term ‘transformation’ highlights that a source object can give rise to assumptions about the features of a target object, and that the latter’s features can change in ways that are similar or different to those of the former. To illustrate, consider our previous example of the halo effect wherein features of a source object (how attractive a person is) influence assumptions about target object features (how intelligent a person is). It may be that an assimilative halo effect emerges for men (the more attractive the male the more intelligent he is considered) whereas contrast effects emerge for women (the more attractive the female the less intelligent she is considered). The term transformation captures both of these possibilities. [↑](#footnote-ref-2)
3. In-line with De Houwer, Gawronski, and Barnes-Holmes (2013) we define attitudes as the study of evaluation which refers to the impact of stimuli on evaluative responses. Thus attitudes are conceptualized as effects (i.e., the impact of the environment on evaluative responses). [↑](#footnote-ref-3)
4. The main aim of this paper is to introduce a new perspective (shared features effect) and to provide an initial test of its basic properties and boundary conditions. We have purposefully described this phenomenon using abstract and theory-neutral terms that make no mention of the potential mental mechanisms that may mediate it. This is not to say that our research does not have implications for the cognitive level of explanation – it does. However, we have opted to unpack these theoretical issues in the General Discussion after we have established and manipulated the basic effect itself (for reasons why see De Houwer, 2011). [↑](#footnote-ref-4)
5. Note that additional exploratory analyses were performed on the exploratory questions for each of the experiments reported in this paper. Interested readers can find these in the Supplementary Materials section. [↑](#footnote-ref-5)
6. We define “symbol” and “symbolic properties” at a very abstract level. On the one hand, we view symbols as stimuli which are in some ways functionally substitutable for other stimuli in the environment (i.e. something that can stand for something else) and symbolic properties as properties of stimuli that are not determined by their mere physical characteristics. By operating at a high level of abstraction, we hope to obtain new insights and reach consensus that might not be achievable when operating at the level of specific theories about learning and liking. That said, there are theories at both the functional (e.g. Hayes et al., 2001) and mental levels (e.g. Deacon, 1997) which offer insight into the origins and nature of symbols and symbolic properties which seem compatible with the general position we forward here. [↑](#footnote-ref-6)
7. In this way the term ‘inference’ describes the outcome of the computation process rather than the computation process itself. The computational process of inferential reasoning can occur on the basis of many different inference rules (e.g., rules that encode general statements about the world [if-then rules] or rules based on mere similarity [analogical mapping rules]) (for more see Van Dessel et al., 2018). [↑](#footnote-ref-7)