Running Head: SHARED FEATURES ACCOUNT

The Shared Features Principle: If Two Objects Share a Feature, People Assume Those Objects Also Share Other Features

Sean Hughes, Jan De Houwer, Simone Mattavelli, & 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 Ghent University Grant BOF16/MET\_V/002 to JDH. Correspondence concerning this article should be sent to sean.hughes@ugent.be.

**Abstract**

In this paper we introduce the *shared features principle* which refers to the idea that, when two stimuli share one feature, people often assume that they share others features. This principle can be recognized in several known psychological phenomena that were until now not appreciated in this way. To illustrate the generative power of the principle, we report five pre-registered studies, plus their meta analyses, in which participants completed an acquisition phase containing three stimuli: a neutral target, a positive source, and a negative source. Our results indicate that behavioral intentions, implicit evaluations, and explicit evaluations of a target object were influenced by the source object with which the target shared a feature. Taken together, the shared features principle appears to be general, reliable, and replicable across a range of measures in the attitude domain. We close with a discussion of its theoretical implications, relevance to many areas of psychological science, as well as its heuristic and predictive value.

*Keywords*: Shared Features, Principle, Attitudes, Implicit, Learning

The Shared Features Principle: If Two Objects Share a Feature, People Assume Those Objects Also Share Other Features

Scientific principles are valuable because they highlight commonalities amongst many different empirical phenomena. In doing so, they not only create order and offer insight (i.e., their heuristic function) but also point us towards new and previously undiscovered instances of that principle (i.e., their predictive function). In this paper, we introduce a new principle to the realm of psychology. This principle, which we refer to as the *shared features principle*, postulates that when stimuli share one feature, people will assume that those stimuli share other features as well. We first provide an overview of existing phenomena in which the shared feature principle can be recognized. We then consider the principle itself in more detail by relating it to concepts that can be applied to a wide variety of phenomena. Finally, we illustrate the predictive value of the principle by demonstrating a novel instantiation that controls for alternative factors.

Let us first consider existing phenomena that represent instances of the shared features principle. Take the minimal group effect in social psychology (Otten, 2016; Tajfel, Billig, Bundy, & Flament, 1971). Research shows that when individuals are arbitrarily assigned to the same group based on some shared feature (e.g., similar clothing item or preference for certain paintings), people assume that those individuals also share other features as well (e.g., that others will share the participant’s own traits; van Veelen, Otten, Cadinu, & Hansen, 2016). Now consider the halo effect: when two individuals share one feature (e.g., they are romantic partners) people will assume that they share other features as well (e.g., when the female partner is physically attractive people will assume that the male partner is also attractive; Sigall & Landy, 1973; also see Forgas & Laham, 2016). Likewise, in the context of stigmatization, the mere proximity effect shows that when a stigmatized and non-stigmatized person share a certain feature (e.g., physical proximity), people assume that they also share other features (e.g., a normal weight individual will be stigmatized more when they stand next to an overweight individual; Hebl & Mannix, 2003). Still another example can be found in matching effects in persuasion: 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., that the message makes more sense or that the source is more trustworthy). These assumptions subsequently increase the chance that the message will persuade the recipient, relative to situations where a mismatch exists (see Clarkson, Tormala, & Rucker, 2011).

Phenomena like this can also be found elsewhere in psychological science. For instance, in consumer and marketing psychology, research on counterfeit brands shows that these brands intentionally imitate the physical properties of (and thus share features with) high status brands in the hope this will influence assumptions about, and ultimately consumption of, the fake brand itself (e.g., assumptions that it is also high in quality, status, and worth purchasing; Phau & Teah, 2009). 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 grandfather 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 grandfather’s actions; Uhlmann, Zhu, Pizarro, & Bloom, 2012). Finally, the shared features principle can also be recognized in learning psychology. In evaluative conditioning (EC), for instance, the fact that a neutral stimulus shares spatio-temporal properties with a valenced stimulus (i.e., the occur together in space and time) often leads the former to acquire properties of the latter (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). This is also true for attribute conditioning: when an unknown person and a known athlete share one feature (spatio-temporal properties), people assume that they also share other features as well (the CS is viewed as being more athletic or healthy: Förderer & Unkelbach, 2015).

If we take a step back, set the specific stimuli and responses to the side, and search for commonalities between the above, then we see that each phenomenon involves a broadly similar situation: one where people make assumptions about the properties of one stimulus (e.g., how positive or negative a person, group, or brand product is) based on the fact that it shares some other feature (e.g., physical or spatio-temporal properties) with a second stimulus (another person, group, or brand product).

If we are correct, then there are remarkable similarities between seemingly different domains in psychological science, which offers many new opportunity for cross-fertilization. Today the aforementioned effects 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) from other intellectual domains. This does not have to be the case. In addition to illuminating previously hidden similarities and differences between psychological phenomena (heuristic value), the shared features principle also opens up entirely new avenues for study (predictive value). Before unpacking its heuristic and predictive value, we will first specify the principle itself by drawing on a recently developed conceptual framework for feature transformation effects (De Houwer, Richetin, Hughes, & Perugini, 2019).

**Feature Transformation**

Because the shared features principle applies to a wide range of known, and still to-be-discovered empirical phenomena, it is best described in abstract terms that do not refer to a specific phenomenon. To do so, we draw upon a conceptual framework recently introduced by De Houwer et al. (2019). In essence, this framework consists of four main concepts: source features, target features, source objects, and target objects. First, ‘objects’ are broadly defined as any potential stimulus or behavior. Objects can thus refer to people, animals, inanimate items, and even responses. Second, ‘features’ are defined 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).

The shared features principle focuses on 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 object that possesses the target feature is called the target object whereas the object that possesses the source feature is called the source object. The target and source objects can differ but they can also be identical. The target feature 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. Take, for instance, the halo effect, wherein source features (e.g., how attractive a person is) lead people to make assumptions about target features (e.g., how intelligent that person is). The source and target feature can belong to the same object (e.g., judging how intelligent an attractive person is) or to different objects (e.g., judging how intelligent the partner of an attractive person is; see Forgas & Laham, 2016, for a review). In studies on the halo effect, the value of a source feature (perceived attractiveness) is varied to investigate if this influences the value of a target feature (perceived intelligence).

Finally, when a source feature influences the assumptions made about a target feature, *feature transformation* is said to have taken place. The term ‘transformation’ highlights that a source feature can give rise to assumptions about the features of a target object, and that the latter can change in ways that are similar or different to the former. To illustrate, consider the halo effect wherein features of a source object (e.g., how attractive a person is) influence assumptions about target object features (e.g., how intelligent a person is). It may be that an assimilative halo effect emerges for men (attractive males are thought to be competent in most job hiring situations) whereas contrast effects emerge for women (attractive females are thought to be less competent in certain job hiring situations; see the ‘beauty is beastly effect’; Paustian-Underdahl, & Walker, 2016). The term transformation captures both possibilities.

**The Shared Features Principle**

The shared features principle provides a functional explanation for a large sub-set of feature transformation effects: it implies that the change in assumptions about the target feature is a function of the fact that the source and target object share another feature. The feature transformation framework offers a parsimonious way to conceptualize and describe both this subset of feature transformation effects, as well as other feature transformation effects (e.g., those in which assumptions about target features change because the source and target object have opposite features). Hence, the shared features principle and the feature transformation framework differ not only in their aims (i.e., explain vs. describe feature transformation effects) but also in their scope (the shared features principle focuses on a subset of feature transformation effects - namely - those that occur on the basis of shared features). The shared features principle is also most easily applied to situations where source and target objects differ, because in these cases, a distinction can easily be made between source and target objects that do or do not share features. That said, one could also conceive of situations where the source and target object are the same, and in these situations, the source and target object (by definition) share features.

Just like any scientific principle, the shared features principle does not *always* hold but does so only under certain conditions (e.g., when the shared feature is salient). In fact, one of the aims of scientific research is to uncover to moderators of the principle. Like all functional scientific principles (e.g., gravity), the principle does not specify the mechanism by which instances of the principle are brought about, nor does it assume that the same mechanism mediates all instances of the principle. Although we will speculate about the mechanism underlying the shared features principles at the end of our paper, the main aim of the paper is to introduce the principle itself and illustrate its heuristic and predictive value.

**Heuristic Value**. The shared features principle has considerable heuristic value. As illustrated above, it can be applied to a wide variety of existing phenomena that were never previously viewed as being connected (e.g., it highlights commonalities and differences between effects in person perception and counterfeit branding). The unifying nature of the principle can be further strengthened by using the terms of the feature transformation framework. Until now, the social, persuasion, marketing, moral, and learning psychology literatures each adopted different terms when describing instances of the shared features principle. As a result, there currently is a multiplicity of concepts that undermines the ability to ‘see the forest through the trees’ (i.e., to identify what is genuinely similar or different between various types of shared feature effects). For instance, evaluative and attribute conditioning research refers to conditioned (CS) and unconditioned stimuli (US), whereas operant conditioning research refers to discriminative stimuli (Sd) and reinforcers (Sr). These terms are rarely used in research on marketing, halo, moral accountability, and person perception. Rather these domains employ idiosyncratic terms that typically refer to the specific properties of the features and objects being studied (e.g., the status of products).

The feature transformation framework in general, and shared features principle in particular, circumvent this issue by providing a common or ‘universal’ language that allows one to describe and functionally explain many 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 example, 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 also be applied to conditioning research. Here too people learn that a source object (US) 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, 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, stigmatizing, prejudice, branding, and so on.

In short, the shared features principle 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 many effects involve 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.

**Predictive Value**. The shared features principle also has predictive value. For instance, although the role of shared features has been implicitly or explicitly recognized in certain domains, it has not been recognized in many others. 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 feature with one another, and in 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 into question – simply our prior explanation of those effects. In other words, we are not questioning that stimulus pairings can lead to changes in liking. Rather we are re-conceptualizing stimulus pairings as just one way to induce a shared feature effect. This new perspective leads to the prediction that EC-like effects can also be found when stimuli share a feature other than their spatio-temporal presence (e.g., the color in which stimuli are presented). Verifying this prediction would support the idea that EC is just one instance of a much broader class of share features effects and would illustrate the predictive power of the shared features principle.

**The Current Research**

With the above in mind we carried out five studies. Each employed a broadly similar format which we will preview here. We first asked participants to complete an acquisition phase. During this phase a series of trials were presented wherein three stimuli simultaneously appeared onscreen: a positive source object, a negative source object, and a neutral target object. We then manipulated the extent to which the target object shared a feature with a certain source object. In Experiments 1-3 the shared feature was the color in which stimuli were presented: half of the trials presented the neutral target object in the same color as the positive source whereas the other half presented the neutral target object in the same color as the negative source object. In Experiment 4 the shared feature was the size of the stimuli: half of the trials presented the neutral target object in the same size as a positive source whereas the other half presented the target object in the same size as a negative source. Finally, in Experiment 5 the shared feature was pre-trained. We first established a relationship between two sets of colors (*Blue-Same-Yellow* and *Green-Same-Purple*) and then, during the acquisition phase, presented a neutral target object in either blue or green, along with a positive source in yellow and a negative source in purple. Following the acquisition phase, we assessed for attitude formation (target object evaluations) using self-report ratings and an Implicit Association Test (IAT). We added an IAT as it is assumed to reflect more automatic evaluations that can influence subsequent behavior in unique ways (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). If changes in liking are driven by simple spatio-temporal contiguity then we would expect to see similar and ambivalent evaluative responses towards both target objects (given that they were both repeatedly paired with positive *and* negative source objects). Yet if those same effects are driven by the fact that the target and source share a feature then we would expect to observe positive evaluations of one and negative of the other. If our account is correct, changes in liking should be moderated by a range of different features that are shared by stimuli.

**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). It involved a single-factor between-subjects design (*Shared Feature*: target shared color with positive vs. negative source object), 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 target object appeared in the same color as positive or negative source objects). The sample size was determined prior to data collection. We stopped data-collection when 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 = 0.05). 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, measures, and studies run. All data were collected without intermittent data analysis.

**Materials**

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

**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: acquisition phase, evaluative measures, and exploratory questions.

**Acquisition phase**. The acquisition phase consisted of three blocks of 16 trials (48 total), with each block containing two types of trials: one trial in which Target Object 1 [TO1] was eventually presented in the same color as positive source objects, and another trial in which Target Object 2 [TO2] was eventually presented in the same color as negative source objects. Specifically, three stimuli simultaneously occurred onscreen during each trial: a neutral target object (either Morag [TO1] or Struan [TO2]) along with a positive and negatively valenced adjective (positive and negative source object [SO]). All three stimuli were initially presented in white against a black background for 3000ms. On certain trials, TO1 and the positive SO both changed to the same color (e.g., blue) whereas the negative SO changed to a different color (e.g., green). On other trials, TO2 and the negative SO both changed to one color (e.g., yellow) whereas the positive SO 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 evaluations were assessed using four different semantic differential scales. On each trial, one of the two target objects 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 TO by averaging scores from these four scales.

**IAT**. We also looked for a shared features effect on implicit evaluations 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 (TO1 or TO2) 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 TO1 and TO2 into their respective categories, with TO1 assigned to the left (‘E’) key and TO2 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 TO1 and positive words using the left key and TO2 and negative words using the right key. The fifth block of 20 trials reversed the key assignments, with TO2 now assigned to the left key and TO1 to the right key. Finally, the sixth and seventh blocks (20 and 40 trials respectively) required participants to categorize TO1 with negative words and TO2 with positive words.

**Behavioral Intentions.** In addition toimplicit and explicit evaluations we also assessed if the acquisition phase altered behavior intentions towards the target objects. Participants were presented with two brand products labeled with either TO1 or TO2. They were asked to indicate which of these products they would be willing to try in a supermarket and given the following options: *I would try Morag*, *I would try Struan*, *I would try Morag and Struan*, *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 TOs-SOs), *color memory* (i.e., the color match between the TOs and SOs), 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*.

**Results**

**Analytic Strategy**

A series of Welch’s independent sample *t*-tests (along with Cohen’s *d* effect sizes and their 95% confidence intervals) were carried out on the rating and IAT data to determine whether evaluations of the two nonsense words (dependent variables) differed as a function of shared features (i.e., the fact that TO1 shared a feature [i.e., color] with positive SOs, and TO2 shared a feature with negative SOs; independent variable). Multi-nominal logistic regression models used for the behavioural intentions data to assess whether participants were more likely to choose one nonsense word over the other on the basis of the shared features. An identical analytic strategy was used in Experiments 2-5.

**Data Preparation**

**Exclusions**. We excluded data from eight participants who did not complete the entire session. The data of 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 TO1 and TO2 were first averaged and then a difference scored was calculated by subtracting scores for TO2 from TO1. Positive values indicate a relative preference for the stimulus that eventually shared a color with positive SOs over the TO that shared a color with negative SOs. 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 relative preference for the nonsense words that eventually shared a color with positive stimuli (i.e., TO1) relative to the nonsense word which eventually shared a color with negative stimuli (i.e., TO2). Negative values indicated the opposite preference (TO2 more positive than TO1).

**Hypothesis Testing**

**IAT.** IAT scores differed as a function of whether the TO shared its color with positive SOs or negative SOs, *t*(98.12) = 6.63, *p* < .001, *d* = 1.31, 95% CI = [0.88, 1.74], BF10 > 105. When TO1 shared a color with a positive SO and TO2 shared a color with a negative SO, participants showed an automatic relative preference for TO1 over TO2 on the IAT (*M* = 0.37, *SD* = 0.46). When the color contingencies were reversed, participants preferred TO2 over TO1 (*M* = -0.23, *SD* = 0.45).

**Self-reported ratings**. Self-reported ratings also differed as a function of whether the TO shared its color with positive SOs or negative SOs, *t*(98.32) = 8.33, *p* < .001, *d* = 1.65, 95% CI = [1.20, 2.10], BF10 > 106. When TO1 shared a color with a positive SO and TO2 shared a color with a negative SO, participants showed a relative preference on the self-report ratings for TO1 over TO2 (*M* = 3.33, *SD* = 4.60). When the color contingencies were reversed, participants preferred TO2 over TO1 (*M* = -4.15, *SD* = 4.48).

**Behavioral intentions**. Data from the behavioral intentions question were entered into a multinomial logistic regression with TO1 as the reference category. Only results from the TO1-TO2 comparison are relevant to the shared feature hypothesis and are reported here (i.e., hypotheses do not refer to the selections of neither word or both words). Results demonstrated that participant’s intentions towards TO1 relative to TO2 differed between the two shared feature conditions, and in a direction that was congruent with prior training, OR = 13.66, 95% CI = [3.56, 52.44], *p* < .001. The proportion of choices in favor of TO1 was higher when TO1 shared the same color as a positive SO compared to when it shared the same color as a negative SO.

**Discussion**

Result provide initial support for the shared feature account in the attitude domain. Although a neutral target object was repeatedly paired with both positive and negative source objects, it acquired the valence of the source that it shared a feature (color) with it. Specifically, target objects that appeared in the same color as positive sources were rated positively whereas target objects that appeared in the same color as negative sources were rated negatively. We also obtained strong evidence for this 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 acquisition phase during which all stimuli were initially presented in white and only later changed to the same or a different color. In Experiment 2, however, 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 source objects, while keeping the color of the other source the same as the target object. This modified design allowed us to test two 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*) entails that people’s attention is fixated on any salient change in the context. Assuming that the effects of spatio-temporal stimulus pairings are magnified when one or both of the events are salient (e.g., Rescorla & Wagner, 1972), one could argue that the change in liking for the target object could have resulted from the mere spatio-temporal pairing of the target object with the salient source object (i.e., the source object whose color changed). This alternative account entails that the observed effect was an instance of EC (i.e., a feature transformation effect due to the sharing of spatio-temporal properties) rather than a feature transformation effect that is due to the sharing of color. If so, then the target should acquire the valence of the source which switches color within the trial (i.e., the salient source stimulus), 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 acquisition phase.

**Acquisition phase**. Training once again consisted of three blocks of 16 trials (48 total), with each block containing two different types of trials: one in which TO1 stayed the same color as positive words, and another in which TO2 stayed the same color as negative words. Unlike Experiment 1, the TO and two SOs were initially presented in the same color for 3000ms. During one type of trial, TO1 and the positive SO remained in the same color (e.g., blue) whereas the negative SO changed color (e.g., purple). During the second type of trial, TO2 and the negative SO remained in the same color (e.g., yellow) while the positive SO 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 12 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**. No evidence of differences in IAT scores as a function of the color that a TO shared with a SO was found, *t*(100.85) = -1.18, *p* = .24, *d* = -0.23, 95% CI = [-0.62, 0.16], BF10 = 0.38. Participants generally showed an implicit evaluation favoring TO2 over TO1 regardless of whether a) TO1 remained in the same color as positive words and the color of negative words changed (*M* = -0.26, *SD* = 0.54), or b) TO2 remained in the same color as negative words and the color of positive words changed (*M* = -0.12, *SD* = 0.60).

**Self-reports.** No evidence of differences in self-reported ratings as a function of the color that a TO shared with a SO was found, *t*(100.98) = -1.09, *p* = .28, *d* = -0.21, 95% CI = [-0.61, 0.18], BF10 = 0.35. Participants generally showed a relative preference for TO2 over TO1 regardless of whether a) TO1 remained in the same color with the positive SO condition (*M* = -2.80, *SD* = 5.33) or b) TO1 remained in the same color as the negative SO condition (*M* = -1.58, *SD* = 6.03).

**Behavioral intentions**. Data were prepared and analyzed as in Experiment 1. Although participants intentions towards TO1 relative to TO2 differed between the two shared features conditions, they did so in the opposite direction as predicted, OR = 0.22, 95% CI = [0.07, 0.66], *p* = .007. The proportion of choices in favor of TO1 was higher when TO1 remained in the same color as a negative SO compared to when it remained in the same color as a positive SO.

**Discussion**

The findings of Experiment 2 contradicted those obtained in Experiment 1. During the acquisition phase a target and two source objects were first presented in the same color, and then one of the source objects changed to a different color. Evidence suggests that participants preferred a target when it shared a color with negative source objects more than when it shared a color with positive sources. Put another way, rather than base their evaluations on a shared feature (i.e., the fact that a source stayed in the same color as the target) participants seemed to have based their evaluations on the source object 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, the distribution of evaluations (see Supplementary Materials), 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 evaluations towards either TO1 or TO2 (for potentially different reasons), those that showed evaluations in line with the shared features account (i.e., target acquires the same valence as the SO it shares a color with) and a third group that showed evaluations in line with a salience hypothesis (e.g., target acquires the same valence as the SO which changes color). 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 relatively simple explanation for the difference in results of Experiments 1 and 2: 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 target object 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 while modifying the instructions to emphasize that the target and source objects 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 acquisition phase.

**Acquisition phase**. Prior to training 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 whereas an additional twelve did not meet the IAT criteria. This led to a final sample of 97 participants.

**Hypothesis Testing**

**IAT**. IAT scores differed as a function of the valence of the SO that had the same color as the TO, *t*(93.42) = 3.29, *p* = .001, *d* = 0.66, 95% CI = [0.25, 1.08], BF10 = 20.10. When TO1 remained in the same color as the positive SO (and the color of the negative SO changed) and when TO2 remained in the same color as the negative SO (and the color of the positive SO changed), participants demonstrated a relative preference for TO1 over TO2 (*M* = 0.21, *SD* = 0.46). When the color contingencies were reversed, participants demonstrated a relative preference for TO2 over TO1 (*M* = -0.15, *SD* = 0.59).

**Self-reported ratings.** Self-reported ratings also varied as a function of TO-SO color relation, *t*(92.94) = 5.52, *p* < .001, *d* = 1.13, 95% CI = [0.69, 1.56], BF10 > 104. When TO1 remained in the same color as the positive SO (and the color of the negative SO changed), and when TO2 remained in the same color as the negative SO (and the color of positive SO changed), participants showed a relative preference for TO1 over TO2 (*M* = 2.92, *SD* = 5.25). When the color contingencies were reversed participants demonstrated a relative preference for TO2 over TO1 (*M* = -2.85, *SD* = 5.02).

**Behavioral intentions**. Data were prepared and analyzed as in Experiments 1-2. Participants intentions towards TO1 relative to TO2 differed between the two shared features conditions, and in a way that was congruent with prior training, OR = 6.94, 95% CI = [2.03, 23.77], *p* = .002. The proportion of choices in favor of TO1 was higher when TO1 shared the same color as a positive SO compared to when it shared the same color as a negative SO.

**Discussion**

When task instructions directed attention towards (rather than away from) the shared feature, a shared features effect emerged. Specifically, targets that shared a color with positive sources were liked more than targets which shared a color with negative sources. We obtained evidence for the shared features effect on explicit, implicit, and behavioral intention measures. Importantly, the effect arose even though the TO and SO shared their color from the start of each trial. Unlike the effect that was observed in Experiment 1, the effect in Experiment 3 can therefore not be explained in terms of mere salience.

**Experiment 4**

Until now we have seen how one particular shared feature (color) comes to moderate implicit and explicit evaluations. 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 (e.g., valence). Therefore, in order to extend and generalize our findings, we swapped one shared 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 acquisition phase in which three stimuli (neutral target, positive source, negative source) were presented onscreen. This time TO1 and positive sources were presented in the same sized font whereas negative sources were presented in a differently sized font. Likewise, TO2 and negative sources shared a common sized font whereas positive sources were always presented in a different sized font. If we are correct, then targets should acquire the same valence as the sources 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 acquisition phase.

**Acquisition phase**. Training consisted of three blocks of 16 trials (48 total) consisting of two different types of trials. During one type of trial TO1 was presented in the same sized font (e.g., 8% of screen height) as positive words and a different sized font as negative words (e.g., 4% of screen height). In another type of trial TO2 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 target and source share a small [4%] font and at other times they shared a large [8%] font; see Figure 3).

**Results**

**Data Preparation**

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

**Hypothesis Testing**

**IAT.** IAT scores differed as a function of the valence of a SO that shared a common size with a TO, *t*(168.75) = 3.79, *p* < .001, *d* = 0.57, 95% CI = [0.27, 0.87], BF10 = 109.05. When TO1 was presented in the same size font as a positive SO and TO2 was presented in the same size font as a negative SO, participants showed an automatic relative preference for TO1 over TO2 on the IAT (*M* = 0.14, *SD* = 0.46). When the size contingencies were reversed, participants demonstrated a relative preference for TO2 over TO1 (*M* = -0.12, *SD* = 0.46).

**Self-reported ratings**. Self-reported ratings differed as a function of the valence of a SO that shared a common size with a TO, *t*(169.77) = 7.66, *p* < .001, *d* = 1.15, 95% CI = [0.83, 1.47], BF10 > 106. When TO1 was presented in the same sized font as a positive SO, and TO2 was presented in the same sized font as a negative SO, participants showed a relative preference for TO1 over TO2 (*M* = 2.34, *SD* = 4.12). When the size contingencies were reversed, participants showed a relative preference for TO2 over TO1 (*M* = -2.38, *SD* = 4.09).

**Behavioral intentions**. Participant’s intentions towards TO1 relative to TO2 differed between the two shared features conditions, in a way that was congruent with prior training, OR = 5.00, 95% CI = [1.91, 13.06], *p* = .001. The proportion of choices in favor of TO1 was higher when TO1 was presented in the same size font as positive sources compared to when it was presented in the same size font as negative sources.

**Discussion**

Results indicate that size can also function as a shared feature that moderates implicit and explicit evaluations as well as behavioral intentions. During the acquisition phase a target was presented with two sources – one positive and another negative. When a target was presented in the same size as positive sources it was liked more than a target that was presented in the same size as negative sources. These findings replicate those obtained in Experiments 1 and 3 and demonstrate that different types of shared features lead to the transformation of evaluations and intentions.

**Experiment 5**

In Experiments 1-4, we exclusively focused on how physical features shared by stimuli (e.g., color or size) influence behavioral intentions, implicit and explicit evaluations. Yet, as we highlighted in the introduction, there are many instances where the features that objects share are conceptual in nature. For instance, minimal group effects can emerge when people are said to share a conceptual relation with one another (e.g., are labeled as ‘overestimators’ or ‘underestimators’; e.g., Tajfel et al., 1971). Moral spill-over effects can occur when people are said to share a conceptual relation (e.g., one is the grandfather of the other; Uhlmann et al., 2012). Likewise, in halo effects, a conceptual feature shared by two people (e.g., being romantic partners) often leads others to assume that those individuals also share additional features (Sigall & Landy, 1973). Thus, the shared features principle accommodates feature transformation on the basis of physical and conceptual shared features.

In Experiment 5 we set out to experimentally model conceptual shared feature effects. Specifically, we first trained a conceptual relation between two sets of colors (e.g., *Blue-Same-Yellow* and *Green-Same-Purple*) followed bya similar acquisition phase to that used in Experiments 1-3. However, this time, we presented a target object in either blue or green along with a positive and negative source that were presented in either yellow or purple. If a target is presented in blue and a positive source is presented in yellow (along with a negative source in purple) then participants should evaluate that stimulus positively (given that blue and yellow were trained to be conceptually similar to one another in the first phase of the experiment). In contrast, if participants encounter a target in green along with a negative source in purple (and a positive source in yellow) then they should evaluate that target negatively (given that green and purple were trained to be similar to one another during the acquisition phase). Such a finding would further replicate our existing findings and expand the remit of the shared features principle by demonstrating that the shared feature moderating attitude formation can be conceptual rather than purely physical in nature.

**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, acquisition, 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 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). Prior research on stimulus equivalence learning shows that such a MTS training procedure results in people responding as if the related stimuli are equivalent (see Hughes & Barnes-Holmes, 2016a).

**Acquisition phase**. Training consisted of three blocks of 16 trials (48 total), with each block containing two types of trials: one type of trial where TO1 was presented in one color (e.g., blue), a positive word was presented in a second color (e.g., yellow), and a negative word was presented in a third color (e.g., purple). In another type of trial TO2 was presented in a fourth color (e.g., green), and the valenced words were presented in the aforementioned colors. Stimulus assignment to the various colors was counterbalanced across participants. 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 *TO-SO color contingency awareness* (i.e., if they could recall what color the TOs and SOs 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**. IAT scores differed depending on the valence of the SO that shared a color connection with a TO, *t*(184.64) = 4.98, *p* < .001, *d* = 0.73, 95% CI = [0.43, 1.02], BF10 > 104. When TO1 was presented in a color that was equivalent to the color that a positive SO was presented in, and TO2 was presented in a color that was equivalent to the color that a negative SO was presented in, then participants preferred TO1 over TO2 (*M* = 0.16, *SD* = 0.48). When the color contingencies were reversed, participants preferred TO2 over TO1 (*M* = -0.18, *SD* = 0.45).

**Self-reported ratings.** Self-reported scores differed depending on the valence of the SO that shared a color connection with a TO, *t*(179.22) = 8.54, *p* < .001, *d* = 1.25, 95% CI = [0.93, 1.56], BF10 > 106. When TO1 was presented in a color that was equivalent to the color that positive words were presented in, and TO2 was presented in a color that was equivalent to the color that negative words were presented in, participants preferred TO1 over TO2 (*M* = 3.57, *SD* = 4.99). When the color contingencies were reversed, participants preferred TO2 over TO1 (*M* = -2.18, *SD* = 4.19).

**Behavioral intentions**. Participant’s intentions towards TO1 relative to TO2 differed between the two shared features conditions, in a manner that was congruent with prior training, OR = 7.09, 95% CI = [2.92, 17.21], *p* < .001. The proportion of choices in favor of TO1 was higher when there was an overlap between TO1 and positive SO color compared to when there was an overlap between TO1 and negative SO color.

**Discussion**

Experiment 5 extends our account further and shows that conceptual shared features give rise to implicit and explicit evaluations in a similar way to merely physical shared features. Prior to the acquisition phase, two relations between colors were trained (i.e., *Blue-Similar-Yellow*, and *Green-Similar-Purple*). Thereafter a target object was simultaneously presented with two sources. Critically, the target was presented in either blue (TO1) or green (TO2), whereas positive sources were presented in yellow and negative sources in purple. Self-reported ratings, IAT effects, and behavioral intention measures all indicated that the target presented in blue was preferred relative to the target presented in green, supporting the idea that a shared conceptual feature can led to a transfer of other properties (i.e., valence).

**Meta-Analysis**

In order to determine the likelihood of observing shared features effects under other experimental conditions (i.e., to provide information about the *generality* of the effect itself), we carried out a random effects meta-analyses on Experiments 1-5. The random effects meta analyses were fitted using the metafor R package (Viechtbauer, 2010) and the maximum likelihood estimator function. A separate meta-analysis was fitted for each outcome variable (IAT, self-report ratings, and behavioral intentions). Our general strategy was to first fit a meta-analytic model and assess for heterogeneity. If heterogeneity was undesirably large then we tested for the presence of outlier experiments using metrics of both excessive influence on the meta analyzed effect size (ΔSDeffect size) and excessive influence on heterogeneity (Δτ2) via leave-one-out analyses. Studies were only labeled as outliers if results from these tests were consistent across all three outcome variables (i.e., IAT, self-reports, & behavioral intentions). Analyses indicated that Experiment 2 was an outlier on the basis of undue influence on both the meta-analyzed effect size and heterogeneity, across all three outcome variables. This is also congruent with our previous observation that the different instructions employed in Experiment 2 may have undermined the effect. As such, it was excluded and a second meta-analytic model was refit in each case. The results of both models are reported below.

**IAT**

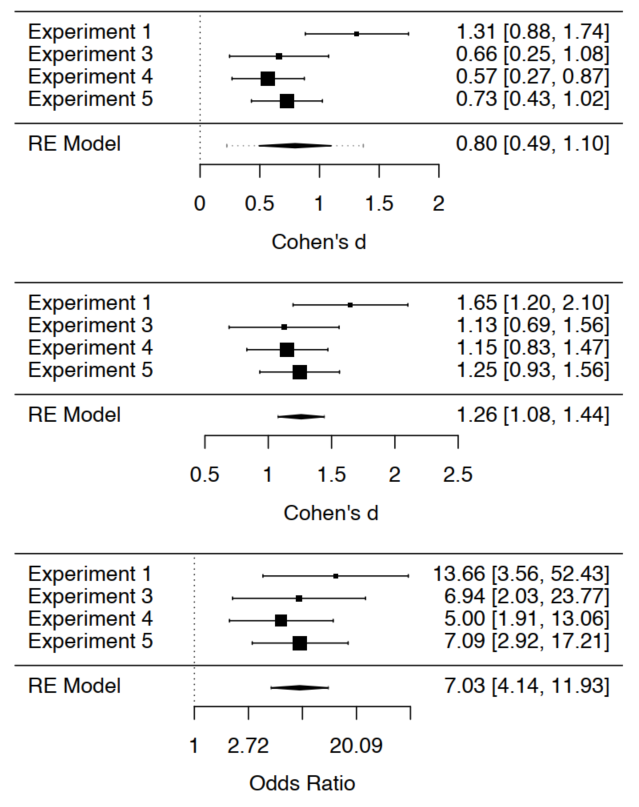
Fitting a meta-analytic model to the IAT revealed a significant effect of medium size (Cohen, 1988): *k* = 5, Cohen’s *d* = 0.61, 95% CI = [0.13, 1.08], 95% CR = [-0.49, 1.70], *p* = .012. However, results were found to contain a high degree of heterogeneity, *Q*(df = 4) = 28.46, *p* = 0.012, τ2 = 0.25, *I*2 = 88.34, *H*2 = 8.57. Following the exclusion of Experiment 2, the meta-analyzed effect size was still found to be significant, and if anything, had moved from medium to large effect size, *k* = 4, Cohen’s *d* = 0.80, 95% CI = [0.49, 1.10], 95% CR = [0.22, 1.37], *p* < .001, and now showed lower heterogeneity, *Q*(df = 3) = 8.08, *p* < .001, τ2 = 0.06, *I*2 = 65.47, *H*2 = 2.90.

**Self-Report Ratings**

Fitting a meta-analytic model to the self-report ratings revealed a significant effect of large size, *k* = 5, Cohen’s *d* = 0.99, 95% CI = [0.38, 1.60], 95% CR = [-0.46, 2.44], *p* = .002, but with a high degree of heterogeneity: *Q*(df = 4) = 48.03, *p* = 0.002, τ2 = 0.45, *I*2 = 92.53, *H*2 = 13.39. After excluding Experiment 2, the meta-analyzed effect size was still significant and now of very large size (Sawilowsky, 2009), *k* = 4, Cohen’s *d* = 1.26, 95% CI = [1.08, 1.44], 95% CR = [1.08, 1.44], *p* < .001, and with negligible heterogeneity, *Q*(df = 3) = 3.66, *p* < .001, τ2 < 0.01, *I*2 = 0.01, *H*2 = 1.00.

**Behavioral Intentions**

Fitting a meta-analytic model to behavioral intentions revealed a non-significant meta-analytic effect of medium size (Chen, Cohen, & Chen, 2010), *k* = 5, OR = 3.71, 95% CI = [0.90, 15.33], 95% CR = [0.14, 99.48], *p* = .07, and with a high degree of heterogeneity: *Q*(df = 4) = 32.06, *p* = 0.07, τ2 = 2.29, *I*2 = 88.29, *H*2 = 8.54. After excluding Experiment 2, the meta-analyzed effect size was significant and now of large size, *k* = 4, OR = 7.03, 95% CI = [4.14, 11.94], 95% CR = [4.14, 11.94], *p* < .001, and with negligible heterogeneity, *Q*(df = 3) = 1.42, *p* < .001, τ2 < 0.01, *I*2 < 0.01, *H*2 = 1.00.



*Figure 1*. Forest plots for IAT (top), self-report (middle), and behavioral intentions (bottom). Note: Experiment 2 was excluded following tests for excessive heterogeneity.

**Robustness Tests**

In our preregistered analytic plan we stated we would examine the robustness of our results to excluding participants based on several exploratory variables, particular those relating to influence and awareness. Given that each individual study lacked the power to address this question we subsequently opted for a meta-analytic approach instead. Several subsets of participants were excluded and separate meta-analytic models were refit. This served to assess the robustness of the shared feature effect when only considering participants who were (a) contingency aware, (b) not demand compliant, (c) not hypothesis aware, or (d) not influence aware. These exclusions are common in the EC literature: contingency awareness is often a necessary condition to observe evaluative effects (thus it was required; Hofmann et al., 2010), whereas the other three factors ensure that our effects were not contaminated by undesirable sources (thus they were excluded). As in the previous meta analyses, robustness tests excluded Experiment 2 as an outlier. Results demonstrate that meta-analytic effects were robust (i.e., congruent with conclusions derived from the full sample) across a) all four exclusion types and b) in each of the three dependent variables (IAT, self-reports, behavioral intentions), all *ps* < .001. There was one exception: differences in behavioral intentions were not robust to excluding influence aware individuals, *p* = .09. The general trend of evidence suggested that learning via shared features is robust to four common exclusions employed in the literature: requiring participants to be contingency aware, not demand compliant, not hypothesis aware, and not influence aware.

**General Discussion**

In this paper we introduced the shared features principle which postulates that when two stimuli share a feature people will assume that they share other features as well. This idea may underpin many phenomena in psychological science. We sought to illustrate the predictive value experimentally in the domain of attitudes, test potential boundary conditions of this effect, and demonstrate that it holds across different physical and conceptual features. Across five studies we exposed participants to an acquisition phase containing three stimuli: a neutral target, a positive source, and a negative source. We then manipulated each trial so that a target would either share a color (Experiments 1-3) or size (Experiment 4) with one of the source objects. To demonstrate that our account speaks to both physical and conceptual features, Experiment 5 created a relation between two sets of colors (*Blue-Similar-Yellow*, and *Green-Similar-Purple*) and then, during the acquisition phase, presented a target in either blue or yellow and the sources in green or purple. In all experiments except Experiment 2, we observed that liking of the target object changed in the direction of the valence of the source object with which the target object shared a feature. The unexpected results of Experiment 2 were likely a consequence of instructions that directed attention away from shared features and towards changes in stimulus features.

Taken together, our results provide strong and repeated support for the shared features principle. Changes in liking were not driven by mere contiguity but instead by the features targets and sources 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 target over the other. They also emerged regardless of the type (color or size) and nature (physical or conceptual) of feature manipulated. These conclusions were further reinforced by our mini meta-analysis where shared features moderated implicit and explicit evaluations across four of the five studies. We can therefore 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 principle itself and said little about why it actually emerges. There are two different levels at which to explain shared features effects (De Houwer, 2011; Hughes, De Houwer, & Perugini, 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: The sharing of features as a contextual relational cue.** Without going into too much detail, functional explanations are not concerned with identifying mental representations and processes. Instead they seek to relate specific effects to more general behavioral principles using terms that refer to the function of events (for a detailed treatment see Hughes & Barnes-Holmes, 2016a; De Houwer & Hughes, 2019). At this level shared features effects could be conceptualized as an instance of relational responding (i.e., a type of behavior that involves ‘responding to the relationship between stimuli’). Relational responses are 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. Take, for instance, a simple contextual cue such as a red traffic light at a busy intersection. This 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.

To illustrate, imagine that you are presented with a positive word along with an unknown word. If this pair of stimuli is accompanied by the word ‘SAME’ this may signal to you that the unknown word has the same (evaluative) meaning as the positive word. As a result you will subsequently like the unknown word more than before. In this example the word SAME functions as a relational contextual cue: it signals that a relation of similarity exists between the unknown and positive word. One could conceptualize shared physical features such as color (Experiments 1-3) and size (Experiment 4) in much the same way: as a relational contextual cue which signaled a relation of similarity between two of the three stimuli presented in an acquisition trial (a neutral target and either a positive or negative source). Once such a relationship was formed other source features were transferred to the target (valence), thus leading to a change in evaluative responding. The fact that conceptual features also moderated evaluations and intentions (Experiment 5) is consistent with past work on the effects of relational contextual cues (Hughes & Barnes-Holmes, 2016a). 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 contextual cues.

**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, that is, 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). [[2]](#footnote-2)

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 might have been these inferences which mediated subsequent evaluations and intentions. For instance, during the acquisition phase in Experiments 1-5, participants may have formed a number of simple propositions concerning the source and target objects (e.g., ‘the target is presented in green’, the positive source is presented green’, and ‘the negative source is presented in purple’). They may have also generated a number of propositions about the source and target object features (e.g., ‘this word [target] is neutral’, ‘that word [source] is positive’ and ‘that word [source] is negative’). These basic propositions may have served as the ‘raw ingredients’ for a relational inference (‘the target and source are similar in that they are both green’), and thus an inference about the target objects features (‘the source is good therefore the target is also good’). Note that these inferences were generated on the basis of both physical and conceptual 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.

**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. The shared features principle may also lead to a change in 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 principle may underpin many phenomenon throughout psychological science.

We also limited our initial efforts to the formation of attitudes. Future work could investigate whether shared features can also be used to strengthen or change existing 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 acquisition phase as we used here to generate an attitude and were then exposed to similar phase designed to reverse those initial attitudes. Researchers could vary this second task so that the target no longer shares a color with either type of source (similar to extinction in EC research; e.g., Gawronski, Gast, & De Houwer, 2015), swap the shared feature contingencies so that the target now shares a feature with the opposite source used in the first phase (similar to counter conditioning in EC research; Kerkhof, Vansteenwegen, Baeyens, & Hermans, 2011) or is exposed to the exact same contingencies as before, but between the formation and change phases, the source the target share a feature with is subjected to a US-revaluation procedure. In each case, they could examine if evaluations of the target 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 evaluations whereas directing attention away from that feature eliminated or even reversed 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 target). 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 opposite 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 an ‘opposite features effect’ waiting to be systematically explored.

**Conclusion**

In this paper we introduce the shared features principle 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 evaluations 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

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics— Simulation and Computation®,* 39(4), 860-864.

Clarkson, J. J., Tormala, Z. L., & Rucker, D. D. (2011). Cognitive and affective matching effects in persuasion: An amplification perspective. *Personality and Social Psychology Bulletin, 37(11),* 1415-1427.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6(2),* 202 -209.

De Houwer, J. (2018). Propositional Models of Evaluative Conditioning. *Social Psychological Bulletin, 13(3)*, e28046. doi:10.5964/spb.v13i3.28046.

De Houwer, J., & Hughes, S. (2019). *The Psychology of Learning: An Introduction from a Functional-Cognitive Perspective*. Textbook submitted for publication.

De Houwer & Hussey (*in prep*).

De Houwer, J., Richetin, J., Hughes, S., & Perugini, M. (2019). *On the assumptions that we make about the world around us: A conceptual framework for feature transformation effects*. Manuscript submitted for publication.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135(3),* 347-368.

Förderer, S., & Unkelbach, C. (2015). Attribute conditioning: Changing attribute-assessments through mere pairings. *The Quarterly Journal of Experimental Psychology, 68(1)*, 144-164.

Forgas, J. P., & Laham, S. M. (2016). Halo effects. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in judgement, thinking and memory* (2nd ed., pp. 276–290). New York, NY: Psychology Press.

Gawronski, B., Gast, A., & De Houwer, J. (2015). Is evaluative conditioning really resistant to extinction? Evidence for changes in evaluative judgements without changes in evaluative representations. *Cognition and Emotion, 29(5),* 816-830.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85(2)*, 197-216.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001) *Relational frame theory. A post- Skinnerian approach to language and cognition*. Kluwer Academic/Plenum.

Hebl, M. R., & Mannix, L. M. (2003). The weight of obesity in evaluating others: A mere proximity effect. *Personality and Social Psychology Bulletin, 29(1),* 28-38.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychological Bulletin, 136(3),* 390-421.

Hughes, S., & Barnes-Holmes, D. (2016a). Relational frame theory: The basic account. In R. D. Zettle, S. C. Hayes, D. BarnesHolmes, & A. Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 129-178). West Sussex, UK: Wiley Blackwell.

Hughes, S., De Houwer, J., & Perugini, M. (2016). The functional‐cognitive framework for psychological research: Controversies and resolutions. *International Journal of Psychology, 51(1),* 4-14.

Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2011). Counterconditioning. *Experimental Psychology, 58,* 31-38.

Otten, S. (2016). The Minimal Group Paradigm and its maximal impact in research on social categorization. *Current Opinion in Psychology, 11*, 85-89.

Paustian-Underdahl, S. C., & Walker, L. S. (2016). Revisiting the beauty is beastly effect: examining when and why sex and attractiveness impact hiring judgments. *The International Journal of Human Resource Management, 27(10),* 1034-1058.

Phau, I., & Teah, M. (2009). Devil wears (counterfeit) Prada: a study of antecedents and outcomes of attitudes towards counterfeits of luxury brands. *Journal of Consumer Marketing, 26(1)*, 15-27.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H., & Prokasy, W. F. (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods, 8*(2), 597-599.

Sigall, H., & Landy, D. (1973). Radiating beauty: effects of having a physically attractive partner on person perception. *Journal of Personality and Social Psychology, 28(2),* 218-224.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1(2),* 149-178.

Uhlmann, E. L., Zhu, L. L., Pizarro, D. A., & Bloom, P. (2012). Blood is thicker: Moral spillover effects based on kinship. *Cognition, 124(2)*, 239-243.

Van Dessel, P., Hughes, S., & De Houwer, J. (2018). How do actions influence attitudes? An inferential account of the impact of action performance on stimulus evaluation. *Personality and Social Psychology Review*. Advance online publication. https://doi.org/10.1177/1088868318795730.

van Veelen, R., Otten, S., Cadinu, M., & Hansen, N. (2016). An integrative model of social identification: Self-stereotyping and self-anchoring as two cognitive pathways. *Personality and Social Psychology Review, 20(1)*, 3-26.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36(3)*, 1-48.

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**

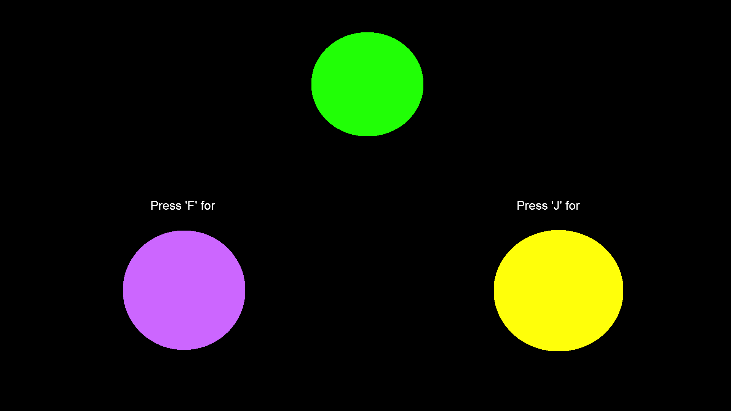
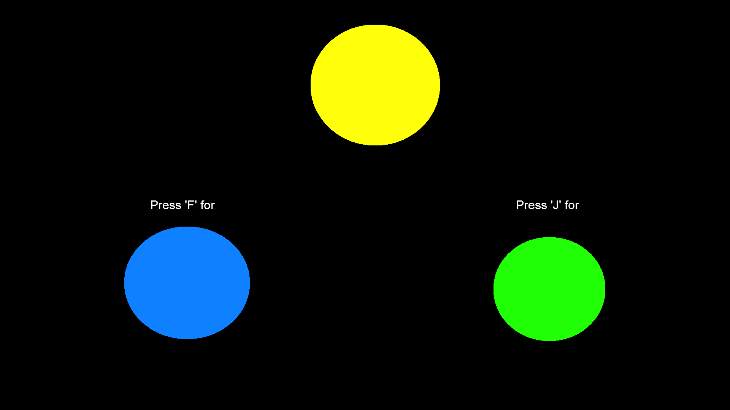
**HELL**

**STRUAN**

**SMILE**

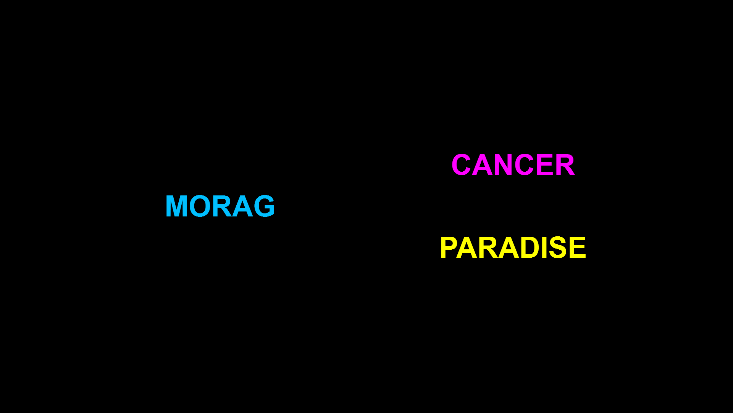
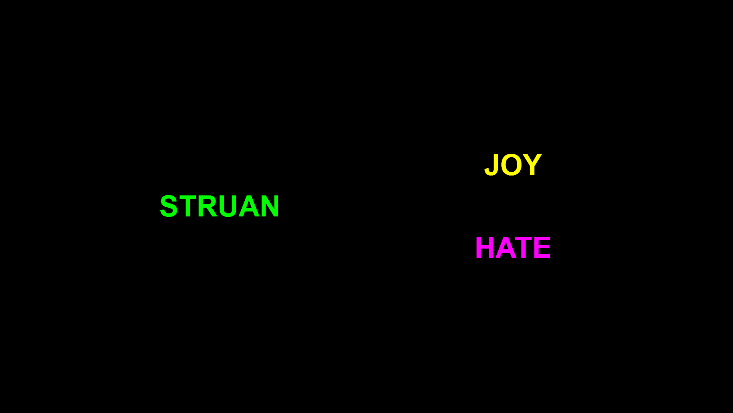


*Figure 3*. Schematic illustration of the two types of trials during the evaluative conditioning phase of Experiment 4. During one type of trial (*left*) one CS and negative USs were presented in the same size (sometimes in large and other times in small font). During a second type of trial (*right*)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

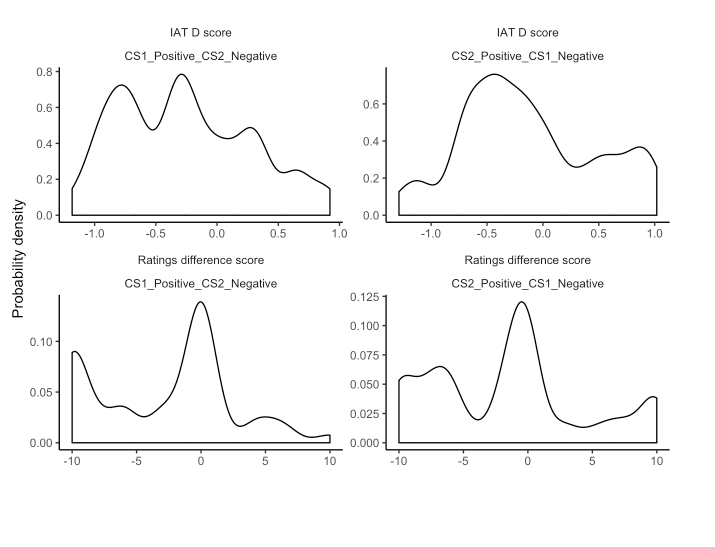


*Figure 4*. Schematic illustration of the two types of trials during the color training phase of Experiment 5. Selecting blue in the presence of yellow, or purple in the presence of green (and vice-versa) was reinforced and any other color relation punished.

*Figure 5*. Schematic illustration of the two types of trials during the evaluative conditioning phase of Experiment 5. During one type of trial (*left*) TO1 and positive USs were presented in colors that had previously been symbolically matched. During a second type of trial (*right*) TO2 and negative USs were presented in colors that had also been symbolically matched during the color training phase.



**Supplementary Materials**



*Figure 6*. Distribution of implicit (upper panels) and explicit (lower panels) evaluation scores for those in Study 2.

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., *under review*). In the case of shared features effects the source and target objects are typically different. [↑](#footnote-ref-1)
2. 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-2)