How do we learn to like and dislike?

Awareness, propositional knowledge and generalization in evaluative conditioning

Inauguraldissertation
zur
Erlangung des Doktorgrades
der Humanwissenschaftlichen Fakultät
der Universität zu Köln
nach der Promotionsordnung vom 18.12.2018
vorgelegt von

Fabia Högden
aus
Duisburg

Juni 2019

Abstract

Evaluative conditioning (EC) is concerned with the learning of likes and dislikes. In EC, neutral stimuli acquire evaluative characteristics through pairings with positive and negative stimuli. EC might be (partly) mediated by a primitive mental process that operates outside of our awareness and control. This process is often characterized as creating simple associations between mental representations of stimuli and is, therefore, often referred to as an associative process.

Associative processes are, among other characteristics, often assumed to operate without awareness and assumed to not capture specific relations between stimuli. My thesis tests whether an associative process contributes to EC in three lines of research. In a first line of research, four experiments showed that awareness of the learning stimuli is necessary to obtain EC. A second line with four experiments showed that attribute conditioning, an effect similar to EC, is sensitive to specific relations between learning stimuli. Both findings are at odds with simple associations underlying the effect. Third, five experiments studied the generalization of EC. This line of research showed that whether acquired rules influence judgments of novel stimuli can depend on characteristics of the judgment task. An important implication from this is that neglect of propositional information does not necessarily indicate the working of associative processes.

In sum, my research does not provide evidence for the contribution of an associative process to EC. Instead, I explain my findings in terms of a propositional or memory process and relate this explanation to contemporary single-process models of EC.

Erklärung über Eigenanteil

Chapter 3 beruht auf folgendem Manuskript:

Högden, F., Hütter, M., & Unkelbach, C. (2018). Does evaluative conditioning depend on awareness? Evidence from a Continuous Flash Suppression paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1641-1657. http://dx.doi.org/10.1037/xlm0000533

Die theoretische Herleitung stammt gemeinsam von mir und Christian Unkelbach. Mandy Hütter hat zur Herleitung beigetragen. Ich habe das konkrete Design entwickelt, die Datenerhebung und –auswertung umgesetzt und das Manuskript geschrieben –in enger Zusammenarbeit mit Christian Unkelbach. Mandy Hütter hat konzeptionell mitgewirkt und am Manuskript mitgearbeitet.

Chapter 4 beruht auf folgendem Manuskript:

Högden, F., & Unkelbach, C. (submitted). The role of relational qualifiers in attribute conditioning: Does disliking an athletic person make you unathletic?

Die zugrundeliegende Idee haben Christian Unkelbach und ich entwickelt. Ich habe das konkrete Design, die Datenerhebung und –auswertung umgesetzt. Das Manuskript wurde von mir und Christian Unkelbach geschrieben.

Chapter 5 beruht auf folgendem Manuskript:

Högden, F., Stahl, C., & Unkelbach, C. (in press). Similarity-based and rule-based generalization in the acquisition of attitudes via evaluative conditioning. *Cognition and Emotion.* https://doi.org/10.1080/02699931.2019.1588709

Die zugrundeliegende Idee stammt von mir und Christoph Stahl. Das konkrete Design, die Datenerhebung und –auswertung wurden von mir umgesetzt. Ich habe das Manuskript in Zusammenarbeit mit Christian Unkelbach geschrieben. Christoph Stahl hat konzeptuell an der Manuskriptgestaltung mitgewirkt

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im Oktober 2019 angenommen

Contents

Introduction

We are born with some basic skills and instincts for survival like breathing, swallowing and a need for bonding but the vast majority of skills that we need in everyday life have to be learned. Learning can be thought of as a means to optimally adapt to the environment in the course of one's life. De Houwer, Barnes-Holmes and Moors (2013) refer to this idea as ontogenetic adaptation. The concept of adaption draws on evolution theory which holds that organisms that are best adapted to the environment will survive. While this notion traditionally concerns phylogenetic adaptation, that is, the adaptation of a species over generations, learning can be thought as adaptation of an individual across its lifespan (Skinner, 1938, 1984). The ability to learn is therefore vital and it is important and interesting to study how humans learn.

Psychological science is, next to others such as neuroscience, biology, philosophy and computational science, a central discipline in understanding human learning. Psychology is concerned with understanding the mind. Hence, its objects of study are often unobservable latent mental constructs, such as intelligence or feelings and it uses observable behavior as a proxy to infer working and characteristics of the mind.[1] Because we do not know yet, how learning works on a cognitive level, I use a definition on the level of observable behavior: Learning is a change in behavior that is due to regularities in the environment (De Houwer et al., 2013).

Within this definition, different types of learning can be characterized and differentiated based on the specific regularities that affect behavior. De Houwer and colleagues (2013) outlined three different types of learning: First, learning can be due to regularities of one stimulus (over time). The most prominent examples are habituation, where a stimulus has less and less influence

[1] This definition applies in particular to cognitive psychology. Other schools of thought have focused on describing and explaining behavior which is also often named as an aim of the psychological science.

on behavior as it appears again and again over time, or its counterpart, sensitization (Mazur, 1994).

Second, learning can be due to regularities between a behavior and a stimulus. This type of learning is referred to as operant conditioning (e.g., Thorndike, 1898). The basic idea is that a behavior that is spatiotemporally close to a pleasant, rewarding stimulus, will be repeated while behavior that is close to an aversive stimulus will be avoided.

Third and central to this thesis, learning can be due to regularities between two stimuli. A very basic and important regularity between two stimuli is their co-occurrence in space and time. This type of learning can be referred to as associative learning because two stimuli are functionally associated (i.e., not associated in the sense that they are mentally connected) by their co-occurrence with each other and therefore have an effect on behavior (cf. Mitchell, De Houwer, & Lovibond, 2009). A vast part of learning research in psychology concerns associative learning. While the most prominent form of associative learning is indisputably classical conditioning, associative learning research has branched out in more narrow lines of research investigating certain paradigms, like for example, predictive learning and category learning. It is worth mentioning that the term associative learning is used inconsistently in the literature. Associative learning as described here refers to the effect that regularities between two stimuli have on behavior. Alternatively, it is used to describe a potential cognitive mechanism that mediates this behavioral effect; a primitive association formation mechanism (Mitchell et al., 2009). The cognitive mechanisms underlying the effects described here will be addressed in Chapter 2 and while I aimed to be explicit about whether I refer to the effect or the cognitive mechanism thoughout the text, the further course of the thesis will mostly use the term “associative” to refer to an association formation mechanism.

For my thesis, it is important to point out another type of learning. The three types described above refer to changes in behavior with regard to stimuli that we have direct learning experience with. In addition, changes in behavior can also come about without direct learning experience. That is, regularities in the environment regarding one stimulus can change behavior towards a different (perceptually similar or otherwise related) stimulus although the latter was not subject to any regularities. This learning phenomenon is referred to as generalization (e.g., Pearce, 1987; Spence, 1937). For sure, this is not an exhaustive list of types of learning. There are several other conceptualizations, for example, learning through observation (Bandura, 1977) or learning through insight (Köhler, 1925) that are not discussed in this thesis.

The current thesis

The focus of my thesis is learning by co-occurrence of two stimuli and learning via generalization. More specifically, I mainly studied evaluative conditioning (EC), that is, the co-occurrence of an affective and a neutral stimulus, and its generalization. The behavior that is changed by this co-occurrence of affective and neutral stimuli is the evaluation of the neutral stimuli. Typically, the neutral stimulus assimilates to the affective stimulus. That is, if a neutral stimulus is paired with a positive stimulus it will be evaluated more positively afterwards. If it is paired with a negative stimulus, in contrast, it will be evaluated more negatively afterwards (De Houwer, Thomas, & Baeyens, 2001; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). Thus, EC can be considered “learning of likes and dislikes”. A similar paradigm I used in several studies is referred to as attribute conditioning (AC). In AC, specific attributes, like athleticism, are conditioned as opposed to a general evaluation as in EC (see Chapter 4 for details).

The current definition of learning is silent with regard to the mental mechanisms involved in learning, for the above outlined reason. It mainly allows for predicting how behavior will change given certain conditions in the environment (e.g., if a neutral stimulus co-occurs with a positive one it will be evaluated more positively). While predicting behavior is a powerful asset my thesis pursues a cognitive approach. I investigate how, by which mental processes, the observable changes in behavior come about. I focused on EC, because it has been argued to be mediated by different cognitive processes than most other types of learning discussed here. More specifically, it has been suggested that EC might be mediated by primitive processes that are, for example, not susceptible to conscious control. Chapter 1 will show that learning research has started out with the idea that learning is a primitive process. In the course of time, however, evidence for the contribution of higher-order reasoning accumulated. Until, recently, purely cognitive explanations of learning fully abandoned the idea of primitive processes (e.g., Mitchell et al., 2009). Thus, it is important and interesting to study the learning processes underlying EC, because it constitutes a potential domain where such primitive processes are at work, challenging purely cognitive views of learning. Chapter 2 will explain why EC is often assumed to be mediated by different processes than other types of learning and review theories of EC. Chapter 3 and 4 will test predictions from those theories. Chapter 3 reports a series of studies on the role of awareness in EC. Chapter 4 reports studies on the effect of relational information between the co-occurring stimuli attribute conditioning. Chapter 5 reports research on the generalization of EC as a novel empirical framework to test the predictions from process theories. Chapter 6 presents an integrated explanation of the findings from the three lines of research and discusses the implications for the broader question whether EC is, unlike most other forms of learning, mediated by a primitive learning mechanism.

Chapter 1: How do we learn? Single- versus dual-process theories in learning

Theories that describe the mental process underlying learning, can be broadly categorized into those that assume that there is one way in which we learn. Opposing that view are dual-process theories that assume that there are qualitatively different mechanisms (at least two) that can produce learning (Mitchell et al., 2009). Typically, single-process theorists nowadays conceptualize learning as the results of an elaborate, reasoning-based process. Dual-process theories usually assume an additional route to learning that operates without much reasoning and cognitive resources. The notion of such a very basic learning process is intriguing because it might influence our behavior without our awareness of it. The idea of dual processes is not only present in learning but pervades many topics in psychology (e.g., Evans, 2008). Depending on the field under investigation, these two processes have been termed differently, for example explicit versus implicit, reflective versus impulsive, heuristic versus systematic or rule-based versus similarity-based (e.g., Chaiken & Trope, 1999). The common distinction in all areas is that one learning mechanism is assumed to be subject to reasoning and to require more effortful processing than the second mechanism which is largely independent of resources, intent and effort.

The notion of such a primitive mechanism is often argued to be based on learning research’s roots in behaviorism (cf. Shanks, 2010). Behaviorist learning theorists, pioneered by Thorndike (e.g, Thorndike, 1898) proposed a view of learning that focused on stimuli in the environment and observable responses to these stimuli. An example is the law of effect which was a milestone in behaviorism: If a response to a certain stimulus is rewarded, stimulus and response will be associated. Once the association was created, the stimulus would elicit the response with a higher probability than before (operant conditioning). Those (and only those) stimulus-response associations were thought to determine behavior and cognitive processes were hardly considered. Behaviorism's heyday was the early 20th century, which makes learning research one of the oldest disciplines in psychological science. Importantly, it existed before what is often referred to as the "cognitive revolution": Around 1960, researchers became increasingly interested in the mental processes underlying behavior and behaviorism's focus on stimulus-response relations was criticized, most prominently by Brewer (1974), and, hence, became less influential. Therefore, the view of learning as a phenomenon detached from cognitive processes like attention, memory and reasoning shifted towards a new, cognitive perspective on learning.

This historical overview shows that the idea of learning as something primitive that needs no higher-order reasoning is as old as learning research itself. It is therefore not surprising that it reappeared in dual-process theories on learning. At least since the cognitive revolution, however, it is generally agreed upon that cognitive processes do contribute to human behavior to some extent (e.g., Shanks, 2010). Therefore, the process debate has mainly focused on the question whether an additional, primitive route exists or not. However, since from a philosophy of science perspective, the nonexistence of a second route can never be proven, the more appropriate question is whether it is necessary and useful to assume an additional mechanism.

Learning research has to a large part been concerned with the phenomenon of classical conditioning which has, thus, been the origin of the single- versus dual-process controversy. I will now explain in more detail which processes have been argued to underlie it. I outline the Rescorla-Wagner model which formally describes the process of learning and has been a very influential model. Importantly for the topic of this thesis, the end of the next section will also show why the processes underlying EC might be different from those underlying classical conditioning.

Classical conditioning: phenomenon and process theories

The phenomenon of classical conditioning traces back to the work of Ivan Pavlov. He discovered that when a stimulus (unconditioned stimulus, US) that immanently elicits a certain response (unconditioned response, UR) is contingently presented together with a second stimulus (conditioned stimulus, CS), the second stimulus will also come to elicit that (or a similar) response (conditioned response, CR). The well-known example of classical conditioning (which is also the original stimuli Pavlov discovered the phenomenon with) is a dog that salivates (UR) when it smells food (US). When the food is repeatedly announced by (i.e. contingently presented with) a bell ring (CS), the bell ring alone, after some time, causes the dog to salivate (CR). A very common variant of the paradigm that is often studied in humans is eyeblink conditioning: An air puff (US) applied to the eye reflexively causes a blink (UR). The air puff might be paired with a light or tone in some trials and that CS will subsequently elicit a blink (CR) on its own.

Stimulus-stimulus versus stimulus-response learning. Pavlov believed that during conditioning, parts of the brain representing the US, UR and CS respectively are simultaneously active. He assumed an innate link between representations of the US and the UR and reasoned that, during conditioning, other links form that enable the CS to also elicit a similar response (Mazur, 2016; Figure 1).

One possibility is that conditioning creates a link between CS and the response. This notion is referred to as a stimulus-response (S-R) link. Alternatively, a link between CS and US, referred to as stimulus-stimulus (S-S) link, could be established. In fact, the question whether S-R or S-S learning underlies classical conditioning effects, has been one of the central research topics in learning. S-R learning is in line with a behaviorist perspective because it refers to stimuli in the environment and observable behavior. S-S learning, on the other hand, implies that there are mental representations of two stimuli. Those representations become connected and thus the CS indirectly causes a similar response as the US. This notion, as opposed to S-R learning, which can basically do without the concept of mental representations, specifies cognitive processes to some extent and can thus be conceptualized as a more cognitive model of classical conditioning than S-R learning (Mitchell et al., 2009).

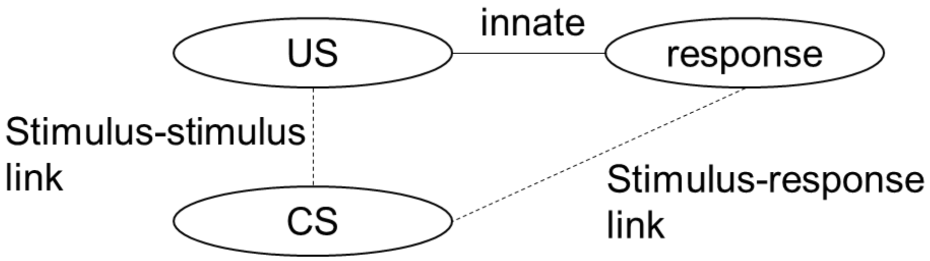


*Figure 1.*Schematic figure of Pavlov's stimulus substitution theory: the elements of conditioning, that is, the unconditioned stimulus (US), the innate response to that stimulus and the conditioned stimulus (CS) are mentally represented. While US and response are inherently associated prior to conditioning, the CS becomes associated with the US (stimulus-stimulus link, S-S) and/or the response directly (stimulus-response link, S-R) during conditioning. Whether an S-S or an S-R link forms during conditioning, has been subject to theoretical debate.

Several studies tested whether S-R or S-S learning underlies classical conditioning. A common approach in this line of research is US devaluation or revaluation. It's core feature is to weaken the link between US and response after conditioning. An aversive auditory US could be devaluated, for example, by habituating participants to the sound. The rationale is that if direct S-R (i.e. CS-response) links drive the classical conditioning effect, then it should be unaffected by postconditioning changes in the US-response link. If conditioning effects are mediated via an S-S (i.e. CS-US) link, then the CS should only elicit the response to the extent that also the US elicits it. That is, conditioning effects should become less pronounced (devaluation) or even reversed (revaluation). Rescorla (1973) was the first to use US devaluation as a test of S-R versus S-S learning. He observed weaker effects of conditioning in rats that had undergone US devaluation as opposed to control rats. Later studies showed similar effects for human subjects in fear conditioning (Davey & McKenna, 1983; White & Davey, 1989). The repeated observation that the emergence, strength and direction of the CR depends on the postconditioning nature of the US-response association, has served as evidence for S-S rather than S-R learning in classical conditioning. Thus, it seems justified to assume an internal representation and connection of some kind of the two stimuli. Concerning the nature of that connection, it has traditionally been assumed that it corresponds to a simple association that connects the mental representations of CS and US. One very influential model that formalized the process of association formation is the Rescorla-Wagner model.[2] Although, as we will see later on, the notion of simple associative links between CS and US has become subject to vigorous debate in more recent research, the model continues to be invoked in modern research on learning.

The Rescorla-Wagner model and the blocking effect. The Rescorla-Wagner model (Rescorla & Wagner, 1972) assumes that every US has a certain maximum associative potential, that is, a certain extent to which it can be predicted by other stimuli, denoted λ_{US}. During conditioning trials, a CS comes to be a predictor of the US, that is, it captures its associative potential to some extent. The predictive power of the CS for the US is denoted V_{CS}. Central to

[2] Some authors argued that the Rescorla-Wagner model is not necessarily a model of association formation but rather a functionally-descriptive model that is silent with regard to the nature of the mental representation and could thus, also be reconciled with an inferential perspective on learning (Mitchell, De Houwer, & Lovibond, 2009).

the Rescorla-Wagner idea is the concept of surprise: The amount of associative strength a certain CS acquires for a certain US in one trial (ΔV_{CS}) depends upon how surprising the occurrence of the US given the CS is in that trial. Surprise is maximal in the first trial of conditioning; the pre-trial expectation of the US following the CS is zero. Over the course of the trials, the occurrence of the US becomes less and less surprising (in a conditioning procedure where the US is always paired with the CS, i.e. contingency is 100%) and the amount of additional associative strength the CS acquires in each trial becomes less and less until it approximates the maximum associative potential of the US, that is $V_{CS} \approx \lambda_{US}$. The full equation of the Rescorla-Wagner model is the following:

$$\Delta \mathrm{V}_{\mathrm{CS(n)}} = \alpha_{\mathrm{CS}} \times \beta_{\mathrm{US}} \times (\lambda_{\mathrm{US(n)}} - \mathrm{V}_{\mathrm{all}})$$

It shows that the associative strength a CS acquires in a certain trial n is determined by the associative potential of the US at trial n that is still “unbound”, that is, not already captured by the same or different CSs in previous trials. Thus, the term $\lambda_{\mathrm{US(n)}} - \mathrm{V}_{\mathrm{all}}$ represents the extent to which the occurrence of the US is surprising. α and β denote learning rates for CS and US respectively, that are constant across trials.

The Rescorla-Wagner model has been very successful in predicting and explaining a variety of effects observed in classical conditioning but it became particularly influential because of its ability to explain blocking (Shanks, 2010). Blocking refers to the phenomenon that the conditioning of one CS with a US is impeded due to previous pairings of another CS with that US. More specifically, blocking is usually shown for components of compound stimuli: If CS_A has previously been paired with the US, then subsequent pairings of the compound CS_{AB} with the US will lead to little learning regarding CS_B. That is, if participants are asked to judge the extent to which they expect the US to occur after CS_B, expectations will be much lower when they had previously seen CS_A-US pairings than if they had not received such pretraining (Kamin, 1969).

The Rescorla-Wagner model explains blocking as follows: During CS_A-US pairings in the first phase, CS_A acquires a substantial part of the associative strength of the US. That is, if the pretraining phase is long enough, V_{CSA} will eventually approximate $\lambda_{US.}$ Thus, there is little or no associative potential left to be subsequently associated with CS_B (because V_{CSB} cannot exceed $\lambda_{US} - V_{all}$, which is very small after CS_A-US pairings). CS_A has become a near-perfect predictor of the US already, so there are no additional occurrences of the US "left to predict" by adding CS_B to the compound. In other words, the occurrence of the US is no longer surprising after CS_A-US pretraining and therefore, little learning takes place.

While the phenomenon of blocking has indisputably been very important for the study of associative processes underlying learning there are two caveats: First, the blocking effect is also conveniently explained in higher-order reasoning terms. A chain of causal reasoning underlying the blocking effect could be: "the US is as likely and strong after CS_A occurred as when CS_{AB} occurred. Thus, CS_B does not predict anything above and beyond CS_A and is therefore not a cause of the US." (De Houwer, Beckers, & Glautier, 2002). There are many studies that support an inferential account of blocking (for an overview see, Shanks, 2010). Second, and importantly, a recent publication called its scope into question because the authors could not replicate the phenomenon of blocking in 15 experiments (Maes et al., 2016).

Evidence for inferential processes underlying classical conditioning. Apart from the various lines of research aiming to explain blocking in terms of inferential reasoning reviewed by Shanks (2010), there have been a number of other approaches to show that classical conditioning is mediated by inferential processes.

First, it has repeatedly been shown that learning by instruction produces very similar results as learning by experience. Cook and Harris (1937), for example, have shown that merely instructing participants that a certain tone will be followed by a shock will lead to an increase in skin conductance (fear conditioning). It has been argued that it is implausible that verbal instructions, that are typically accompanied by elaborate, inferential reasoning, can result in primitive association formation (Mitchell et al., 2009; Shanks, 2010).

Further, researchers have studied the role of awareness of the CS-US pairings and the effect of cognitive load in conditioning (for a review see Lovibond & Shanks, 2002 and Mitchell et al., 2009). The idea of both manipulations is that association formation processes require less cognitive resources than higher-order cognition. Therefore, an associative perspective on learning would predict that conditioning effects also emerge when little resources are available.

Concerning awareness, research has repeatedly shown that awareness of the CS-US pairings during conditioning (measured via contingency memory in most of the cases) is a necessary precondition for classical conditioning effects to emerge (Dawson & Shell, 1985; Lovibond & Shanks, 2002). This is commonly interpreted as evidence for the contribution of cognitive processes in conditioning. The same conclusion is suggested by research studying classical conditioning’s sensitivity to load manipulations: Reduced attention to CS-US pairings has been shown to reduce learning (e.g. Dawson & Biferno, 1973). Also, Shanks and Darby (1998) showed that abstract rules influenced learning beyond the simple pairings of CS and US. This study will be discussed in more detail in Chapter 5.

Evaluative conditioning and the Perruchet effect as evidence for primitive learning processes. While research has been rather consistent in pointing out the contribution of cognitive processes in learning, there are some findings that are not easily reconciled with this view. The

Perruchet effect and EC are two paradigms that continue to be mentioned as evidence that is compatible with an association formation perspective on learning and have therefore been a focus of interest (Lovibond & Shanks, 2002; Mitchell et al., 2009; Shanks, 2010).

The Perruchet effect refers to the following observation: During reinforcements of a CS with an air puff US, Perruchet (1985) observed an eyeblink response to increase. When the CS was not followed by an air puff it decreased (extinction). Importantly, however, expectancy ratings of the US showed the opposite pattern: After a couple of reinforced trials, participants expected the US to a lesser extent in the subsequent trial than when the CS had not been followed by the US previously ("gambler's fallacy"). The dissociation between the two dependent variables is striking because the automatic eyeblink reflex was strongest when participants cognitively considered the likelihood of the air puff to occur to be lowest. This shows that the eyeblink response was learned independent of the cognitive expectation of the air puff and thus attests to the contribution of a noncognitive learning mechanism that follows the rules of simple association formation as described by, for example, the Rescorla-Wagner model (note, however, that the explanation of the Perruchet effect has subsequently been subject to debate, e.g., Mitchell, Wardle, Lovibond, Weidemann, & Chang, 2010; cf. Perruchet, 2015, for an overview).

The second phenomenon which is often brought forward as evidence for a noncognitive route to learning is EC (Lovibond & Shanks, 2002; Mitchell et al., 2009; Shanks, 2010). The literature on EC is much broader than that on the Perruchet effect and has thus, arguably, played a more central role in the study of learning processes. That is, EC can be understood as the single most important paradigm in justifying the assumption of a route to learning that is qualitatively different from learning based on cognitive reasoning. If there is convincing evidence for a primitive way of learning in EC this would support dual-process theories of learning. More broadly, it would call into question the monopoly of controlled cognitive processes and contribute to an understanding of the human mind and behavior that acknowledges the role of primitive processes that might be outside of our awareness and control.

I will now proceed to define EC in more detail and point out why it is considered more likely than other forms of learning, to be based on little-cognitive learning. Then, I will present theories on the processes underlying EC and describe my own empirical work that tests central predictions that follow from those theories. We will see that very similar aspects like those outlined in this subsection (awareness, cognitive load, propositional information) have played a central role in investigating the processes underlying EC.

Chapter 2: Evaluative conditioning

Evaluative conditioning (EC) is concerned with the learning of evaluative characteristics via conditioning. On an effect level, it is defined as a change in the evaluation of an initially neutral stimulus (CS) after that stimulus was (repeatedly) paired with one or more positive or negative stimuli (US). A CS that was paired with a positive US will subsequently be evaluated more positively than a CS that was paired with a negative US (De Houwer, 2007). Although the first studies describing EC effects date from the 50s (e.g. Staats, Staats, & Heard, 1959) it is mainly recently that it has become a very popular research area (cf. Corneille & Stahl, 2018).^This strong interest in EC might – apart from its theoretical importance for theories of learning explained in the previous chapter - partly be due to its vicinity to social psychological attitude research. Research on attitudes is very wide, ranging from measurement and functions of attitudes to their relation to behavior. EC can be located at the study of the formation of attitudes. Attitude researchers have come to agree that learning, especially associative learning (the effect of regularities between two stimuli, not the mechanism) plays a role in the formation of attitudes (Vogel & Wänke, 2016). Besides insights for attitude research and related areas like stereotype and prejudice research, EC has a high relevance for the understanding of phobias (i.e. clinical research), marketing and consumer research. EC is a robust phenomenon, as attested by multiple studies showing EC with a variety of stimuli and modalities (for a review see De Houwer, Thomas, & Baeyens, 2001), and is widely applicable.

Experiments studying EC usually employ a learning phase in which CS and US are paired. EC effects have been shown with forward (i.e. the CS is followed by the US), backward (i.e. the US is followed by the CS) and simultaneous pairings of CS and US; they have been shown with only one single CS-US pairing (De Houwer et al., 2001; Hofmann et al., 2010) and also both with a “one-to-one” CS-US assignment (i.e. a CS is paired with only one specific US) and “one-to-many” assignment (i.e. a CS can be paired with multiple different USs of the same valence, Stahl & Unkelbach, 2009). Hence, the EC effect seems to be rather stable across different procedural specificities.

After a learning phase, evaluative characteristics of the CSs established during conditioning are usually measured with direct and/or indirect measures of liking. As direct measurement, usually rating scales are used; indirect measures can, for example, be based on affect misattribution (affective misattribution procedure, AMP, Payne, Cheng, Govorun, & Stewart, 2005) or reaction times (evaluative priming, Fazio, Jackson, Dunton, & Williams, 1995, implicit association test, IAT, Greenwald, McGhee, & Schwartz, 1998). While some researchers hold the view that direct and indirect measures capture different types of evaluation (e.g. explicit versus implicit attitudes) a more cautious assertion would be that indirect measures are less obtrusive and thus less prone to distortion on the part of the participants than direct ratings. They mostly also measure faster and hence more spontaneous responses towards the CSs. The distinction between direct and indirect measures has brought forward some interesting patterns of findings that have fueled theorizing in EC (e.g. dissociation between direct and indirect measures). Before I turn to those process theories of EC, I will review empirical findings that are considered unique to EC and have therefore fostered the idea that underlying EC might be processes different from other learning paradigms.

Evidence distinguishing EC from classical conditioning

Levey and Martin (1987) were among the first to describe EC and term it as such because of its procedural similarity to classical conditioning. Due to this similarity it is comprehensible that it was questioned that EC is an independent phenomenon that qualitatively differs from classical conditioning (e.g., Davey, 1994; Lipp & Purkis, 2005). However, at least four points are repeatedly being raised to distinguish EC from classical conditioning: first, its resistance to extinction, second and relatedly, its sensitivity to co-occurrences as opposed to contingencies of CS and US, third its affective nature and fourth its supposed independence of CS-US memory.

Extinction, as a procedure, refers to the inclusion of CS alone trials in the learning phase - intermixed with CS-US trials or as an extinction block. In classical conditioning, the effect of extinction is a reduced CR. That is, if a tone-CS is not followed by an air puff for a sequence of trials, then the likelihood and magnitude of the eyeblink response will decrease (cf. Perruchet effect). Concerning EC effects, however, several studies observed that they are not or only very slightly decreased by extinction (Baeyens, Crombez, Hendrickx, & Eelen, 1995; Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Baeyens, Eelen, van den Bergh, & Crombez, 1989; Blechert, Michael, Williams, Purkis, & Wilhelm, 2008; De Houwer, Baeyens, Vansteenwegen, & Eelen, 2000; Díaz, Ruiz, & Baeyens, 2005). Researchers concluded that EC, in contrast to classical conditioning, does not involve expectancy learning. That is, if CS-US pairings in EC do not lead to an expectation of the US given the CS occurs, then this expectancy cannot be violated by the non-occurrence of the US in CS alone trials. In other words, the absence of the US is not surprising because participants never learned to expect the US in the first place. Importantly, more recent research has called EC’s resistance to extinction into question and provided alternative explanations for the findings listed here. They will be reviewed in the Discussion in Chapter 6. The point here is that the notion that EC does not constitute expectancy learning continues to be brought forward as a central argument why EC, unlike other forms of learning, could be mediated by primitive processes.

The notion of expectancy learning is also mirrored in the related debate about whether EC is dependent on CS-US contingencies or merely CS-US co-occurrence (contiguity). The former means that learning should be based on the probability of the US given the CS (i.e. contiguity) and the probability of the US given the CS was not present (i.e. US base rate). Contingency between CS and US can be formalized as $\mathrm{P(US|CS)} - \mathrm{P(US|\neg CS)}$ and indicates how good of a predictor the CS is for the US. While classical conditioning has repeatedly been shown to depend on CS-US contingency, a study by Baeyens, Hermans and Eelen (1993; see also Baeyens, Eelen, Crombez, & van den Bergh, 1992) has raised doubts whether the same is true for EC. Rather, EC seemed to be only driven by the contiguity, that is, the number of co-occurrences of CS and US. This finding has given rise to association formation interpretations of EC as being based on a simple mechanism that merely (and likely unconsciously and unintentionally) registers the co-occurrence of CS and US.

On a conceptual rather than an empirical level, it has repeatedly been argued that a differentiation of EC and classical conditioning might also be justified because EC is affective in nature (cf. Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012). The affective “system” (i.e. part of human thought and behavior) is often contrasted with a cognitive system and is thus ascribed attributes like being primitive, and subject to intuition rather than conscious reasoning.

Finally, there have been studies showing EC’s independence of memory for the CS-US pairings that tended to be interpreted as evidence that the processes underlying EC – unlike those underlying classical conditioning – are independent of awareness of the CS-US pairing (see De Houwer, Hendrickx, & Baeyens, 1997; Field, 2000). These findings have subsequently been subject to debate, however, and continue to do so. Hence, whether awareness of the CS-US pairings is necessary for EC to emerge is also a central focus of this thesis and the related research and conceptual debate will be discussed in more detail in Chapter 3. Again, the point here is that the findings described in this section, lead researchers to develop modern theories of the processes underlying EC that differ from the theories of classical conditioning. The next section will show that early theories of EC aimed to explain it fully in terms of primitive processes. However, it was soon acknowledged that also higher-order processes contribute to EC and, hence, dual-process perspectives emerged. Recently, there are also purely cognitive or “propositional” accounts of EC and the contemporary debate mainly takes place between proponents of dual-process and those cognitive single-process accounts.[3]

Process theories of EC

The primitive versus more elaborate mental processes potentially underlying EC are commonly referred to as associative versus propositional processes. In parallel to the single- versus dual-process debate outlined in Chapter 1, contemporary single-process models of EC aim to account for the effect by a single propositional learning process while dual-process models of EC assume an additional associative process that is qualitatively distinct from the propositional route. Gawronski and Bodenhausen (2009, 2011) recently suggested a taxonomy to characterize associative and propositional processes: They differentiated between the operating conditions and the operating principles of the learning processes (see also Bargh, 1992). Operating conditions refer to the conditions under which processes operate; operating principles refer to the way processes work. The advantage of the taxonomy is that conditions under which and principles of how certain processes operate can be gathered across different specific theories of EC. It can thus be considered a rather theory-independent approach that allows for general

[3] Recently, a functional perspective on EC emerged (e.g., De Houwer & Hughes, 2016; Hughes, De Houwer, & Barnes-Holmes, 2016) that is concerned less with the mental processes underlying it but focuses on its functional relation to other forms of attitude formation and change (e.g., persuasion). This approach is not directly relevant to the current thesis, however, because it pursues a cognitive approach.

conclusions about the characteristics of the learning processes in EC. Those general conclusions, in turn, can inform specific theories of EC – new theories can emerge and existing ones can be altered to better fit the empirical evidence. Furthermore, the taxonomy provides an apt structure along which I can present the empirical work I conducted.

The conditions under which associative and propositional processes are considered to operate are commonly equated with the characteristics of (non-) automaticity which have been put forward by conceptual analyses of automaticity. Operating principles, on the other hand, have been suggested by different accounts of EC that make specific assumptions about how learning takes place. Table 1 summarizes the operating conditions and principles (by theory) that characterize associative and propositional processes respectively and the following sections will give a more detailed overview.

Table 1. Operating conditions and operating principles (as implied by theories of evaluative conditioning) of associative and propositional processes in evaluative conditioning

	Associative process	Propositional process
Operating conditions	Operates when participants are **not aware** of CS-US pairings	Operates when participants are **aware** of CS-US pairings
	Operates when participants **do not intent to learn** about evaluative characteristics of the CS	Operates when participants **intent to learn** about evaluative characteristics of the CS
	Operates when participants direct **little attention** to or have **little cognitive resources** available to encode CS-US pairings	Operates when participants direct **sufficient attention** or have **sufficient cognitive resource**s available to encode CS-US pairings
	Operates when encoding of CS-US pairings is **outside of participant's control**	Operates when **participants can control** (i.e., stop or alter) encoding of CS-US pairings

		Associative process	Propositional process
Operating principles	Holistic account	Formation of an integrated representation of CS, US and evaluation; presentation of the CS activates that holistic representation and thus also activates the evaluation	
	Referential account	Formation of a referential association of CS and US that only depends on CS-US co-occurrence, CS presentation associatively activates US and corresponding affective attributes	
	Implicit misattribution account	Affect elicited by US presentation is implicitly misattributed to simultaneously present CS, presentation of CS directly activates that affective response	
	Conceptual categorization account	Comparison of CS and US leads to categorization of CS as member of the same affective category as US through increased salience of US-congruent features of CS	
	Propositional account		Formation of statements about the relation between CS and US, upon CS presentation, the statement is retrieved and influences judgments
	APE	(Formation and) activation of mental associations between CS and US	Validation of activated information

Operating conditions of associative and propositional processes. Associative processes are usually assumed to operate under conditions of automaticity (e.g., Gawronski & Bodenhausen, 2011). In a comprehensive conceptual analysis, Moors and De Houwer (2006) pointed out that automaticity is unintentional and, more broadly, independent of goals, uncontrolled or uncontrollable and unconscious and therefore autonomous, purely stimulus-driven, efficient and fast. A summarized version of those features of automaticity is often used in the learning and attitude literature, referred to as the "four horsemen of automaticity" (Bargh, 1994). Those are awareness, efficiency, intention and control. Applied to EC, they refer to a learning process that a) operates without or independent of participants' awareness of the pairings of CS and US during conditioning, b) depends only to a small extent on cognitive resources (at least to a smaller extent than propositional processes) and takes places c) without participants' intention to learn about evaluative characteristics of the CS and d) without participants being able to prevent that learning from happening. Propositional processes, on the other hand, are typically assumed to be at work when participants a) are aware of the CS-US pairings, b) have sufficient cognitive resources available to process them, c) intent to learn and d) are under control of learning (De Houwer, 2018).

Operating principles of associative and propositional processes. As mentioned above, operating principles that characterize associative and propositional processes have been suggested by different theoretical accounts of EC (see, e.g. Hofmann et al., 2010, for an overview). The accounts can be roughly divided into single-process associative theories that explain EC only in terms of an associative learning process, single-process propositional theories that explain EC only in terms of propositional reasoning and dual-process theories that assume that both kinds of learning processes can contribute to EC effects.

Single-process associative theories of EC. Levey and Martin (1987) were among the first to describe EC effects. They paired CSs with idiosyncratically chosen postcard USs and observed an evaluative change in the CSs in the direction of the paired US. They later explained their finding in terms of an integrated representation of CS, paired US and their evaluation that forms during conditioning. Hence, if a certain CS is subsequently presented, it activates the CS-US-evaluation conglomerate and the CS is evaluated in line with the US. In a narrow sense, this account, referred to as the holistic account of EC, is not an association formation one because it does not assume the formation of an association but rather a newly formed holistic representation of the elements involved in conditioning. It bears a lot of similarity to modern association formation theories, however, for example due to the assumed primitive nature of the process.

One of the more modern association formation accounts is referred to as the referential account of EC: A large number of experiments on properties of EC - among them many findings described in this Chapter's section "Evidence distinguishing EC from classical conditioning" - lead Baeyens and colleagues (e.g., Baeyens et al., 1992) to conclude that it is based on a simple learning mechanism that results in an association between CS and US. Unlike classical conditioning, the CS does not become a predictor of the US, however, and is not influenced by statistical contingency of CS and US but only the number of their co-occurrences. Similar to the holistic account, subsequent presentation of a CS is assumed to activate the corresponding US and its evaluation along the established association. Hence, the referential account constitutes a typical example of an S-S theory of learning. Note that, although this account describes two different routes to learning; one based on expectancies and another one based on mere references (or associations) of the CS to the US, it is not a dual-process theory in the current sense. It assumes that the expectancy type of learning underlies classical conditioning and that the referential type of learning underlies EC. Thus, it aims to account for EC effects by one process only and can hence be understood as a single-process theory of EC.

A recent explanation of EC describes the working of an associative process in terms of misattribution of affect (Jones, Fazio, & Olson, 2009; March, Olson, & Fazio, 2018). This framework is referred to as implicit affective misattribution account and is – in its core- a theory of S-R learning: During CS-US pairings, the affective response elicited by the US is implicitly (i.e. not consciously) misattributed to the CS, creating a direct link between CS and affective response. That is, participants are assumed to experience a positive or negative feeling that is actually caused by the US, but is falsely encoded as being caused by the CS because the US is, for example, less salient than the CS. Thus, in contrast to the holistic and the referential account, subsequent CS presentations will directly activate the associated affective response, bypassing the representation of the US.

As pointed out earlier, by now evidence accumulated (e.g. reviewed in Chapter 3) and there is a general agreement that propositional, cognitive processes do contribute to evaluative learning. That is, single-processes associative theories all have difficulties accounting for the broad range of findings in EC research on their own (e.g. the empirical findings presented in this thesis). To fully account for evaluative learning via EC, they need to assume an additional, more cognitive route and, hence, become a dual-process model. Thus, while their explanatory potential is limited, single-process associative theories are useful because they specify operating principles of associative processes in EC. The most influential single-process propositional theory by De Houwer (2009), on the other hand, genuinely claims to account for all findings concerning EC with one process. This account together with the conceptual categorization account will be discussed in the next subsection.

Single-process propositional theories of EC. The conceptual categorization account is an early explanation of EC effects that was put forward by Field and Davey (1999) as an alternative to traditional association formation accounts. It assumes that EC effects emerge because the CS is mentally categorized as belonging to the same affective category as the US during conditioning. They suggest that in the presence of the affective US, those features of the CS that are similar to the US become more salient and thus lead to a perception and encoding of the CS as similar to the US. The authors do not specify explicitly whether propositional reasoning is at the basis of this categorization process, therefore it cannot be classified as a propositional theory in a strict sense. But I listed it in this section because it is explicitly not an association formation theory.

The propositional account put forward by De Houwer (2009) accounts for EC effects in terms of the formation and evaluation of propositions. Propositions are logical statements that can be (believed to be) true or false or something in between. This is commonly expressed by describing propositions as having “truth values”. The validity or truth of a certain proposition is evaluated by taking into account other propositional knowledge about the world, for example prior knowledge or inferential reasoning. Thus, all factors that affect whether a proposition is believed to be true or not should also affect EC (De Houwer, 2009). According to the propositional account, participants form a proposition about the relation between CS and US during conditioning, for example “CS and US co-occur”. Subsequently, when asked to evaluate the CS, participants draw on the established propositions. The proposition that “CS and US co-occur” might translate into liking of the CS if the pre-existing proposition about the world that “similar stimuli tend to co-occur” is also taken into account (cf. Van Dessel, Hughes, & De Houwer, 2018). Hence, it might be inferred that “the CS has a similar valence as the US” (De

Houwer, 2018). Importantly, another, recent single-process perspective on EC underlines the contribution of memory-based processes to the effect (Gast, 2018; Stahl & Aust, 2018). This conceptualization will be discussed in Chapter 6 because it provides an explanatory framework for the empirical findings in this thesis.

Dual-process theories of EC. The most influential dual-process model is referred to as the associative and propositional processes in evaluation (APE) model and was put forward by Gawronski and Bodenhausen (2006). As the name suggests it was originally mainly concerned with the processes operating during evaluations of CSs, not the processes operating during learning. Associative processes are, according to the model, those that draw solely on activated associations. Propositional processes go beyond that, because they are concerned with the validity ("truth value") of the activated memory content. That is, associative processes happen first and are potentially followed by propositional processes that check the validity of the activated associations. For example, upon presentation of a CS in the judgment phase, the US it was previously paired with will be associatively activated and its evaluative connotation influences judgments. However, if the learning phase also included information on the validity of the CS-US association, for example, an instruction that CSs have the opposite valence of the USs they are paired with, the CS-US association might be discarded by propositional processes. Importantly, the model further proposes that associative processes are mainly reflected in indirect measures while direct measures reflect propositional processes. Hence, in the outlined example, an indirect evaluative measure might reflect the CS-US association (i.e., standard EC effect) while a direct measure reflects the instructed validity information (i.e., a reversed EC effect, cf. Rydell & McConnell, 2006 for a similar model that proposes two separate cognitive systems).

Summary and predictions. Associative processes are generally assumed to operate under conditions of automaticity while propositional processes operate under conditions of non-automaticity. Thus, if EC emerges from a learning phase that creates conditions of automaticity, this would be evidence that EC can emerge from associative processes. Chapter 3 presents empirical work that studies a prominent feature of automaticity namely awareness. My coauthors and I used a novel method, called Continuous Flash Suppression, with which we could present CSs to participants without their awareness. We tested in four experiments whether EC effects can emerge for CSs that participants were not aware of during conditioning.

A further important difference is that associative processes, according to all theories that describe them, only produce and/or draw on a mental link (“reference” or “association”) of CS and US. Propositional processes produce/ draw on more informative statements that can, for example, specify the way in which CS and US are related. That is, associative processes cannot differentiate between a CS that causes a US and a CS that prevents a US. Thus, if EC effects do not reflect the specific relation between CS and US this would speak for a contribution of associative processes in EC. While EC’s sensitivity to the specific CS-US relation has been shown in numerous studies (cf. Chapter 4), Chapter 4 tests this prediction in a paradigm closely related to EC, namely attribute conditioning (AC). AC is concerned with attributes that are more specific than a mere positive or negative evaluation. In AC, the US is a stimulus that has a certain attribute, for example, is athletic. After CS-US pairings, the CS will also be judged to possess that attribute.

Importantly, while AC is not affective in nature, it might also be mediated by primitive, associative processes for the following reasons: First, similar to EC, it has been shown that AC is not sensitive to extinction and blocking (Förderer & Unkelbach, 2015), suggesting that it is not a form of expectancy learning. Second, research in a similar paradigm (spontaneous trait transference, Carlston & Skowronski 2005) suggests the operation of associative as opposed to deliberate processes (see Chapter 4’s General Discussion, for details).

To study associative processes in AC, my coauthors and I tested in four experiments whether AC effects are sensitive to qualifiers that specify the relation of CS and US to be positive (i.e. they like each other) or negative (i.e. they dislike each other).

Lastly, Chapter 5 describes a series of studies in a novel domain that might allow for new predictions regarding associative and propositional processes - the generalization of EC. We tested in five experiment whether and when liking of novel stimuli is influenced by their similarity to learned stimuli or influenced by acquired rules.

Chapter 3: Does evaluative conditioning depend on awareness? Evidence from a Continuous Flash Suppression Paradigm

Abstract

The role of awareness in evaluative learning has been thoroughly investigated with a variety of theoretical and methodological approaches. We investigated evaluative conditioning (EC) without awareness with an approach that conceptually provides optimal conditions for unaware learning - the Continuous Flash Suppression paradigm (CFS). In CFS, a stimulus presented to one eye can be rendered invisible for a prolonged duration by presenting a high-contrast dynamic pattern to the other eye. The suppressed stimulus is nevertheless processed. First, Experiment 3.1 established EC effects in a pseudo-CFS setup without suppression. Experiment 3.2 then employed CFS to suppress conditioned stimuli (CSs) from awareness while the unconditioned stimuli (USs) were visible. While Experiment 3.1 and 3.2 used a between-participants manipulation of CS suppression, Experiments 3.3 and 3.4 both manipulated suppression within participants. We observed EC effects when CSs were not suppressed, but found no EC effects when the CS was suppressed from awareness. We relate our finding to previous research and discuss theoretical implications for EC.

Does evaluative conditioning depend on awareness? Evidence from a Continuous Flash Suppression Paradigm

Likes and dislikes pervasively influence human cognition and behavior. In studying how likes and dislikes emerge, evaluative conditioning (EC) has played an important role. EC is the change in liking of initially neutral stimuli (conditioned stimuli, CSs) after pairings with liked or disliked stimuli (unconditioned stimuli, USs; De Houwer, 2007). Formation of likes and dislikes via EC has been shown in different domains, with a variety of stimuli and procedural variations. EC is highly robust, widely applicable, and based on a very simple paradigm, namely the co-occurrence of two stimuli (see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010, for a metaanalytic review).

Yet, despite a substantial amount of research, there is still a lively debate about the necessary and sufficient conditions under which the pairing of two stimuli leads to an EC effect. One major question has been whether awareness of the pairings is such a requirement; that is, whether EC may or may not occur without participant's awareness of the CS-US pairings (Sweldens, Corneille, & Yzerbyt, 2014). Answering this question has important implications for theories of learning and beyond. For example, the distinction between dual- and single-process theories is present in many domains of psychology, such as persuasion (e.g., Kruglanski & Thompson, 1999; Petty & Cacioppo, 1986), person perception (e.g., Fiske & Neuberg, 1990), or reasoning (Sloman, 1996). For learning paradigms such as EC, dual-process theories distinguish between two qualitatively different kinds of learning - one that necessitates awareness and one that does not necessitate awareness (e.g., Gawronski & Bodenhausen, 2006, 2011). Single-process theories, on the other hand, assume only one learning mechanism that necessitates awareness (Lovibond & Shanks, 2002; Mitchell et al., 2009)

EC is highly informative to distinguish between these theoretical approaches. Successful EC without awareness would strongly support dual-process theories. Failures to show EC without awareness would support single-process theories; because EC is such a robust effect rooted within such a simple paradigm (mere co-occurrence; Shanks, 2005), one may even interpret such failures as evidence against dual-process theories on a larger scale. If research fails to show basic learning effects without awareness, it is unlikely to find more complex effects without awareness (cf. Mausfeld, 2003).

In the remainder of the introduction, we will review research on the role of awareness in EC and discuss its scope and intricacies. We organized it with regard to methodological approaches starting with the early correlational approaches and proceeding to experimental approaches. Then, we will present the aim and approach of our research.

Correlational Approaches to Awareness of the CS-US Pairs. Initially, research investigating the role of awareness in EC adopted a correlational approach: Participants' memory for the CS-US pairings was used as a proxy for awareness during conditioning and was correlated with the size of the EC effect on a person or stimulus level. This correlational approach yielded both evidence for EC without awareness (Olson & Fazio, 2001; Walther & Nagengast, 2006) and evidence against EC without awareness (Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Stahl & Unkelbach, 2009; Stahl, Unkelbach, & Corneille, 2009). However, studies relying on memory for pairings cannot provide conclusive evidence for EC without awareness. An EC effect without "awareness" might be due to awareness during learning or memory failure at test. Lack of memory for CS-US pairings after conditioning is no indicator of awareness for CS-US pairings during conditioning (Gawronski & Walther, 2012).

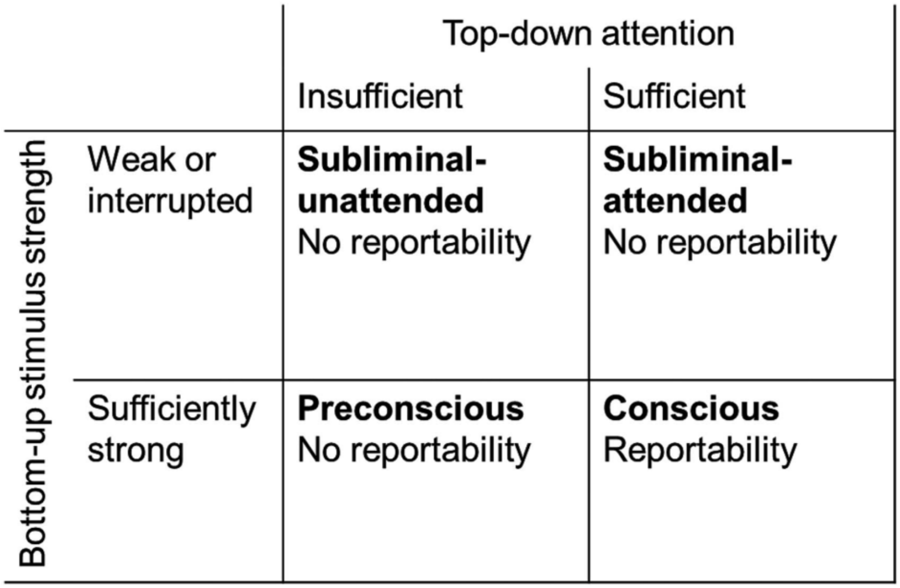
A methodological advancement was the adaptation of Jacoby's (1991) process dissociation procedure to separately assess memory of the CS-US pairs and evaluations that do not depend on CS-US memory. This approach provided evidence for EC without awareness and fueled the single- versus dual-process debate in EC (Hütter & Sweldens, 2013; Hütter et al., 2012). Yet, the process dissociation approach has similar

limitations as the correlational approaches, as it also depends on memory for the pairings at the time of measurement, which is again a proxy for awareness during conditioning. Hence, EC in the absence of memory of the CS-US pairs could be because of forgetting as opposed to learning without awareness.

Experimental Approaches to Manipulate Awareness of the CS-US Pairs. Given the discussed limitations, Gawronski and Walther (2012) encouraged experimentally manipulating awareness during conditioning. However, manipulating awareness is a methodological challenge; each approach possesses specific advantages and shortcomings.

Here we shortly discuss the most frequent approaches along a taxonomy suggested by Dehaene, Changeux, Naccache, Sackur and Sergent (2006). The taxonomy outlines different states of consciousness based on bottom-up stimulus strength and top-down attention (Figure 2). While Dehaene and colleagues discuss consciousness rather than awareness, a distinction between awareness and consciousness seems to be inconsequential for the present purpose. The operational definition we will use in the following is that participants are aware of a stimulus or consciously perceive a stimulus when they can report its presence.

As the taxonomy shows, a stimulus presumably enters consciousness when it is perceptually strong and when attention is allocated to the stimulus. When bottom-up stimulus strength is weak or interrupted, the stimulus is considered to be subliminal. If top-down attention is allocated to subliminal stimuli, they may affect behavior, for example in priming tasks (e.g., Payne, Brown-Iannuzzi, & Loersch, 2016, for a recent demonstration). When stimulus strength is sufficiently strong but there is insufficient attention allocated to the stimulus, the stimulus is considered to be preconscious. Preconscious stimuli cannot be reported but have the potential to affect behavior. In the following, we accordingly use the term awareness to describe a state of reportability, which is equivalent to the "conscious" quadrant in the taxonomy by Dehaene and colleagues (2006). We now discuss the pertinent research that may fall into the remaining three quadrants.



		Top-down attention	
		Insufficient	Sufficient
Bottom-up stimulus strength	Weak or interrupted	**Subliminal-unattended** No reportability	**Subliminal-attended** No reportability
	Sufficiently strong	**Preconscious** No reportability	**Conscious** Reportability

Figure 2. Four states of consciousness differing in bottom-up stimulus strength and top-down stimulus attention. Adapted from Dehaene et al. (2006).

Manipulating Stimulus Strength in EC. One method to reduce bottom-up stimulus strength and thereby awareness is a short stimulus presentation. The logic is straightforward: If stimuli (CSs or USs) are presented so briefly that participants cannot report seeing them, they cannot be aware of the CS-US pairings. A range of studies observed EC effects for subliminally presented USs. For example, De Houwer, Baeyens, and Eelen (1994, Experiment 1) presented positive and negative words (USs) for 28.5 ms after neutral words (CSs) and found a clear EC effect (for similar findings, cf. De Houwer, Hendrickx, & Baeyens, 1997; Rydell, McConnell, Mackie, & Strain, 2006). However, some of these studies were criticized for lack of methodological rigor (e.g., presentation times that allow stimulus perception) and other peculiarities (e.g., between-participants manipulation of valence, Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Sweldens, Corneille, & Yzerbyt, 2014). Avoiding some of these shortcomings and peculiarities, Stahl, Haaf, and Corneille (2016) recently found no EC effects for subliminally presented CSs in a set of six studies with 27 experimental conditions.

A general limitation to these studies is that the awareness manipulation relies on stimulus duration. Exposure duration impacts a multitude of processes that are necessary but not sufficient for awareness (e.g., Bar & Biederman, 1999; Moutoussis & Zeki, 2002; see Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006, for an overview). Therefore, manipulating awareness via presentation times seems to be a nonoptimal approach.

Manipulating Top-Down Attention in EC. Another method to manipulate awareness is depleting participants’ attentional resources during conditioning, for example, with a secondary task. This manipulation causes top-down attention to shift away from EC pairings while the stimuli themselves are sufficiently strong for perception. Implementations of this method obtained heterogeneous results. For example, Fulcher and Hammerl (2001; Experiment 1) depleted participants' resources by having them do math tasks presented via headphones. The control condition explicitly informed participants that stimuli are paired and they should "take note" of the pairings. They observed an EC effect in the load condition and a reversed EC effect in the control condition.

Dedonder, Corneille, Yzerbyt, and Kuppens (2010) used an auditory two-back task. Participants wearing headphones had to press the spacebar each time they would hear a number identical to the one they heard two places before during the CS-US pairings. Participants in a control condition only listened to music via headphones during conditioning. The authors observed no EC effect in the two-back task condition, while the music condition showed a clear EC effect (for similar findings, cf. Pleyers, Corneille, Yzerbyt, & Luminet, 2009). Moreover, using the adaptation of the process dissociation procedure in the same paradigm, Mierop, Hütter, and Corneille (2017) observed resource depletion to decrease the EC effect.

Depletion and load manipulations also raise theoretical concerns about the manipulation's validity. For example, when attentional load is manipulated between participants, absence of EC under depletion conditions might be because of differential goals between conditions. Bearing in mind that EC is sensitive for a focus on valence (Gast & Rothermund, 2011b) it is plausible that a secondary task distracted participants' attention from evaluative aspects of the presented stimuli, which might prevent EC effects independent of awareness concerns (Dedonder, Corneille, Bertinchamps, & Yzerbyt, 2014; Sweldens et al., 2014). Moreover, implementing them as a within-participants manipulation on a trial-by-trial basis might inflict task-switching costs which might interfere with or conceal an effect of CS-US pairings.

Dedonder and colleagues (2014) applied another method to reduce awareness in EC by presenting CSs outside of the focal gaze. According to Figure 2's taxonomy, such parafoveal

CSs are in a preconscious state because they are presented sufficiently strong but attention is focused elsewhere (foveally). They observed EC effects for CSs in the control condition that were presented foveally but not for CSs presented parafoveally. This approach, however, confounds the state of awareness with spatial CS-US proximity. As USs were always presented foveally, CSs in the control condition were located closer to the USs than CSs that were presented parafoveally. Therefore, the causal factor might be CS-US proximity (Jones et al., 2009) rather than awareness of the CS-US pairings.

In summary, a number of studies investigated the question whether EC can emerge without awareness of the CS-US pairs, but the findings diverge. Furthermore, the variety of methodological approaches and their specific shortcomings underline the difficulty of experimentally studying processes that are assumed to occur without awareness.

In addition, the study of unaware processing faces a statistical problem: finding no EC effects without awareness is based on accepting null findings (e.g., absence of EC for subliminally presented CSs). Conventional statistical analyses forbid drawing conclusions in favor of the null hypothesis. Thus, conventional analyses cannot statistically substantiate null findings. Bayesian analyses offer a solution to this issue because they allow for the quantification of evidence for the null hypothesis and, hence, for the absence of EC effects without awareness. The Bayes approach has only recently gained popularity in psychological research and has therefore only been applied in one recent study on EC (Stahl et al., 2016).

Building on the discussed research, the present research attempts to create ideal conditions for evaluative learning without awareness that avoids most of the methodological and statistical pitfalls. Please note that in the course of the debate outlined previously, the present research team has made arguments and presented empirical data for both EC only with awareness (Stahl & Unkelbach, 2009; Stahl et al., 2009) and for EC without awareness (Hütter & Sweldens, 2013; Hütter et al., 2012). Thus, we did not favor any outcome a priori, although we considered an EC effect without awareness the more interesting case, because it would have more evidential value; a null result would not refute the possibility of EC without awareness, because it is logically impossible to proof the nonexistence of a potential empirical effect. However, we also believe that accumulating evidence for failures of EC without awareness makes the possibility more and more unlikely.

Optimal Conditions for Unaware EC: Continuous Flash Suppression. Besides organizing the existing research, the taxonomy of Dehaene and colleagues (2006) provides some insights for optimal conditions under which EC without awareness might emerge. First, evaluative learning with stimuli of low bottom-up strength that are not attended seems a priori unlikely. From a functionalist perspective, one may ask why such a form of learning should exist at all and what the possible function may be. Second, evaluative learning with stimuli of low bottom-up strength that are sufficiently attended to have an a priori higher chance to yield effects, as stimuli have been shown to influence cognitive processes under these conditions (e.g., in semantic priming). Yet, as we have argued above, the typical manipulation of awareness via presentation time has a number of other problems.

We believe the most promising approach is located within the third quadrant. To realize these conditions, under which a stimulus is sufficiently strong, but does not pass the awareness threshold, we employed the method of Continuous Flash Suppression (CFS).

A typical CFS setup works as follows: A stationary stimulus carrying little visual information (e.g., black and white, low contrast) is presented to one eye, while the other eye perceives a flashing (i.e., rapidly changing) sequence of colored pixel masks (Tsuchiya & Koch, 2005). To present images to one eye and another image to the other, most experiments use stereoscopes (Figure 5). The visual system cannot merge the simultaneous conflicting input from both eyes into one coherent representation. Therefore, information from the eye receiving less informative input is usually suppressed from awareness. Suppressed stimuli are assumed to be nevertheless encoded and processed (e.g., Tsuchiya & Koch, 2005; Yang, Brascamp, Kang, & Blake, 2014), but as long as visually more informative flashing stimuli are presented to the other eye participants are not aware of the suppressed stimuli in the sense that they cannot report seeing them.

CFS is thereby a powerful method for suppression because it allows for long stimulus presentations (up to three minutes, Tsuchiya & Koch, 2005) without awareness of the stimulus. It was used effectively in numerous studies, for example to investigate visual aftereffects (e.g., Kanai, Tsuchiya, & Verstraten, 2006; Kaunitz, Fracasso, & Melcher, 2011), priming (e.g., Faivre, Berthet, & Kouider, 2012), fear conditioning (Raio, Carmel, Carrasco, & Phelps, 2012), and perceptual learning (Seitz, Kim, & Watanabe, 2009; for a review, see Yang et al., 2014). CFS is further ideal to study EC without awareness, because one may create pairing conditions that do not confound the state of awareness with stimulus exposure duration or stimulus proximity and allows for a within-participants manipulation of awareness.

Present Research. The following experiments used CFS to investigate the possibility of EC without awareness. Our approach to awareness is experimental; we do not rely on subjective self-reports or post-pairing memory measures, but aim to manipulate participants' awareness during learning. We mainly use measures of stimulus awareness as a manipulation check. Regarding the statistical considerations elaborated above, we used Bayesian analyses to obtain evidence for the nonexistence of an EC effect when necessary. There are three ways to implement EC in a CFS setup; first, one might present the pairing of a CS and US to one eye and suppress the pairing with flashes to the other eye. Second, one might suppress the US and present the CS to the other eye, thereby preventing awareness of the pairing. And third, one might suppress the CS and present the US to the other eye, also preventing awareness of the pairing.

We opted for the third option for three reasons. First, prior research has shown that high-level affective information is not processed under CFS (Yang et al., 2014; Yang, Hong, & Blake, 2010). Second, Stahl and colleagues (2016) argued that presenting USs without awareness as opposed to CSs is problematic in that "an absence of EC effect [. . .] may be attributed to a lack of affective reactions to the US, instead of a lack of learning per se" (Stahl et al., 2016, p. 1108). Research on CFS has also shown that affective stimuli can be suppressed only for shorter periods of time than nonaffective stimuli (e.g., Gayet, Paffen, Belopolsky, Theeuwes, & Van der Stigchel, 2016; Stein & Sterzer, 2012; Yang, Zald, & Blake, 2007). Tore, affective USs would interfere with suppression during learning. Third, USs are usually visually interesting stimuli and are therefore optimally suited to suppress awareness of the CSs.

We aimed to suppress the CSs by presenting them as stationary, low contrast stimuli to one eye while flashing a sequence of US photos and colored pixel masks to the other eye. The rationale is that EC effects for suppressed CSs would provide evidence for evaluative learning without awareness, thereby supporting a dual-process account of EC and dual-process theories in general (Gawronski & Bodenhausen, 2006, 2011).

EC procedures typically involve CS-US pairings that are easily perceived and processed by participants. Studying EC with CFS, however, requires presenting the US for very short durations only and interrupting their presentation with pixel masks. Please note that this problem arises for the aware, not the suppressed stimulus. Thus, the stimulus presentation alone, even without the suppression manipulation, makes the stimuli more difficult to perceive. Therefore, Experiment 3.1 established the basic EC effect in a pseudo-CFS setup omitting the suppression. That is, the presentation omitted the stereoscopes and thereby avoided the competition for awareness. Experiment 3.2, 3.3, and 3.4 included suppression by using a stimulus presentation similar to Experiment 3.1 but presenting it dichoptically (i.e., CS and US flash were simultaneously presented to different eyes using stereoscopes). We used evaluative ratings of like and dislike as a direct measure and responses in the affective misattribution procedure (AMP; Cronbach's alpha = 0.84–0.87[4]; Payne et al., 2005) as a more indirect measure.

We report all data exclusions (if any), all manipulations, and all measures in the experiments. We conducted three additional studies that are not described here: two preliminary experiments that aimed to show an EC effect in a pseudo-CFS paradigm without using suppression (like Experiment 3.1). Those experiments employed a "one-to-many" EC procedure (see Stahl & Unkelbach, 2009): one CS was paired with multiple USs. None of the experiments yielded an EC effect, and we thus did not further pursue the one-to-many EC approach. A further intermediate study aimed to manipulate suppression within participants on a trial-by-trial basis. This trial-by-trial procedure impeded suppression, however, and was therefore discarded. Experiment 3.3 and 3.4 nevertheless implement a within-manipulation of awareness using a block-wise design.

[4] The Cronbach alpha values reported here were computed from the data from Experiment 3.2. We decided to use these data because a) in Experiment 3.2 we had only two within-participant conditions (no between-participant condition), resulting in the maximum number of observations going into the reliability scores and because b) we obtained only two different alpha scores – one for the positively paired CSs and one for the negatively paired CSs - which made reporting on those scores more concise than reporting on eight or four different scores from Experiment 3.1, 3.3 or 3.4. Every CS was used eight times as a prime in the AMP. We aggregated those eight trials resulting in a score indicating the proportion of positive responses (ranging from 0 to 1) for every CS for every participant. We obtained eight proportions (for eight CSs) per participant. The four proportions of positively paired CS (four "items") and the four proportions of negatively paired CS were submitted to two separate Cronbach alpha analyses. The Cronbach alpha scores reported here are in line with scores reported by the authors who first introduced the AMP (Payne et al., 2005b).

Experiment 3.1

The aim of Experiment 3.1 was to identify stimulus presentation parameters within a CFS paradigm (i.e., unusual and interrupted stimulus presentations) that lead to reliable EC effects. Two additional studies (see last paragraph of Present Research) showed that one-to-many pairings impede EC effects under these conditions. Experiment 3.1 therefore used a “one-to-one” approach: Each CS was paired with a single unique US. In addition, we identified presentation time and focus on valence as two potential factors. We manipulated presentation duration of the pairings and whether participants classified US valence in every trial, because valence-related tasks have been shown to increase the focus on valence and, hence, the EC effect (Gast & Rothermund, 2011b).

Method.

Participants and design. One hundred twenty-two students participated (79 women, 43 men, average age 23.24 years). We manipulated presentation time (400 ms vs. 2,000 ms; see Material and procedure paragraph) and valence classification (classify vs. do not classify US) between participants and US valence (positive vs. negative) within participants. We measured two dependent variables: evaluative ratings and proportion of “positive” responses in the AMP for every CS; the AMP has been a sensitive indirect measure of evaluations in our lab (e.g., Förderer & Unkelbach, 2013, 2015, 2016).

A sensitivity analysis with G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) showed that this sample allows detecting a small overall effect (d = 0.27) with a power of .85 and α = .05 (two-tailed one-sample t test, see Results section).

Material and procedure. We created a computer program to conduct the experiment with OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). Upon arrival, experimenters seated participants at PC work stations and started the program; up to six people participated in experimental sessions. After providing informed consent and general demographic information, participants rated 60 animal photos regarding their pleasantness (1 = very unpleasant to 9 = very pleasant), thereby calibrating participants to the rating scale and allowing idiosyncratic stimulus selection (see Unkelbach, Stahl, & Förderer, 2012). For each participant, the computer program selected the four photos rated most positively and the four photos rated most negatively as positive and negative USs, respectively. As CSs, we used eight pictures of gray geometric shapes (Appendix A). The program assigned CS-US pairs randomly and presented each pair 10 times, resulting in 80 conditioning trials. During a trial, the CS appeared on one side of the screen for 400 ms (short presentation time) or 2,000 ms (long presentation time) while, on the other side of the screen, there was a rapidly changing sequence of the assigned US (75 ms short, 375 ms long) and a colored pixel mask (25 ms short, 125 ms long) repeated four times. Figure 3 illustrates this setup. The program counterbalanced arrangement of CS and US-flash on the left and right side of the screen to avoid systematic influences of interindividual dominance of one eye over the other. This counterbalancing is mainly important for subsequent experiments in which eyes compete for visual dominance. Participants all read the following instructions (translated from German): "We will now use pictures that you evaluated positively and negatively to investigate how they influence the processing of other stimuli. On one side of the screen, a fixation cross will be presented. Afterwards an animal picture and a colored pixel picture will be shown (very shortly) one after another on the same side, while on the other side of the screen a geometric shape will be visible. At the end of the experiment, we would like to examine how the animal photos interact with the perception of the geometric shapes. Therefore, we ask you to direct your gaze to the fixation cross and watch the pictures and shapes attentively, hereafter. If you do not have any question, you can start the task with the space key." "Very shortly" was only inserted for the short presentation times condition. Participants in the valence classification condition also indicated after every trial whether the US was positive or negative in a forced two-choice task. The program added the following sentence to their instructions: "You will be asked after every trial to classify the animal picture as positive or negative." Participants in the other condition pressed space to continue to the next trial.

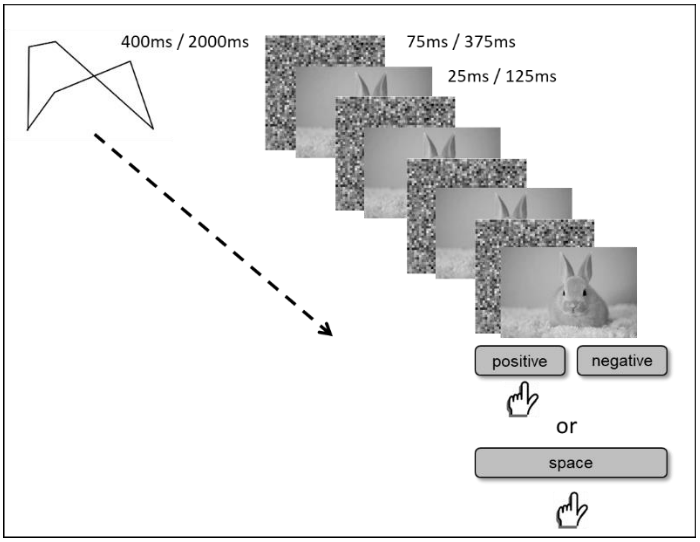


Figure 3. Stimulus presentation combining CFS with EC: a grey shape (conditioned stimuli, CS, presentation time depends on the condition) is presented on one side of the screen while a rapidly changing sequence of an animal photo (unconditioned stimuli, US) and colored pixel masks is presented on the other side. This picture sequence constitutes the flash – a core element of CFS. Depending on the condition, participants then either classify the valence of the US or press space to get to the next trial.

After conditioning, participants rated all CSs regarding their pleasantness on a Likert scale ranging from 1 = very unpleasant to 9 = very pleasant. After the rating, they completed the AMP (Payne et al., 2005), using the CSs as the potentially affective stimuli. A given AMP trial presented a CS for 75 ms, which was immediately replaced by a blank screen for 125 ms, followed by a Chinese character (Kanji) for 100 ms, followed by a black-and-white mask. Participants indicated with a keypress whether they perceived the Kanji as pleasant or unpleasant in a forced two-choice task. The rationale is that participants misattribute the affective reaction caused by the CS to the Kanji. Therefore, higher liking of a CS that was paired with a positive US should result in a higher proportion of "positive" responses to Kanjis shown after that CS, and vice versa for a CS that was paired with a negative US. The program provided 10 training trials and then 64 critical trials (each CS presented eight times). Key assignment of the response categories "positive" and "negative" was counterbalanced. Upon completion, participants were debriefed, thanked, and paid.

Results.

Evaluative ratings. Figure 4's upper half shows participants' mean ratings of CSs paired with positive USs (CSs+) and of CSs paired with negative USs (CSs-) as a function of presentation time and valence classification. We computed participants' mean rating differences of CSs+ and CSs- so that positive values indicate a standard EC effect and tested this difference against zero. We observed an overall significant EC effect, $M_{\text{Diff}} = 1.06$, $SD = 1.79$, $t(121) = 6.58$, $p < .001$, $d = 0.60$, 95% confidence interval (CI) [0.44, 0.74].[5]

Next, we analyzed the EC effects in a presentation time (short vs. long) x valence classification (classify vs. do not classify US) ANOVA. This analysis yielded two main effects. First, as Figure 4 suggests, the valence classification task significantly *reduced* this EC effect; participants in the valence classification conditions showed a smaller EC effect ($M = 0.46$, $SD =$

[5] The effect size d we report in all out experiments denominates Cohen's dz which is computed with the formula Cohen's $d_z = \frac{M_{diff}}{\sqrt{\frac{\Sigma(X_{diff} - M_{diff})^2}{N-1}}}$ implemented in the function "cohensD()" in the R package "lsr" (Lakens, 2013; Navarro, 2015). We bootstrapped 95% confidence intervals around dz with the R package "bootES" (Gerlanc & Kirby, 2015).

1.41) than participants who did not classify the USs, *M* = 1.65, *SD* = 1.93, *F*(1,118) = 14.82, *p* < .001, η^2 = 0.12, *95%CI* [0.03, 0.22]. Second, participants in the long presentation time conditions showed a stronger EC effect (*M* = 1.41, *SD* = 1.94) than participants in the short presentation time condition, *M* = 0.74, *SD* = 1.58, *F*(1, 118) = 4.97, *p* = .028 , η^2 = 0.03, *95%CI* [0, 0.13]. There was no interaction between valence classification and presentation time, *F*(1, 118) = 0.79, *p* = .377, η^2 = 0.01, *95%CI* [0, 0.06].[6]

AMP. Figure 4's lower half shows participants' mean proportion of "positive" responses in the AMP of CSs+ and CSs- as a function of presentation time and valence classification. We again computed mean differences in the proportion of "positive" responses towards CSs+ and CSs- so that positive values indicate a standard EC effect. We observed an overall EC effect, *M* = 0.05, *SD* = 0.19, *t*(121) = 2.91, *p* = .004, *d* = 0.26, *95%CI* [0.07, 0.43]. In a presentation time (short vs. long) x valence classification (classify vs. do not classify US) ANOVA, we detected no significant main or interaction effects, all *F* < 3.1, all η^2 < 0.03.[7] Proportion of "positive" responses in the AMP and explicit ratings were correlated, *r(*120*)* = 0.44, *p* < .001.

[6] Accordingly, testing the EC effect against zero within every condition separately, yielded an effect in the condition with long presentation times and valence classification (*t*(27) = 2.98, *p* = .006, *d* = 0.56, *95%CI* [0.26, 0.86]), with long presentation times and without valence classification (*t*(30) = 4.91, *p* < .001, *d* = 0.88, *95%CI* [0.52, 1.21), and short presentation times without valence classification (*t*(30) = 4.58, *p* < .001, *d* = 0.82, *95%CI* [0.48, 1.15]), but not in the condition with short presentation times and valence classification (*t*(31) = 0.23, *p* = .823, *d* = 0.04, *95%CI* [-0.32, 0.41]).

[7] Looking at each condition separately, there was an EC effect only in the short presentation times without valence classification condition, *t*(30) = 2.77, *p* = .010, *d* = 0.50, *95%CI* [0.14, 0.83], all other conditions: all *t*s < 1.7, all *d*s < 0.4.

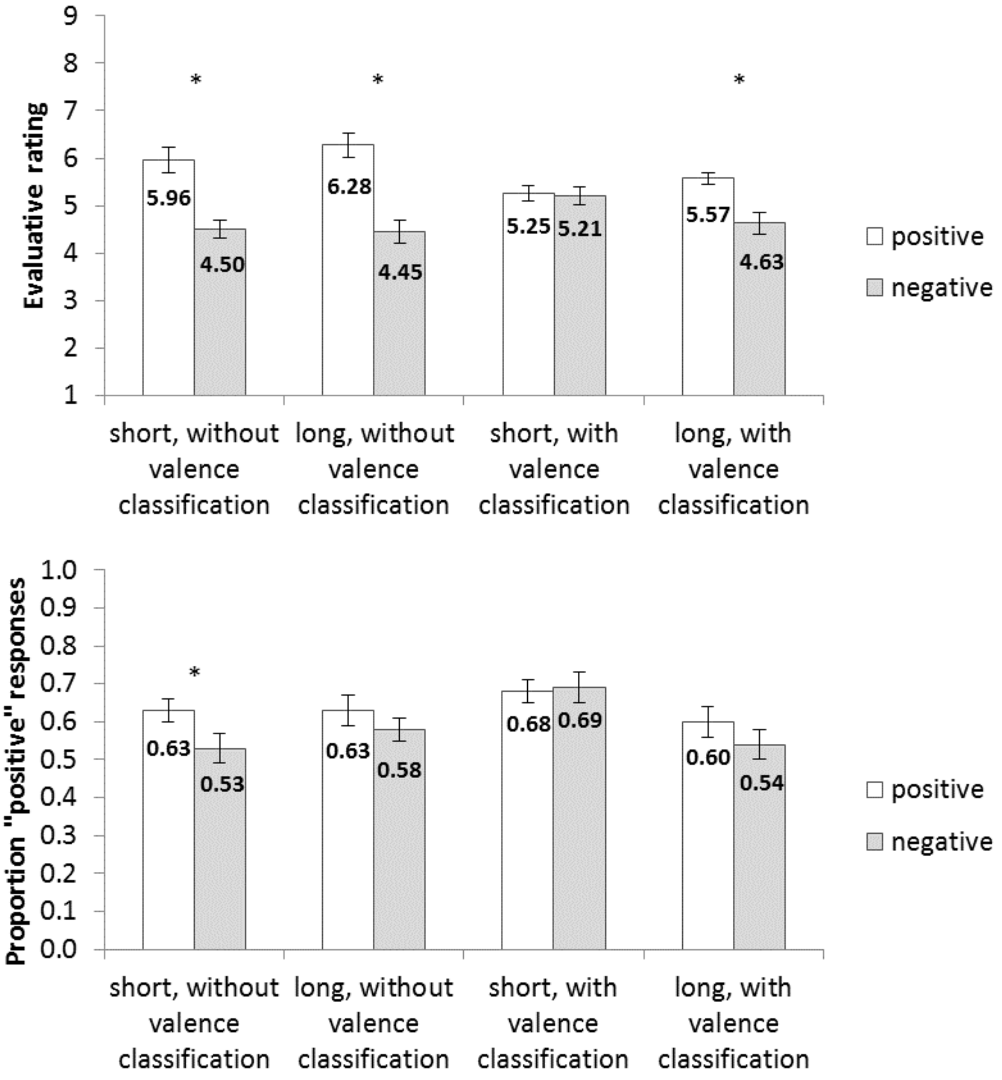


Figure 4. Evaluative ratings (upper figure, 1 = “very unpleasant”, 9 = “very pleasant”) and proportion of “positive” responses in the AMP of CSs that were paired with positive and negative USs as a function of presentation time of the CS-US pairs (short vs. long) and whether participants classified US valence in every trial. Error bars indicate the standard error of the mean; asterisks indicate significant EC effects.

Discussion. We observed reliable EC effects for direct (d = 0.60) and indirect (d = 0.24) measures under the unusual stimulus presentation situation of the present pseudo-CFS setup. For explicit ratings, long presentation times and no valence classification task promoted EC effects. The positive effect of long presentation times on EC is in line with the literature (Hofmann et al., 2010). We did not expect the negative effect of the valence classification task on EC. However, this finding is in line with the general notion that secondary tasks tie up cognitive resources and therefore reduce effects associated with primary tasks (Pleyers et al., 2009).

We decided to use the parameters of the short presentation times-no valence classification condition in Experiment 3.2. Although long presentation times were conducive of EC effects, we opted for short presentation times, because CFS works better with faster than slower flashes. This condition also yielded a strong EC effect in explicit ratings and was the only condition to produce a significant EC effect on the indirect evaluative measure.

Experiment 3.2

Experiment 3.2 investigated whether an EC effect would still be present when CSs were suppressed from awareness via CFS. Thus, participants viewed the CS-US pairings through stereoscopes, which presents the CS to one eye and the US to the other. In addition, we included an awareness measure at the end of the experiment to assess whether suppression was successful.

Method.

Participants and design. Sixty-eight students participated in Experiment 3.2 (46 female, 22 male, average age: 22.47 years). This sample allows detecting a small effect ($d = 0.37$) with 85% power and $\alpha = .05$ for a two-tailed one-sample t-test (see results section). We manipulated valence of paired US (positive vs. negative) within participants and assessed two dependent variables: evaluative rating and proportion of positive classifications in the AMP for every CS. As a manipulation check, we measured recognition of suppressed stimuli in an “offline” (i.e., post evaluative ratings) awareness test.

Material and procedure. Materials and procedure were highly similar to the short presentation times-no valence classification condition of Experiment 3.1, with two exceptions:

First, participants viewed stimulus presentation in the conditioning phase through stereoscopes to create dichoptic vision. We used ScreenScope mirror stereoscopes that comprise four mirror tiles arranged inside a viewer to split the visual field into two halves; the visual field of the left eye is restricted to the left side of the screen, and the visual field of the right eye is restricted to the right side of the screen (see Figure 5). This apparatus allows simultaneous presentation of different images to both eyes, which is necessary for CFS. The experimenters adjusted the stereoscopes for every participant individually before the experiment's learning phase. They were removed after conditioning, so that participants completed the evaluative ratings and the AMP under normal viewing conditions. As participants no longer viewed CS and US simultaneously, the instructions were changed to the following: "We will now use pictures that you evaluated positively and negatively to investigate your perception of disrupted pictures. You will see a fixation cross on the screen and afterwards an animal picture and a colored pixel picture will be shown very shortly one after another. We ask you to direct your gaze to the fixation cross and watch the pictures attentively, hereafter."

Second, to assess whether suppression was successful, participants completed an awareness test of 80 trials after the ratings and the AMP, using again the mirror stereoscopes. The awareness test was highly similar to the conditioning procedure but used novel grey geometric shapes instead of the CSs. Instead of the USs, the program used eight animal photos that participants had rated most neutral in the rating phase at the beginning of the experiment. Thus, parallel to conditioning, the presentation in the awareness test also involved pairs of geometric shapes on the suppressed eye and a flash of pixel masks and animal photos on the unsuppressed eye; the stimulus timing was the same as during conditioning.

We implemented this post-experiment test with stimuli other than the CSs to rule out correct recognition of CSs due to an affective characteristics acquired without awareness that informs recognition judgments instead of genuine recognition due to processing with awareness during conditioning (“inference account”, Stahl et al., 2009). Before viewing the pairings in the awareness test, participants again donned the stereoscopic viewers and removed them afterwards. Then they were asked to recognize out of four answer options the geometric shape that had appeared in the awareness test. They completed 16 recognition trials, eight with the target stimuli (i.e., where one of the four answer options was correct) and eight with foils (i.e., where none of the four answer options was correct). Participants could also indicate that none of the shapes had appeared in the awareness test.

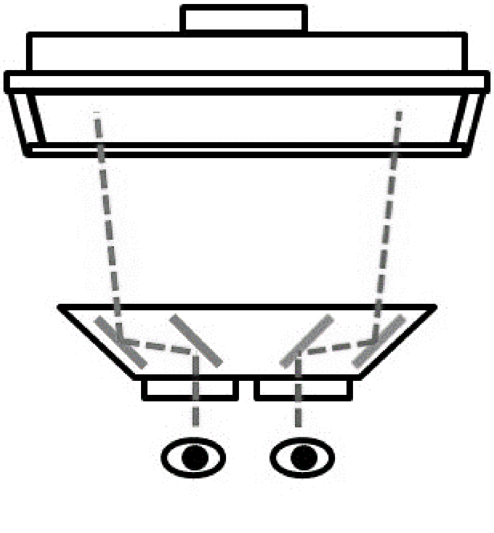


Figure 5. Schematic figure of a stereoscopic viewer through which participants watched the stimulus presentation on the screen. The arrangement of the four mirrors (grey bars) limits the visual field of the left eye to the left half of the screen and equally for the right eye.

Results.

Awareness manipulation check. In the majority of trials (*M* = 0.59, *SD* = 0.24) of the awareness test, participants indicated that they had seen none of the geometric shapes before. This was different from 0.5 which was the correct proportion of foil trials in the recognition task, *t*(67) = 3.15, *p* = .001, *d* = 0.38 *95%CI* [0.13, 0.63]. We then analyzed the remaining trials - those in which participant chose one of the four answer options. In these trials, participants, on average, performed on chance level (0.25), *M* = 0.24, *SD* = 0.22, *t*(62) = -0.51, *p* = .694, *d* = -0.06, *95%CI* [-0.35, 0.19]. Five participants indicated on all 80 trials that they had seen none of the geometric shapes before. Thus, they did not yield any data for the objective recognition test and were not regarded in this analysis, hence the reduced degrees of freedom. These data suggest that suppression was successful given the parameters of the experiment.

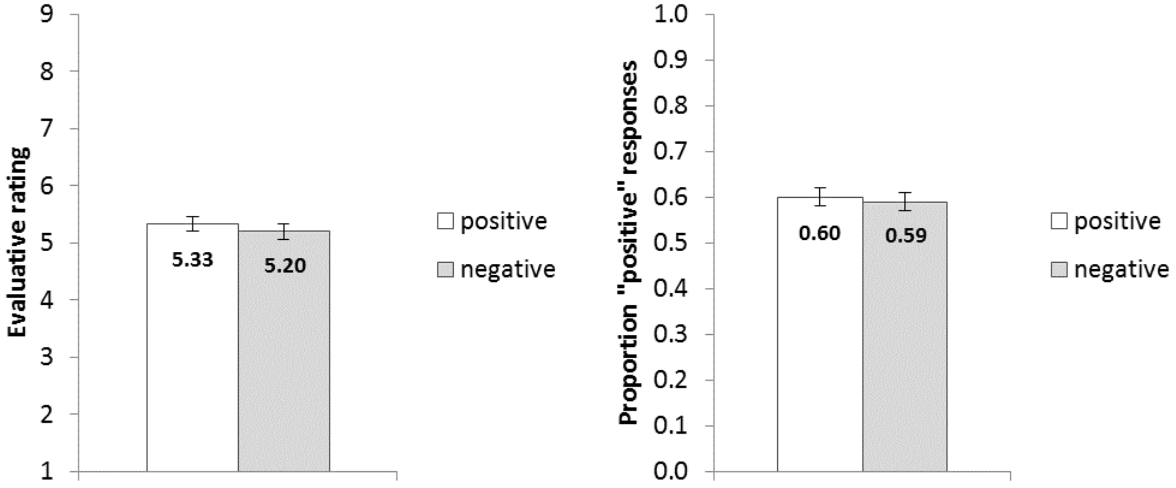


Figure 6. Evaluative ratings (left figure, 1 = “very unpleasant”, 9 = “very pleasant”) and proportion of “positive” responses in the AMP of CSs that were paired with positive and negative USs when CSs were suppressed from awareness via CFS. Error bars indicate the standard error of the mean; asterisks indicate significant EC effects.

Ratings and AMP. Figure 6 shows participants' mean evaluative ratings and mean proportion of "positive" responses in the AMP of CSs+ and CSs-. We conducted the same analyses as for Experiment 3.1. We observed no significant EC effect in the evaluative rating (*M* = 0.13, *SD* = 1.17, *t*(67) = 0.90, *p* = .373, *d* = 0.11, *95%CI* [-0.13, 0.36]) or the proportion of "positive" responses, *M* = 0.00, *SD* = 0.12, *t*(67) = 0.24, *p* = .809, *d* = 0.03, *95%CI* [-0.21, 0.28]. We additionally computed Bayes Factors (BF_{01}) which quantify the evidence for the null hypothesis relative to the alternative hypothesis (Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007). We observed BF_{01} = 5.11 for the evaluative rating and BF_{01} = 7.30 for the proportion of "positive" responses, which can be interpreted as substantial evidence against an EC effect in both dependent measures (Jeffreys, 1998).[8] The correlation between AMP responses and explicit ratings did not differ significantly from zero, *r*(66) = 0.21, *p* = .085.

Discussion. Experiment 3.2 showed that when CSs are suppressed from awareness via CFS, there is evidence against an EC effect. That is, the CSs did not show a standard EC effect and a Bayesian analysis indicated that the null-hypothesis is considerably more likely than the alternative hypothesis. From participants' performance on the awareness test, we concluded that suppression was successful. That is, for the majority of suppressed stimuli participants indicated that they had not seen them before and when participants indicated that they did, their recognition performance was at chance level.

Crucially, numerous studies show that processing of CFS-suppressed stimuli is not simply entirely abolished. Previous research using CFS showed effects of the suppressed stimuli

[8] We used the function "ttestBF()" and from the R package "BayesFactor" (Morey & Rouder, 2015) to compute Bayes factors in all experiments. We used the default medium prior that corresponds to an rscale parameter of $\sqrt{2} \div 2$. This means that 50% of the true prior standardized effect sizes lie between −0.7071 and 0.7071. This prior has been identified as a "reasonable" default prior for psychological research because it covers common effect sizes and is computationally convenient (also for ANOVA designs; Rouder, Morey, Speckman, & Province, 2012).

on visual aftereffects (Tsuchiya & Koch, 2005), perceptual learning (e.g. Seitz et al., 2009), fear conditioning (Raio et al., 2012), or priming (e.g. Faivre et al., 2012). This demonstrates that stimuli suppressed via CFS are processed to an extent that should allow for EC. The findings so far support the following conclusion: Viewing CSs with awareness (Experiment 3.1) leads to an EC effect, even under nonoptimal presentation conditions, but viewing CSs without awareness (Experiment 3.2) does not. For sure, there is another factor besides awareness that systematically varies between Experiment 3.1 and 3.2, namely the use of stereoscopes. It might have distracted participants' attention away from the CS-US pairs. Furthermore, while in Experiment 3.2, CS and US flash were presented at the same retinal location due to CFS, CS and US flash were viewed at an angle in Experiment 3.1. Lastly, we did not obtain a recognition measure for suppressed stimuli in Experiment 3.1. Therefore, the two experiments could not be compared with regard to recognition performance. Thus, Experiment 3.3 aimed to replicate Experiment 3.2 and in addition, show an EC effect with stereoscopic vision without suppression.

Furthermore, in Experiment 3.2, we obtained an awareness estimate for suppressed recognition stimuli other than the CSs to preclude the possibility that acquired valence is used as a cue in the recognition task in the absence of genuine recognition. This awareness test, however, only allowed us to draw conclusions about the success of suppression on a participant level. In order to analyze CS evaluations conditional on whether they were processed with or without awareness (i.e. stimulus-level; Pleyers et al., 2007), we opted for a measure of recognition of the CSs proper in Experiment 3.3.

Experiment 3.3

Experiment 3.3 manipulated suppression of CSs within participants. Both the CSs perceived with and without awareness were viewed through stereoscopic viewers and we obtained awareness data for all CSs. This within-participants approach created experimental conditions that only differed with regard to the state of awareness of the CSs, thus enabling strong conclusions about the role of awareness in EC.

Method.

Participants and design. Seventy-six students participated in Experiment 3.3 (44 female, 32 male, average age: 21.92 years). This sample allows detecting a small effect (d = 0.35) in every condition and a small difference between conditions (*d* = 0.35) with 85% power and *α* = .05 (two-tailed one-sample and paired t-test, see results section). We manipulated within participants, whether a CS was suppressed from awareness via CFS or not, and whether a CS was paired with a positive or negative US. Suppression was manipulated in a block-wise manner: In the first block, all CSs were suppressed from awareness; in the second block they were not. Block order was not counterbalanced, as a reversed order potentially prevents suppression. We assessed two dependent variables: evaluative rating and proportion of positive responses in the AMP for every CS. As a manipulation check, we measured recognition of suppressed and unsuppressed CSs.

Material and procedure. We used 16 geometric shapes as CSs and eight positive and negative animal photos as USs. To reduce the experiment’s duration, we used ratings of the animal photos from Experiment 3.1 to determine positive USs and negative USs. We selected the photos that were rated most positively (*M* = 8.04, *SD* = 1.36) and negatively (*M* = 2.95, *SD* = 2.01) as USs. As the prerating phase was dropped, the instructions did not mention positivity or negativity of the pictures anymore but were otherwise largely identical. The computer program randomly assigned CS-US pairs to each other and determined which CSs were to be suppressed from awareness or not. The suppressed CSs were presented in grey color to make suppression more feasible while the unsuppressed CSs were presented in black color.

Upon arrival, experimenters welcomed and seated participants at the lab computers. If participants gave informed consent, they donned the stereoscopic viewers and started with the first block of conditioning which consisted of 80 CS-US pairings. Different from Experiment 3.1 and 2, we increased overall trial length to 2000 ms. A flash of US (100 ms) and pixel masks (100 ms) was presented to one eye and was repeated ten times. The CS appeared on the other eye with a delay of 200 ms (after one flash) and was presented for 1800 ms.

After the first block of conditioning, participants took off the stereoscopes and completed a recognition measure. They indicated in a forced multiple-choice task which of four geometric shapes had been presented during conditioning. In all eight trials, one of the four answer options had been presented as CS in the previous conditioning block. Thus, unlike in Experiment 3.2, there were no foil trials where no option was correct and consequently there was also no option to indicate that none of the shapes had been presented. The distractors were also grey geometric shapes. The program randomly determined the position of the target for every trial. After the recognition task, participants completed the evaluative ratings and the AMP, using the same parameters as in Experiments 3.1 and 3.2.

For the second, unsuppressed block of conditioning, participants also wore stereoscopic viewers. Two minor modifications differentiated the second from the first block: CSs were presented in black instead of grey color (i.e., with higher contrast) and the program presented a flash of US photos and blank screens to the unsuppressed eye instead of pixel masks. The repeated time windows of 100 ms of blank screen were supposed make the CS visible. After the second block of conditioning, participants again completed the recognition task with the eight unsuppressed CSs, the evaluative ratings, and the AMP. Upon completion, experimenters debriefed and rewarded participants.

We opted for the changes in stimulus timing in comparison to Experiment 3.1 and 3.2 to make the emergence of an EC effect in the control condition (unsuppressed block) feasible: First of all, we increased trial length (and thereby, increased the number of repetitions of the flash). As a compromise between long presentation times that are desirable in the unsuppressed condition, and the effectiveness of suppression that is desirable in the suppressed condition, we delayed CS onset by 200 ms. Delayed (or gradual) onset of the suppressed stimulus is considered to be conducive of suppression (e.g., Sklar et al., 2012). Furthermore, we adapted the presentation times within the flash so that US and pixel masks /blank screen were both repeatedly shown for 100 ms (as opposed to 75 ms for US and 25 ms for pixel masks in Experiment 3.1 and 3.2). We deemed this necessary to give the visual system enough time to register the CS stimulus on the other side of the screen in the unsuppressed condition. However, Experiment 3.4 will replicate Experiment 3.3 with the timing parameters of Experiment 3.2.

Results.

Recognition. Participants recognized suppressed CSs less often (*M* = 0.50, *SD* = 0.29) than CSs that were not suppressed from awareness (*M* = 0.81, *SD* = 0.23), *t*(75) = 8.38, *p* < .001, *d* = 0.96, *95%CI* [0.65, 1.26]. However, both suppressed and unsuppressed CSs were recognized above the 0.25 chance threshold; suppressed: *t*(75) = 7.63, *p* < .001, *d* = 0.88, *95%CI* [0.64, 1.11], unsuppressed: *t*(75) = 21.09, *p* < .001, *d* = 2.42, *95%CI* [1.82, 3.08].

Ratings. Figure 7’s left panel shows participants’ mean ratings of CSs+ and CSs- as a function of whether they were suppressed from awareness or not. We computed a paired sample *t*-test with EC effects as dependent variable and observed a larger EC effect for unsuppressed

CSs (M = 0.83, SD = 2.11) than for suppressed CSs, M = 0.11, SD = 1.29, t(75) = 2.63, p = .010 , d = 0.30, *95%CI* [0.08, 0.50]. One-sample *t*-tests within each condition showed an EC effect for the unsuppressed CSs (t(75) = 3.42, p = .001, d = 0.39, *95%CI* [0.13, 0.63] and no EC effect for the suppressed CSs, t(75) = 0.73, p = .467 d = 0.08, *95%CI* [-0.15, 0.31]. In a Bayesian *t*-test, we observed BF_{01} = 6.12 for the suppressed CSs, which can be interpreted as substantial evidence against an EC effect (Jeffreys, 1998).

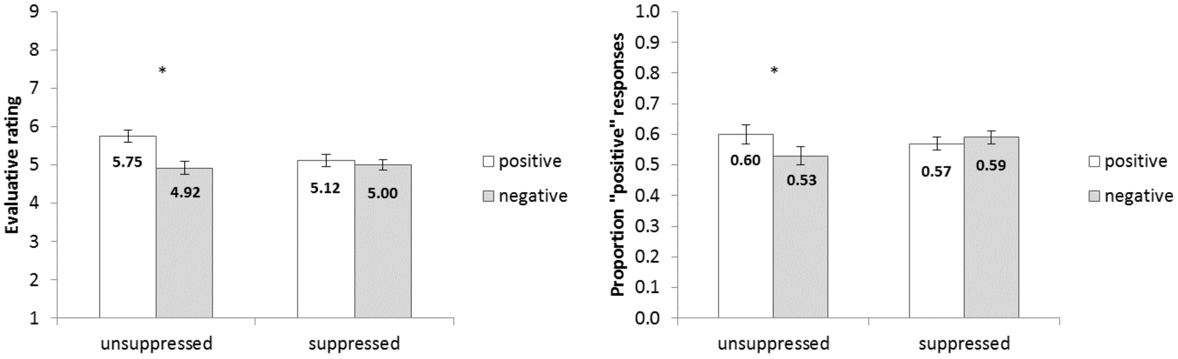


Figure 7. Evaluative ratings (left figure, 1 = “very unpleasant”, 9 = “very pleasant”) and proportion of “positive” responses in the AMP of CSs that were paired with positive and negative USs as a function of whether CSs were suppressed from awareness via CFS or not. Error bars indicate the standard error of the mean; asterisks indicate significant EC effects.

AMP. Figure 7’s right panel shows participants’ mean proportion of “positive” responses of CSs+ and CSs- as a function of whether they were suppressed from awareness or not. We computed a paired sample *t*-test with EC effect in the AMP as dependent variable and again observed a larger EC effect for unsuppressed CSs (M = 0.07, SD = 0.24) than for suppressed CSs, M = -0.02, SD = 0.13, t(75) = 2.78, p = .007, d = 0.32, *95%CI* [0.08, 0.53]. We also ran one-sample *t*-tests within each condition, testing the EC effect in the AMP against zero. We observed an effect for the unsuppressed CSs (t(75) = 2.69, p = .009 , d = 0.31, *95%CI* [0.05, 0.51]) and no EC effect for the suppressed CSs, *t*(75) = -1.20, *p* = .234, *d* = -0.14, *95%CI* [-0.35, 0.09]. We observed BF_{01} = 3.98 in a Bayesian *t*-test, which can be interpreted as substantial evidence against an EC effect for suppressed CSs (Jeffreys, 1998). In the unsuppressed condition, AMP and rating responses correlated positively (*r*(74) = 0.57, *p* < .001), while in the suppressed condition they correlated negatively, *r*(74) = -0.27, *p* = .018.

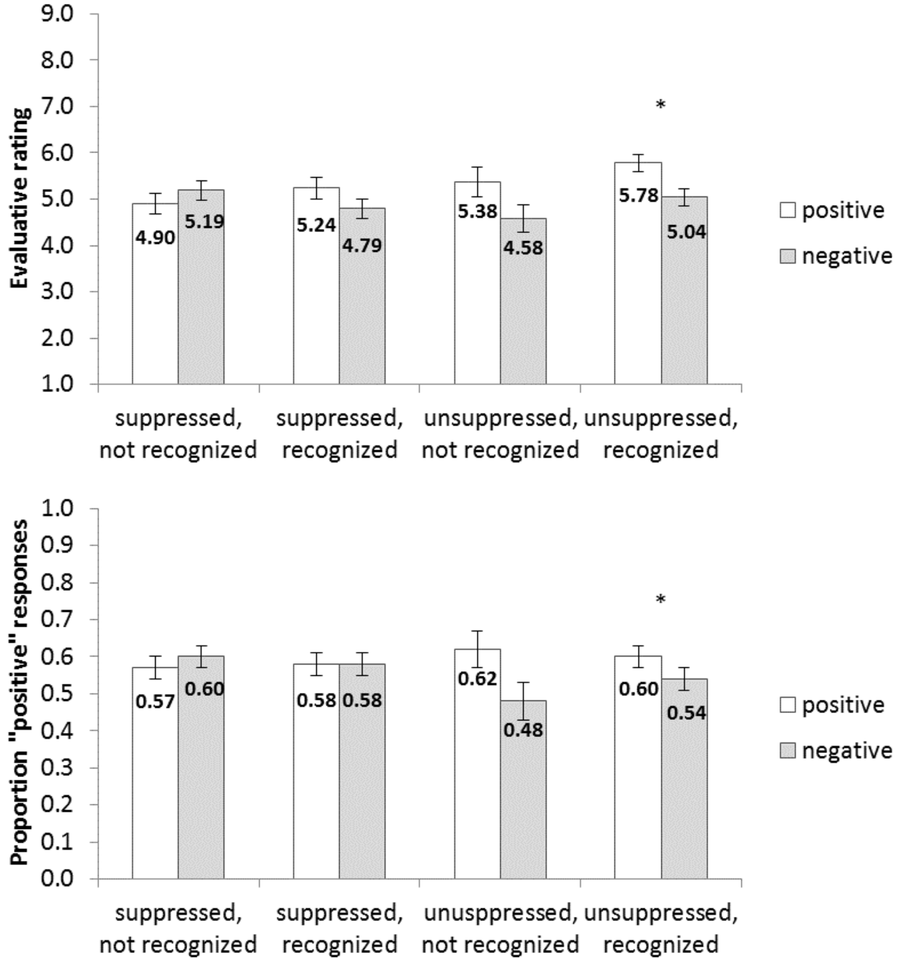


Figure 8. Evaluative ratings (upper figure, 1 = "very unpleasant", 9 = "very pleasant") and proportion of "positive" responses in the AMP of CSs that were paired with positive and negative USs as a function of whether CSs were suppressed from awareness via CFS or not and whether they were recognized correctly in a post-conditioning recognition test. Error bars indicate the standard error of the mean; asterisks indicate significant EC effects.

Analyses conditional on recognition. Different from Experiment 3.2, the recognition of suppressed CSs was not on chance level. Thus, we analyzed the data conditional on (un)successful recognition of the CSs (see Pleyers et al., 2007), supplementing the present experimental with a correlational approach (see Introduction). We thereby computed the mean evaluations based on the recognized and not recognized CSs in the suppressed and unsuppressed conditions. Figure 8 shows participants' respective mean evaluations in rating (upper part) and AMP (lower part) of CSs+ and CSs-. We computed EC effects for CSs in all of those four conditions for both dependent variables separately and tested them against zero. First, we analyzed CSs that were (not) recognized "in line" with what we expected from the experimental manipulation. We observed an EC effect for CSs that were not suppressed and were therefore expectedly recognized correctly in the rating (*M* = 0.76, *SD* = 2.32), *t*(71) = 2.78, *p* = .007, *d* = 0.33 *95%CI* [0.08, 0.56]) and the AMP, *M* = 0.07, *SD* = 0.25, *t*(71) = 2.42, *p* = .018, *d* = 0.29, *95%CI* [0.02, 0.50]. Furthermore, as expected, CSs that were suppressed and therefore not recognized did not show an EC effect in the rating (*M* = -0.19, *SD* = 1.67), *t*(57) = -0.87, *p* = .387, *d* = -0.12, *95%CI* [-0.38, 0.15], BF_{01} = 4.85) or in the AMP, *M* = -0.01, *SD* = 0.20, *t*(57) = -0.52, *p* = .605, *d* = -0.07, *95%CI* [-0.32, 0.20], BF_{01} = 6.12.

Then, we analyzed CSs that were not "in line" with the manipulation. CSs that were not suppressed from awareness but were nevertheless not recognized did not yield an EC effect in the rating (*M* = 0.13, *SD* = 2.28), *t*(23) = 0.27, *p* = .791, *d* = 0.06, *95%CI* [-0.37, 0.47], BF_{01} = 4.51) nor the AMP, *M* = 0.10, *SD* = 0.31, *t*(23) = 1.51, *p* = .144, *d* = 0.31, *95%CI* [-0.10, 0.61], BF_{01} = 1.72. And CSs that were suppressed but nevertheless recognized – arguably the most interesting analysis – did not show an EC effect in the rating (*M* = 0.29, *SD* = 2.40), *t*(55) = 0.92, *p* = .363, *d* = 0.12, *95%CI* [-0.16, 0.39], BF_{01} = 4.59) nor in the AMP, *M* = -0.02, *SD* = 0.18, *t*(55) = -0.99, *p* = .325, *d* = -0.13, *95%CI* [-0.39, 0.13], BF_{01} = 4.29.[9]

Discussion. Using a within-participants approach, Experiment 3.3 showed that EC effects do not emerge when CSs are suppressed from awareness. Suppressed CSs were recognized above chance, though. This could be due to insufficient suppression via CFS or processing to some extent of the recognized stimuli. As the US flash and CS had simultaneous stimulus offset, it is possible that participants saw afterimages of the CSs, for example. Either way, despite recognition above chance, suppressed CSs did not show an EC effect. A stimulus-level analysis showed that even those CSs that were recognized correctly despite suppression did not show EC effects (see also Stahl et al., 2016). These findings suggest that even relatively weak suppression abolishes EC. Note, however, that findings from these stimulus-level analyses should be interpreted with caution, because the number of observations in every condition differed strongly resulting in differential power to detect an effect.

Experiment 3.3 employed a longer overall trial length and a different timing within the flash than Experiment 3.1 and 3.2, in order to make EC effects with unsuppressed CSs more feasible. One could argue, though, that these timing changes were somewhat arbitrary. Furthermore, the different color of presentation of suppressed and unsuppressed CSs (grey vs. black) could constitute a potential confound between the experimental blocks. Even though it is difficult to imagine why a difference in contrast should lead to the absence or presence of an EC effect, we opted to eliminate this confound in the next experiment.

[9] However, suppressed but recognized CSs did not differ significantly from unsuppressed recognized CSs regarding the size of the EC effect in the rating (t(52) = -0.67, *p* = .509, *d* = -0.09, *95%CI* [-0.37, 0.19], BF_{01} = 5.41) nor the AMP, t(52) = -1.71, *p* = .093, *d* = -0.24, *95%CI* [-0.50, 0.06], BF_{01} = 1.72.

Experiment 3.4

Experiment 3.4 aimed to replicate the findings obtained in Experiment 3.3, testing their robustness across different timing parameters. We again manipulated suppression of CSs within participants but used the timing parameters of Experiment 3.2 to enhance comparability. Furthermore, suppressed and unsuppressed CSs were both presented in black color as opposed to grey and black color in Experiment 3.3, ruling out a potential confound.

Method.

Participants and design. Ninety-one students participated in Experiment 3.4 (54 female, 37 male, average age: 24.79 years). This sample allows detecting a small effect (d = 0.32) in every condition and a small difference between conditions (*d* = 0.32) with 85% power and *α* = .05 (two-tailed one-sample and paired t-test, see results section). Design and measure were identical to Experiment 3.3.

Material and procedure. The material and procedure of Experiment 3.4 were very similar to the one used in Experiment 3.3 with two exceptions. First, the stimulus timing was the same we used in Experiment 3.1 in the short presentation time condition and in Experiment 3.2. The CS was presented for 400 ms while the flash on the other eye alternated four times between the assigned US (75 ms) and a mask (25 ms). Second, both suppressed and unsuppressed CSs were presented in black color.

Results.

Recognition. Suppressed CSs were recognized correctly less often (*M* = 0.53, *SD* = 0.29) than unsuppressed CSs (*M* = 0.76, *SD* = 0.23), *t*(90) = 7.25, *p* < .001, *d* = 0.76, *95%CI* [0.53, 0.98]. CSs that were not suppressed from awareness were recognized above chance (0.25

threshold, $t(90) = 20.90$, $p < .001$, $d = 2.19$, *95%CI* [1.67, 2.77]) and so were CSs that were suppressed from awareness, $t(90) = 9.35$, $p < .001$, $d = 0.98$, *95%CI* [0.76, 1.21].

Ratings. Figure 9's left panel shows participants' mean ratings of CSs+ and CSs- as a function of suppression. Replicating Experiment 3.3, we observed a larger EC effect for unsuppressed CSs ($M = 1.20$, $SD = 2.01$) than for suppressed CSs, $M = 0.22$, $SD = 1.46$, $t(90) = 3.86$, $p < .001$, $d = 0.40$, *95%CI* [0.20, 0.60]. Within each condition, we again observed EC effects for the unsuppressed CSs ($t(90) = 5.20$, $p < .001$, $d = 0.55$, *95%CI* [0.36, 0.72] but not for the suppressed CSs, $t(90) = 1.42$, $p = .159$, $d = 0.15$, *95%CI* [-0.06, 0.33]. We observed $BF_{01} = 3.28$ in a Bayesian *t*-test for the suppressed CSs, which can be interpreted as substantial evidence against an EC effect (Jeffreys, 1998). We then pooled responses for suppressed CS from Experiment 3.3 and 3.4 and ran the same analyses to obtain a larger power to detect EC. Unchangedly, the EC effect was not significant, $M = 0.17$, $SD = 1.38$, $t(166) = 1.57$, $p = .\ 119$, $d = 0.12$, *95%CI* [-0.03, 0.26], $BF_{01} = 3.50$.

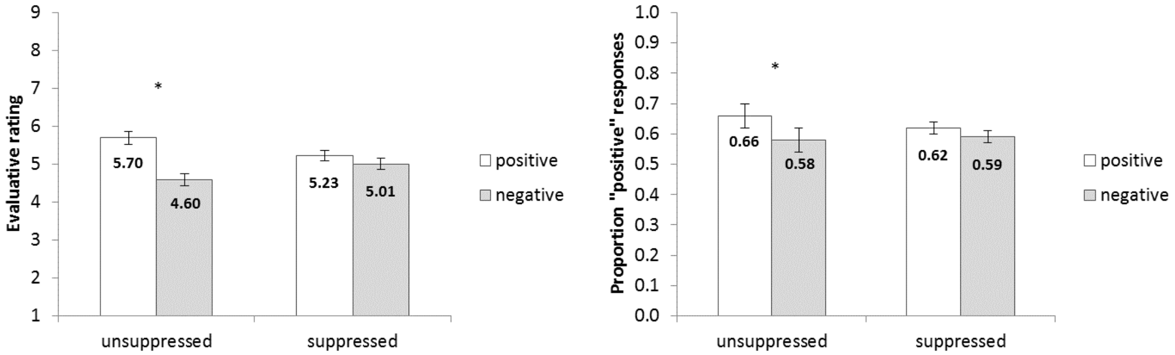


Figure 9. Evaluative ratings (left figure, 1 = "very unpleasant", 9 = "very pleasant") and proportion of "positive" responses in the AMP of CSs that were paired with positive and negative USs as a function of whether CSs were suppressed from awareness via CFS or not. Error bars indicate the standard error of the mean; asterisks indicate significant EC effects.

AMP. Figure 9's right panel shows participants' mean proportion of "positive" responses of CSs+ and CSs- as a function of suppression. A paired sample *t*-test with EC effect in the AMP as dependent variable showed a larger EC effect for unsuppressed CSs (*M* = 0.08, *SD* = 0.23) than for suppressed CSs, *M* = 0.03, *SD* = 0.15, *t*(90) = 2.19, *p* = .031, *d* = 0.23, *95%CI* [0.04, 0.39]. Subsequent one-sample *t*-tests compared the EC effect in the AMP against zero within each condition. We observed an EC effect for the unsuppressed CSs (*t*(90) = 3.52, *p* < .001, *d* = 0.37, *95%CI* [0.19, 0.51]) and no EC effect for the suppressed CSs, *t*(90) = 1.85, *p* = .068, *d* = 0.19, *95%CI* [-0.01, 0.38]. In a Bayesian *t*-test, we observed BF_{01} = 1.69, which can hardly differentiate between the null and the alternative hypothesis (Jeffreys, 1998). When we pooled AMP responses for suppressed CS from Experiment 3.3 and 3.4 to achieve greater power, the results were clearer: we observed substantial evidence against an EC effect, *M* = 0.01, *SD* = 0.14, *t*(166) = 0.68, *p* = . 500, *d* = 0.05, *95%CI* [-0.10, 0.20], BF_{01} = 9.26.

For unsuppressed CSs, the proportion of "positive" responses in the AMP and rating responses were correlated, *r*(89) = 0.52, *p* < .001. For suppressed CSs there was no such correlation, *r*(89) = 0.02, *p* = .872.

Analyses conditional on recognition. As in Experiment 3.3, we additionally analyzed rating and AMP data conditional on recognition of the CS. Figure 10 shows participants' mean evaluations in rating (upper half) and AMP (lower half) of CSs+ and CSs-. We used EC effects for CSs in all four conditions as dependent variables and tested them against zero. For CSs that were not suppressed and were therefore recognized correctly, we observed an EC effect in the rating (*M* = 1.35, *SD* = 2.17), *t*(86) = 5.79, *p* < .001, *d* = 0.62 *95%CI* [0.41, 0.81]) and in the AMP, *M* = 0.09, *SD* = 0.22, *t*(86) = 3.73, *p* < .001, *d* = 0.40, *95%CI* [0.24, 0.53]. Concerning suppressed CSs that were, in line with the manipulation, not recognized, we did not observe an

EC effect in the rating (*M* = 0.24, *SD* = 2.11), *t*(65) = 0.93, *p* = .355, *d* = 0.12, *95%CI* [-0.13, 0.35], BF_{01} = 4.89) or in the AMP, *M* = -0.002, *SD* = 0.17, *t*(65) = -0.11, *p* = .909, *d* = -0.01, *95%CI* [-0.27, 0.23], BF_{01} = 7.36.

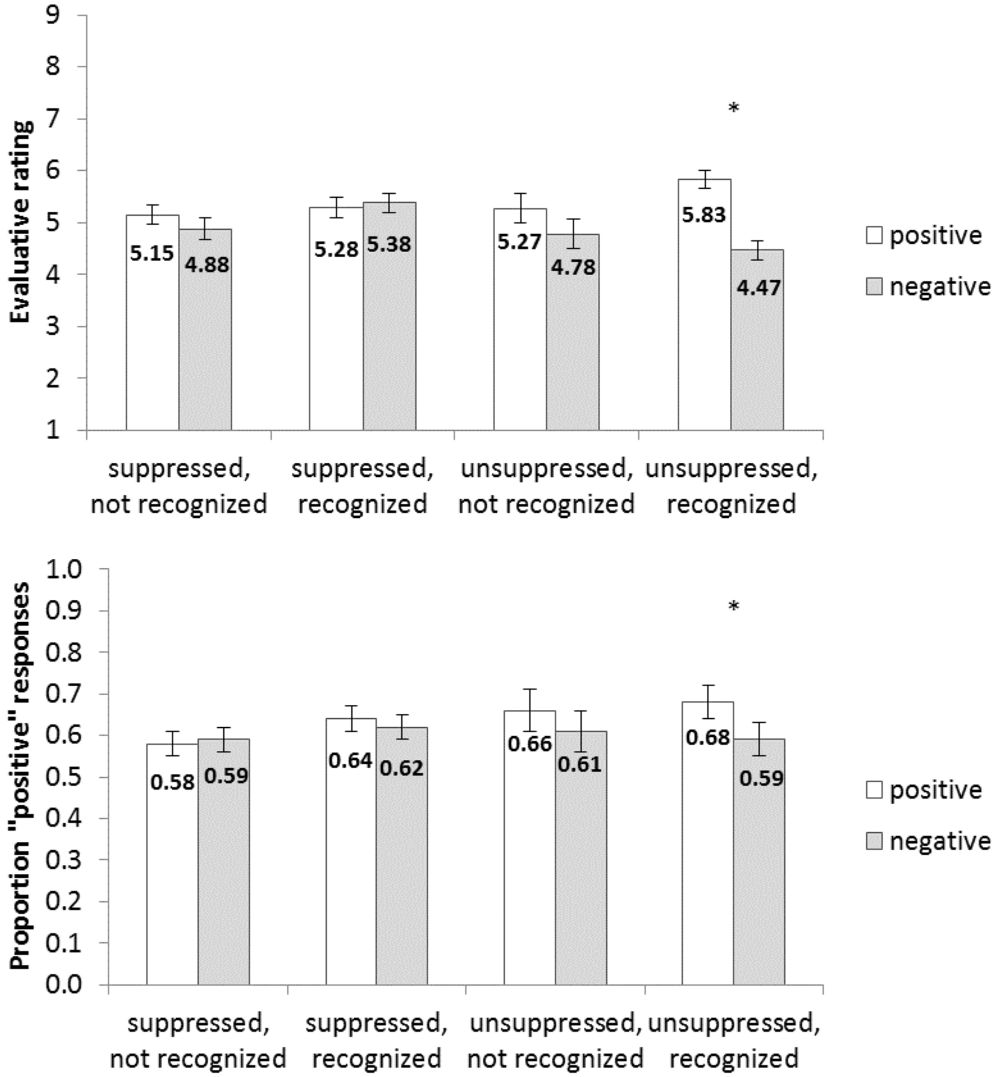


Figure 10. Evaluative ratings (upper figure, 1 = “very unpleasant”, 9 = “very pleasant”) and proportion of “positive” responses in the AMP of CSs that were paired with positive and negative USs as a function of whether CSs were suppressed from awareness via CFS or not and whether they were recognized correctly in a post-conditioning recognition test. Error bars indicate the standard error of the mean; asterisks indicate significant EC effects.

Analyzing CSs that were not "in line" with the manipulation, we first observed that unsuppressed CSs that were nevertheless not recognized did not yield an EC effect in the rating (*M* = 0.44, *SD* = 2.54), *t*(33) = 1.01, *p* = .319, *d* = 0.17, *95%CI* [-0.18, 0.49], BF_{01} = 3.40) nor the AMP, *M* = 0.04, *SD* = 0.32, *t*(33) = 0.81, *p* = .424, *d* = 0.14, *95%CI* [-0.21, 0.46], BF_{01} = 4.02. Finally, suppressed CSs that were nevertheless recognized did not show an EC effect in the rating (*M* = 0.06, *SD* = 2.05), *t*(71) = 0.26, *p* = .796, *d* = 0.03, *95%CI* [-0.21, 0.26], BF_{01} = 7.47) nor in the AMP, *M* = 0.03, *SD* = 0.18, *t*(71) = 1.23, *p* = .224, *d* = 0.15, *95%CI* [-0.09, 0.37], BF_{01} = 3.76.[10]

In an effort to achieve greater power, we jointly analyzed data of suppressed but recognized CSs from Experiment 3.3 and 3.4. However, we observed no EC effect in the rating (*M* = 0.16, *SD* = 2.20, *t*(127) = 0.84, *p* = .401, *d* = 0.07, *95%CI* [-0.10, 0.25], BF_{01} = 7.20) nor the AMP, *M* = 0.004, *SD* = 0.18, *t*(127) = 0.25, *p* = .802, *d* = 0.02, *95%CI* [-0.16, 0.20], BF_{01} = 9.87.

Discussion. Experiment 3.4 replicated the findings from Experiment 3.3 using the stimulus timing of Experiments 3.1 and 3.2 and keeping the CSs color constant across experimental blocks. It shows that evaluative learning does not emerge when CSs are suppressed using the CFS paradigm. In line with Stahl and colleagues (2016), this conclusion holds both for CSs that were recognized and for CSs that were not recognized.

General Discussion

The question whether evaluative learning is possible without awareness is at the heart of a lively debate about single- vs. dual-process theories in psychology. We aimed to show evaluative

[10] In contrast to Experiment 3.3, suppressed but recognized CSs differed significantly from unsuppressed recognized CSs regarding the size of the EC effect in the rating (t(68) = -4.24, *p* < .001, *d* = -0.51, *95%CI* [-0.74, -0.27]) and in the AMP, t(68) = -2.05, *p* = .044, *d* = -0.25, *95%CI* [-0.44, -0.02].

learning without awareness in EC. To create optimal conditions for EC effects without awareness, we presented CSs to participants using Continuous Flash Suppression. This technique keeps stimulus durations, locations, and fixations constant between suppressed and unsuppressed conditions and trials. In addition, there is substantial evidence for the influence of suppressed stimuli on cognition. Despite these, in our opinion, optimal conditions, we observed no evidence for EC effects for suppressed CSs. Therefore, our results do not support the existence of evaluative learning without awareness.

Experiments 3.1 and 3.2 investigated EC in a CFS paradigm with and without suppression. Black-and-white CSs were presented on one side, flashes of pixel masks and US on the other side of the screen. In Experiment 3.1, participants simply viewed this stimulus presentation perceiving both CS and US with awareness. In Experiment 3.2, participants viewed the same stimulus presentation dichoptically; one eye saw the CS, the other eye saw the US flash. This dichoptic viewing made the CS and the US flash compete for visual awareness. Because the US flash carried more visual information than the stationary black-and-white CS, the CS was suppressed from awareness. Recognition performance of the suppressed CSs was at chance level and there was evidence against an EC effect in both, a direct and an indirect measure, whereas the unsuppressed CSs in Experiment 3.1 showed a strong EC effect.

The interim conclusion from Experiment 3.1 and 3.2 was that suppressing CSs from awareness abolishes the EC effect. Alternatively, however, the use of stereoscopes accompanied by participants' distraction from the stimulus presentation might account for the disappearance of the EC effect. Experiment 3.3 and 3.4 thus employed a within-participant manipulation of suppression via CFS, keeping the dichoptic viewing conditions constant. Results corroborated the conclusion that CSs that are suppressed from awareness via CFS do not produce an EC effect. Furthermore, suppressed CSs were recognized correctly less often than unsuppressed CSs. Even those suppressed CSs that were nevertheless correctly recognized in a forced-choice task did not show an EC effect. However, omitting the suppression (i.e., the flash), we observed reliable EC effects, even when participants saw the CSs and USs via stereoscopes.

The present results are therefore best explained by a single-process account of EC. Our findings are in line with two recent findings using subliminal and parafoveal presentation of CSs that also pointed out awareness as a necessary condition for EC (Dedonder et al., 2014; Stahl et al., 2016). Going beyond this research, the employed CFS methodology does not rely on stimulus exposure duration or spatial location to manipulate awareness. Both are considered problematic because they afford alternative explanations for a lack of unaware EC (Jones et al., 2009). Second, CFS allows to meet criteria such as simultaneous presentation of CS and US, which have been identified as facilitating EC without awareness (Jones et al., 2009, see also Hütter & Sweldens, 2013a). Third, CFS can be classified as a method creating conditions of “preconscious” perception (s. Figure 2; Also Dehaene et al., 2006; but see Bahrami, Carmel, Walsh, Rees, & Lavie, 2008; Bahrami, Lavie, & Rees, 2007, for findings showing that the role of attention in CFS is more complex than simply being absent). Preconscious stimuli have a sufficient bottom-up strength but do not pass the awareness threshold and can therefore not be reported. They have been shown to elicit more intense cortical activation and to have more behavioral potential than subliminal stimuli (Dehaene et al., 2006). Hence, “preconscious” presentation of stimuli seemed like a highly promising avenue to show evaluative learning without awareness. Nevertheless, we failed to find evidence for evaluative learning without awareness across three experiments.

Limitations. Any potential conclusions from these results must assume that stimuli suppressed via CFS are processed to an extent that allows for evaluative conditioning. This crucial requirement is attested by numerous studies using CFS to study influences without awareness on behavior such as perceptual learning (e.g. Seitz et al., 2009) or priming (e.g. Faivre et al., 2012). Additional evidence comes from neuropsychological studies repeatedly showing that CFS-suppressed stimuli are processed in high-level brain areas (e.g., Jiang & He, 2006; Vizueta, Patrick, Jiang, Thomas, & He, 2012). Even if only very low-level features of suppressed stimuli are processed, as some studies suggest (Gray, Adams, Hedger, Newton, & Garner, 2013; Stein & Sterzer, 2011), there is no reason to preclude the hypothetical possibility of EC under suppression. The effective pairing would be a pairing of low-level CS features and a US which could result in a more positive or negative evaluation of the CS. Finally, our own recognition data from Experiments 3.3 and 3.4 suggest that participants encoded the CSs at least to some extent.

One might also argue that suppressing USs instead of CSs from awareness would have been preferable. Affective information might be more relevant for the organism and might therefore be more readily processed without awareness. As outlined in the introduction, however, processing of affective stimulus properties seems to be abolished by CFS (Yang et al., 2010). If affect is processed, on the other hand, suppression is abolished (e.g., Gayet et al., 2016; Stein & Sterzer, 2012; Yang, Zald, & Blake, 2007). This would render it difficult to investigate EC effects when USs are suppressed.

Another limitation of CFS is the difference in presentation modalities (black and white and stationary vs. colored and flashed); this might also limit the degree to which CS and US are processed in an associative and thus assimilative manner (Fiedler & Unkelbach, 2011;

Unkelbach & Fiedler, 2016). One possibility would consist in suppressing CS-US pairs, rather than either the CS or the US. However, the limitations discussed for the suppression of USs also apply to such a variant of the CFS paradigm.

A further caveat is our use of a memory-based awareness measure after conditioning as opposed to an online measure of awareness. It has been pointed out that the former may be affected by forgetting and is therefore suboptimal to assess awareness of the CS during conditioning (Balas & Gawronski, 2012; Gawronski & Walther, 2012). Note that this does not fundamentally challenge our conclusions, though: We do not rely on measured awareness as an experimental variable because we straightforwardly manipulated awareness experimentally. Rather, our recognition measure of CSs served as a manipulation check. Beyond that, Hütter and colleagues (2012) showed that awareness measures administered after conditioning might not only reflect explicit memory of the conditioning stimuli. In the absence of memory, participants' conditioned attitudes might also inform their recognition judgement in a way that stimuli that acquired valence are selected as the ones that were presented even in the absence of recognition. We precluded that possibility in Experiment 3.2 because we obtained an estimate for the visibility of CFS-suppressed stimuli from different stimuli than the CSs. Those stimuli underwent the same CFS procedure as CSs previously did, only they were paired with neutral instead of affective stimuli on the unsuppressed eye. The recognition data for these stimuli can therefore not be confounded with conditioned attitudes. In Experiment 3.3 and 3.4, in contrast, we assessed recognition of the CSs proper for the sake of analyses of CS evaluations conditional on whether they were recognized or not. In this measure, we assessed the proportion of recognized CSs. Experiment 3.3 and 3.4's awareness measure thus concerns the identity of the CS rather than the US (valence) associated with the CS. Thereby we deviate from the typical valence awareness measure used in EC research. However, as the recognition of the CS is a necessary precondition for the identification of the US, we chose the measure that is most informative with regard to the success of our manipulation.

Finally, the central limitation of this work is the logical impossibility to prove the non-existence of a phenomenon. Aware of this impossibility, the present efforts aimed to show EC without awareness. Yet, despite our best efforts, we might have failed to create the ideal conditions to obtain EC without awareness; or the effect might be so small that it was not possible to detect it given the present sample sizes. The accumulation of null findings (e.g., Dedonder et al., 2014; Stahl et al., 2016) might reduce the likelihood of learning without awareness; and one might then consider a single process approach more parsimonious, but ultimate proof will remain logically impossible.

Implications and conclusion. Studying EC effects with CFS simulates how preference formation without awareness could happen in everyday life. Stimuli in our environment usually have a bottom-up strength that is sufficient for perception (e.g., billboards) but our attention is focused on something else (e.g., navigating traffic). Therefore, suppressing stimuli with CFS is a highly controlled, yet externally valid experimental analog of such situations. This analogue, however, does not support the formation of likes and dislikes without awareness in our experiments. Nevertheless, our findings neither imply that the acquisition process cannot be defined by other automaticity criteria, nor that the retrieval and usage of evaluations cannot operate in an automatic manner. For instance, the influence of evaluations on behavior may be controllable or uncontrollable, depending on the state of cognitive load. Our conclusion is thus both a strong and a cautious one: The accumulated scientific evidence at the present stage does not support an evaluative learning process that operates without awareness.

Chapter 4: The role of relational qualifiers in attribute conditioning: Does disliking an athletic person make you unathletic?

Abstract

In attribute conditioning (AC), stimuli (CSs) acquire attributes through mere pairings with other stimuli possessing that attribute (USs). If neutral “Neal” is paired with athletic “Wade”, participants judge Neal as more athletic compared to when Wade would be unathletic. Prior research suggests that a CS-US link mediates AC effects, but the link’s nature is unclear. The link may be merely referential or propositional. Building on evaluative conditioning research, we introduced relational qualifiers between CS and US to probe the link’s nature; concretely, CSs either liked or disliked USs. Four experiments (n = 811) showed a moderation of AC by this relation: When Neal disliked athletic Wade, he was judged as unathletic. This was partly due to (dis)liking signaling (dis)similarity between Neal and Wade. Thus, CS and US seem to be propositionally linked. We discuss other processes that might contribute to AC and the paradigm’s relation to spontaneous trait transference.

The role of relational qualifiers in attribute conditioning: Does disliking an athletic person make you unathletic?

How do people, consumer goods, or stimuli in general acquire their attributes or traits? How does a brand become elegant, a cereal healthy, or a person athletic? One simple way how stimuli acquire attributes is Attribute Conditioning (AC; Förderer & Unkelbach, 2015): By merely pairing a stimulus with another stimulus possessing a certain attribute, the first stimulus acquires this attribute. For example, showing a neutral person (e.g., a picture of a face) together with another athletic person (e.g., a picture of a person playing soccer) makes participants' assessment of the initially neutral person more athletic (Förderer & Unkelbach, 2011). In conditioning terms, the former is the conditioned stimulus, CS (i.e., neutral before the pairing), and the latter is the unconditioned stimulus, US (i.e., athletic before the pairing). While AC is well-established on an effect level (Staats & Staats, 1957; see Unkelbach & Högden, in press, for an overview), the mental processes underlying it are not clear yet (see Unkelbach & Förderer, 2018). In particular, Förderer and Unkelbach (2016) proposed that AC effects are due to a link between the CS and US's mental representations (see below). However, they did not specify the nature of the link; for example, it may be a mere associative link (Gawronski & Bodenhausen, 2011) or a propositional link (Mitchell et al., 2009). The present research addresses the nature of this link by introducing semantic qualifiers of the CS-US relations; specifically, whether the CS likes or dislikes the US. In more colloquial terms, does disliking an athletic person make you unathletic? By answering this question, we constrain the potential processes that may underlie AC.

In the remainder, we provide a short overview of AC research up to date, and in particular, how AC relates to Evaluative Conditioning (EC). Then, we delineate how relational qualifiers may inform the processes underlying attribute conditioning. Finally, we present four experiments (total n = 811) investigating if and how relational qualifiers change CS assessments after pairing them with athletic or unathletic USs.

AC effects and potential AC processes. AC is a reliable and versatile phenomenon (Unkelbach & Högden, in press; Unkelbach & Förderer, 2018). The effects are found on direct measures such as explicit ratings, and more indirect measures such as semantic priming or semantic misattribution (Förderer & Unkelbach, 2011). AC pairings have been shown to change people's stimulus assessments of "speed" or "softness" (Kim, Allen, & Kardes, 1996), "size" (Olson, Kendrick, & Fazio, 2009; Exp. 2), or "humor", "attractiveness", "intelligence", or "athleticism" (Förderer & Unkelbach, 2014).

However, all these dimensions have evaluative connotations; "fast", "funny", or "athletic" are typically positive attributes. Thus, it is important to show that AC goes beyond creating overall positive or negative evaluations as in EC. In EC, people's evaluation of a CS typically changes in the direction of the evaluation of a paired US (Gast, Gawronski, & De Houwer, 2012). Thus, AC effects may be generalized effects of conditioned valence on the provided rating dimension (i.e., "halo" effects; Gräf & Unkelbach, 2016; Nisbett & Wilson, 1977). To differentiate AC effects from EC, Förderer and Unkelbach (2011) showed that AC effects are still present if one controls statistically for the evaluation of a given CS (see Förderer & Unkelbach, 2014; for an experimental approach). Thus, AC cannot be fully accounted for by general liking or disliking, but is a genuine phenomenon in its own right.

Building on similar theorizing in EC research (Baeyens et al., 1992), Unkelbach and Förderer (2018) suggested a referential learning process that creates a link between CS and US. This link was experimentally tested by Förderer and Unkelbach (2016) in a *revaluation* paradigm (e.g., Walther, Gawronski, Blank, & Langer, 2009). They paired CSs with athletic or unathletic USs. As expected, participants judged the CSs in accordance with the US attributes (i.e., athletic or unathletic), both on direct and indirect measures. Yet, after the pairings, the USs changed their attributes from athletic to unathletic (e.g., a runner becoming visibly chubby) and vice versa (e.g., a chubby person becoming visibly muscular and lifting weights). Without further pairings, this revaluation influenced CS assessments. That is, a CS that was paired with a formerly athletic US which had then turned unathletic was judged as less athletic than a CS that was paired with a formerly unathletic US which had turned athletic. This implies a CS-US link. Next, we address the potential nature of this link.

Potential CS-US links. In EC, two candidates for CS-US links are associations and propositions (Gawronski & Bodenhausen, 2011). In their model of AC effects, Unkelbach and Förderer (2018) used the more general term *referential links* instead of associations because the latter are historically grounded in the idea of spreading activation: When the CS is “activated”, for example by presentation or recall, activation spreads along the links and activates the US, which in turn influences evaluations. However, spreading activation models face substantial theoretical challenges (Ratcliff & McKoon, 1981). Referential links can also be thought of from a distributed memory perspective (e.g., McClelland, McNaughton, & O’Reilly, 1995). This framework conceptualizes the CS-US link as a shared context of CS and US in memory as opposed to an association between the mental representations of CS and US (Gast, 2018; Unkelbach & Förderer, 2018). The relevant distinction between the two candidates is the same, though: Referential links are unqualified connections while propositional links carry meaning. They allow information about the CS-US link, they can be true or false, and they can be subject to logical reasoning (Mitchell et al., 2009).

One way to probe the nature of the CS-US link are relational qualifiers; that is, semantic qualifications of the link. In EC research, a number of such qualifiers have been introduced, such as CSs being friends versus enemies of USs (Fiedler & Unkelbach, 2011), CSs loving or hating USs (Förderer & Unkelbach, 2012), or CSs starting or stopping USs (Moran & Bar-Anan, 2013). If CS-US links are propositional, semantic qualifiers should influence EC effects, while mere referential links should be insensitive to such information (Hu, Gawronski, & Balas, 2017a).

Numerous studies showed a moderation of EC by relational information (e.g., Moran & Bar-Anan, 2013; Unkelbach & Fiedler, 2016), with attenuation or complete reversals of EC effects given negative CS-US relations (e.g., “stops”, “hates”, “dislikes”). Thus, EC effects seem to involve propositional information. It is important to note though that data on relational qualifiers cannot distinguish between propositional models (e.g. Mitchell et al., 2009) and dual-process models of both associations and propositions (e.g., Gawronski & Bodenhausen, 2018). Evidence for the presence of one process does not preclude the existence of another process.

The explanation for such moderation by propositional information in EC is straightforward: Participants should dislike stimuli that are in a negative relation to stimuli they like and vice versa. Concretely, if an initially neutral person X (i.e., the CS) dislikes a friendly person (i.e., the US), person X becomes less likeable. For EC, liking and disliking are general evaluative relations and may, thus, inform evaluative judgements. For AC, the situation is more complex. If an initially neutral person Y (CS) dislikes an athletic person (US), person Y should not necessarily become unathletic; although one may argue that many attributes have evaluative implications, in principle, one may construe an orthogonal relation of specific attributes and evaluations; that is, one may like or dislike both athletic and unathletic people, without implications for one’s own perceived athleticism.

There are several ways a moderation by relational qualifiers may nevertheless come about for AC effects. We will use a negative relation (“dislikes”) as an example. First, attributes also have positive or negative connotations (e.g., being “athletic” is rather good than bad and being “unathletic” is rather bad than good). If Person Y dislikes another athletic person, Person Y is evaluated more negatively, which in turn makes Person Y unathletic (e,g,, “Person Y dislikes something good and is therefore bad; hence, Y is unathletic”). Thus, potential moderating influences of relational qualifiers might be a generalized halo effect (Nisbett & Wilson, 1977). Second, if participants observe person Y in a negative relation with an athletic person (e.g., a US person doing sports), participants might infer a negative relation between the CS and the activity (e.g., “Person Y does not like sports; hence, Y is unathletic”). And third, participants might construe a “dislike” relation as a dissimilarity relation (e.g., “Person Y is different from the person doing sports”), due to the subjective link of interpersonal attraction (i.e., “dislike”) and similarity (Alves, Koch, & Unkelbach, 2016; Berscheid, 1985).

It is thus an intriguing theoretical and empirical question if and how relational qualifiers influence AC effects. In the following four experiments we tested whether and how relational qualifiers between CS and US influence AC.

Overview of the experiments. We used athleticism as the target attribute, based on the materials by Förderer and Unkelbach (2016); that is, participants observed pairings of CSs with either athletic und unathletic USs and assessed CSs’ athleticism, both on direct and indirect measures. We used “like” versus “dislike” as relational qualifiers.

Experiment 4.1 showed that relational qualifiers influence AC effects, using a direct rating measure. We also collected participants’ CS liking ratings and found that the relations’ influence did not depend on participants’ evaluations. Experiment 4.2 replicated the influence using an indirect measure, a semantic variant of the affective misattribution procedure (AMP, Payne, Cheng, Govorun, & Stewart, 2005a). To preclude that participants infer a direct relation of CSs with the attribute ("CS dislikes sports"), Experiment 4.3 and 4.4 used a second-order conditioning procedure. That is, we conditioned athleticism to a CS_1 and then participants observed a positive or negative relation between CS_1 and CS_2; thus, the relevant CS_2 never occurred together with sports. Experiment 4.4 addressed whether participants interpret the like/dislike relation as CS-US similarity or dissimilarity and if this may account for the relational qualifiers' influence on AC.

We report 4 experiments out of 6 that we conducted in this research line. We conducted an additional study before Experiment 4.3 with the same parameters as Experiment 4.3. The difference was that we used twelve CS-US pairs like in Experiment 4.1 and 4.2. Although the results descriptively mirrored those reported for Experiment 4.3, we did not observe any significant effects when analyzing the main DV. We therefore ran the same study with a reduced number of pairings to reduce attention and memory load in participants and to shorten the duration of the online experiment. A pooled analysis of attribute ratings from the additional study and Experiment 4.3 showed the same pattern of findings as reported for the data of Experiment 4.3 alone. Another experiment was conducted between Experiment 4.3 and 4.4. Similar to Experiment 4.4, it aimed to test whether participants construe the like/dislike relation between CS and US as (dis)similarity. The experiment had a design flaw, though. The question probing if participants interpret (dis)liking as (dis)similarity used the same stimuli as those presented during conditioning. This confounds the potential similarity-liking interpretation with previous responses (i.e., participants who showed strong AC effects should also show a high similarity-liking relation, as the CS and US became subjectively more similar/dissimilar on the athleticism dimension). The reported Experiment 4.4 avoids this problem. While the omitted experiment did not allow inferences regarding participants' interpretation of the relation, it significantly replicated the influence of relational qualifiers on AC.

For all reported experiments, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures. In all experiments, we routinely asked participants in an open question what they thought the purpose of the study was. We did not analyze this data and do not report. We aimed to impede possible demand effects by using an indirect measure in Experiment 4.2 and a second-order conditioning procedure in Experiment 4.3 and 4. Further, we excluded participants who indicated to not have taken part seriously (none in Experiment 4.1, three in Experiment 4.2, two in Experiment 4.3, one in Experiment 4.4; see Aust, Diedenhofen, Ullrich, & Musch, 2013).

Experiment 4.1: Relational qualifiers influence AC effects

Experiment 4.1 tested whether a like/dislike relation between CS and US influences participants' athleticism assessment in a rating measure in a standard AC paradigm.

Method.

Participants and design. One hundred and fourteen people participated in Experiment 4.1 (mean age: 33.38 years, 48 female, 65 male, 1 unspecified). To determine our sample size, we took a priori into consideration the effect size of AC that we previously observed in our lab (e.g., Förderer & Unkelbach, 2011, 2016) and the effect size of relational qualifiers (e.g., Förderer & Unkelbach, 2012). We typically observe significant moderation of EC effects by relational qualifiers (i.e., the interaction) using a within-participants design with 40 to 60 participants in the laboratory. Assuming that the online data collection introduces more noise, we aimed for a sample of around 100 people to yield sufficient power to detect a potential effect of relations on

AC. We recruited participants on Amazon Mechanical Turk and they received a small monetary reward.

We manipulated within participants whether CSs were paired with athletic or unathletic USs and whether CS and US liked or disliked each other. Participants assessed CS athleticism and evaluated CS likability on explicit rating scales.

Procedure and material. We programmed an online experiment with SoSci Survey (Leiner, 2016). The program displayed information about the experiment and confidentiality of participants' data and asked participants for their consent to participate. They stated their age and gender and proceeded to the instructions for the learning phase. The instructions informed them that they would see "photos and drawings of men who either like or dislike each other" and that they should watch the pictures attentively. The last sentence read "You will be asked about the pictures at the end of the experiment".

The learning phase paired 12 black-and-white photos of men (CSs) with six drawings of men performing athletic activities (USs; e.g., cycling, running) and six drawings of men performing unathletic activities (e.g., watching TV, eating on the couch). CS men and US men were both displayed with a random male name to facilitate differentiation between the CSs. Orthogonally, CSs were assigned to "like" or "dislike" the USs. The program randomly created fixed CS-US pairs ("one-to-one" procedure; see Stahl & Unkelbach, 2009) and presented each pair six times, resulting in 72 attribute conditioning trials. Figure 11 shows an example trial. For each trial, a CS appeared alone for 500 ms on the left of the screen, then the relation appeared on its right side. CS and relation were shown together for another 500 ms and then the US appeared on the right side of the screen. CS, relation, and US were shown together for three seconds. There was an inter trial interval of 500 ms. The presentation order was random.

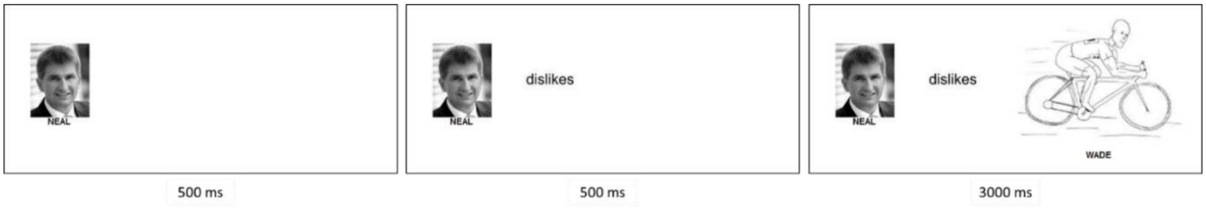


Figure 11. Example trial of "Neal disliking Wade on a bicycle" from Experiment 4.1 and 4.2. The CS is shown alone for 500 ms, then the relation appears and they are shown together for another 500 ms and lastly, the US appears. All three elements are shown for 3 s.

After conditioning, the program showed the instructions for the attribute ratings. Participants' task was to "to indicate how athletic you think certain persons are". During attribute ratings, the program showed a CS on the upper half of the screen and a continuous rating slider below, ranging from 1 ("unathletic") to 101 ("athletic", only the labels were visible to participants). All CSs were rated once, resulting in 12 attribute rating trials. CS presentation order was random. Next, the program asked participants to "indicate how much you like certain persons". Liking ratings of all 12 CSs in random order followed, with similar presentation parameters as for attribute ratings except that the labels of the rating slider were "not at all" to "very much".

After liking ratings, the program asked participants to indicate with a slider whether they like to exercise ("not at all" to "very much") and how athletic they are ("not at all athletic" to "very athletic").[11] Finally, participants were thanked and rewarded.

[11] For all four experiments we ran regressions, with the categorical variables attribute and relation as predictors and either responses to the question whether they like to exercise or responses to the question how athletic they are as continuous predictors. As dependent variables we used attribute ratings and attribute ratings controlled for liking. In none of the analyses any of the two control variables had any effect.

Results.

Attribute ratings. Figure 12 shows participants' mean attribute ratings of CSs liking and disliking athletic and unathletic USs. We conducted a 2 (attribute: athletic vs. unathletic) x 2 (relation: like vs. dislike) repeated measures ANOVA with athleticism ratings as dependent measure. Overall, participants rated CSs paired with athletic USs as more athletic (*M* = 51.18, *SD* = 10.37) than CS paired with unathletic USs (*M* = 48.36, *SD* = 9.76), *F*(1, 113) = 6.27, *p* = .014, η_p^2 = .05, 95% confidence interval (CI) [0.00, 0.15]. This attribute conditioning main effect was qualified by an interaction with relation, *F*(1, 113) = 8.80, *p* =.004, η_p^2 = 0.07, *95%CI*[0.01, 0.18]. The main effect of relation in the overall ANOVA was not significant, *F*(1,113) = 1.99, *p* = .161, η_p^2 = 0.02, *95%CI*[0.00, 0.09]. Thus, the relational qualifiers significantly moderated the AC effect.

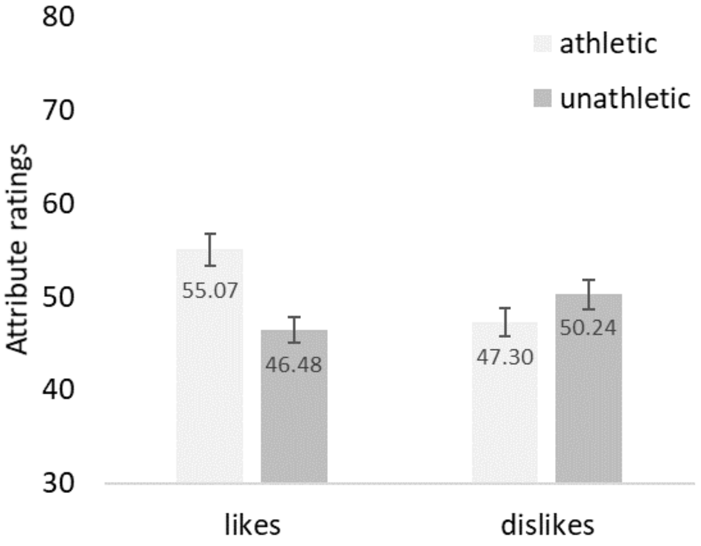


Figure 12. Experiment 4.1: Attribute ratings of CSs that were paired with an athletic or unathletic US and that either liked or disliked the US. Error bars represent the standard error of the mean.

Liking ratings. We conducted the same ANOVA as described above with liking ratings as dependent measure. We observed a main effect of relation. Participants rated CSs liking USs as more likable (*M* = 52.44, *SD* = 14.28) than CSs disliking USs (*M* = 45.15, *SD* = 14.25), *F*(1,113) = 19.86, *p* < .001, η_p^2 = 0.15, *95%CI*[0.05, 0.27]. This main effect of liking is a standard result for relational qualifiers (e.g., Fiedler & Unkelbach, 2011; Walther, Langer, Weil, & Komischke, 2011). No other effects were significant, all *F*s < 1.00, all *p*s > .34, all η_p^2s < 0.01.

Attribute ratings controlling for liking. To investigate if the interaction in attribute ratings might be due to "evaluative" variance (i.e., accounted for by liking of the CS), we conducted a multi-level regression analysis. We treated CS attribute ratings as the dependent variable and liking rating of the same CSs as predictor and included a random intercept for participants. We used the residuals of this analysis, which statistically correct for evaluations of a given CS, in the same ANOVA as described above (see Förderer & Unkelbach, 2011). We still observed an attribute main effect. Participants rated CSs paired with athletic USs as more athletic than CSs paired with unathletic USs, *F*(1, 113) = 6.80, *p* = .010, η_p^2 = 0.06, *95%CI*[0.00, 0.15]. The attribute by relation interaction was also still present, *F*(1, 113) = 8.31, *p* = .005, η_p^2 = 0.07, *95%CI*[0.01, 0.17]. Thus, even when we statistically controlled for liking, the relational qualifiers significantly moderated the AC effect.

Discussion. On an effect level, Experiment 4.1 showed that relational qualifiers influence AC effects. When CS and US "liked" each other, CSs paired with athletic USs were evaluated as more athletic than CSs paired with unathletic USs. When they "disliked" each other, this pattern was reversed as shown by the interaction.

However, Experiment 4.1 raises two concerns. First, we observed the interaction on an explicit rating measure. A concern in pairing paradigms is that participants simply understand the experiment's purpose and control their responses accordingly; that is, they show "demand effects". Therefore, we aimed to replicate the pattern with an indirect measure in Experiment 4.2. Second, as Figure 11 illustrates, participants might misunderstand the CS – likes/dislikes – US relation as the CS likes/dislikes the athletic activity. Experiments 4.3 and 4.4 will address this concern.

Experiment 4.2: Relational qualifiers influence AC effects on an indirect measure

Experiment 4.2 replicated Experiment 4.1 in a laboratory setting and included a semantic variant of the affective misattribution procedure (AMP; Payne et al., 2005) to ensure that AC effects and the relational qualifier influence are not due to demand effects. We called this the semantic misattribution procedure (SMP; Förderer & Unkelbach, 2011). Payne and colleagues developed the AMP to indirectly assess peoples' affective reactions. In the AMP, participants see an affective picture flashed, immediately followed by a Chinese character (Kanji). Then they decide if the Kanji indicates something positive or negative. The idea is that participants misattribute their affective reaction caused by the affective picture (prime) to the Kanji (target). Instead of affective pictures, we used the CSs and asked participants if the Kanji represents a word with an athletic or nonathletic meaning. Assuming an AC effect proper rather than a demand effect, people should decide that a Kanji following an athletic CSs has an athletic meaning more often, compared to Kanji following a nonathletic CSs. The procedural details (see below) of this method make strategic behavior in the sense of demand effects unlikely.

Method.

Participants and design. We aimed for a similar sample size as in Experiment 4.1. One hundred and one students participated in Experiment 4.2 (mean age: 22.05 years, excluding two participants who gave nonsensical age information, 56 female, 45 male). We recruited participants on campus and they participated for course credit or a small monetary reward.

The design was highly similar to Experiment 4.1; the sole design difference was the inclusion of the semantic misattribution procedure as additional dependent variable.

Procedure and material. We conducted the experiment in the laboratory. We programmed the experiment with Open Sesame (Mathôt et al., 2012). We used the same stimuli and setup as in the online study in Experiment 4.1. The procedural changes were as follows: We translated the instructions into German and changed the rating scales for attribute and liking ratings to a range from 1 to 9, with higher numbers indicated higher athleticism and likeability. The first measure we assessed after the learning phase was the semantic misattribution procedure (SMP; Förderer & Unkelbach, 2011, 2016).

The program instructed participants about the procedure of a SMP trial and that their task was to evaluate the Kanji that is presented at the end of each trial. The cover story was to indicate whether they believed the Kanji's meaning was rather athletic or unathletic. They were asked to respond spontaneously and to not be distracted by the preceding photos of men that were presented at the beginning of each trial.

An SMP trial looked as follows: A fixation cross appeared in the center of the screen. After 300 ms, the CS replaced the cross. The CS stayed onscreen for 75 ms. Then, a blank screen followed for 125 ms, followed by a Kanji for 100 ms. A black-and-white pixel image masked the Kanji that was displayed until participants gave a response. Participants responded with the keys "A" and "L" on a German keyboard. We counterbalanced key assignment to the categories "athletic" and "unathletic". Each CS appeared eight times, resulting in 96 SMP trials. Presentation order was random.

After the SMP, we measured attribute ratings, liking ratings, and how much participants like to exercise and how athletic they are themselves.

Results.

Semantic misattribution procedure. Figure 13's upper panel shows the mean proportion of "athletic" responses in each condition. We submitted these to a 2 (attribute: athletic vs. unathletic) x 2 (relation: like vs. dislike) repeated measures ANOVA. We only observed an interaction, *F*(1,100) = 5.17, *p* = .025, η_p^2 = 0.05, *95%CI*[0.00, 0.15]. No other effects were significant, all *F*s < 1.43, all *p*s > .23, all η_p^2s < 0.02.

Attribute ratings. Figure 13's lower panel shows CSs' attribute ratings as a function of paired US and CS-US relation. We analyzed the data with the same ANOVA as in Experiment 4.1. Overall, we found a standard AC effect: Participants rated CSs paired with athletic USs as more athletic (*M* = 5.07, *SD* = 1.10) than CSs paired with unathletic USs (*M* = 4.70, *SD* = 1.06), *F*(1, 100) = 7.80, *p* = .006, η_p^2 = 0.07, *95%CI*[0.01, 0.18]. This AC main effect was qualified by an interaction with relation, *F*(1, 100) = 6.36, *p* = .013, η_p^2 = 0.06, *95%CI*[0.00, 0.17]. Different from Experiment 4.1, we also found a relation main effect. CSs that liked a US were evaluated as more athletic (*M* = 5.01, *SD* = 1.13), than CSs that disliked a US (*M* = 4.76, *SD* = 0.97), *F*(1,100) = 4.03, *p* = .047, η_p^2 = 0.04, *95%CI*[0.00, 0.13].

Liking ratings. We found an attribute main effect. Overall, CSs paired with athletic USs were rated as more likable (*M* = 5.09, *SD* = 1.05) than CSs paired with unathletic USs (*M* = 4.80, *SD* = 1.15), *F*(1,100) = 6.39, *p* = .013, η_p^2 = 0.06, *95%CI*[0.00, 0.17]. There was also an interaction with relation, *F*(1,100) = 4.50, *p* = .036, η_p^2 = 0.04, *95%CI*[0, 0.14]. CSs that liked athletic USs were rated more positively (*M* = 5.31, *SD* = 1.28) than those that liked unathletic

USs (*M* = 4.77, *SD* = 1.39). CSs disliking athletic USs were also rated slightly more positively (*M* = 4.86, *SD* = 1.42) than CSs disliking unathletic USs (*M* = 4.82, *SD* = 1.55).

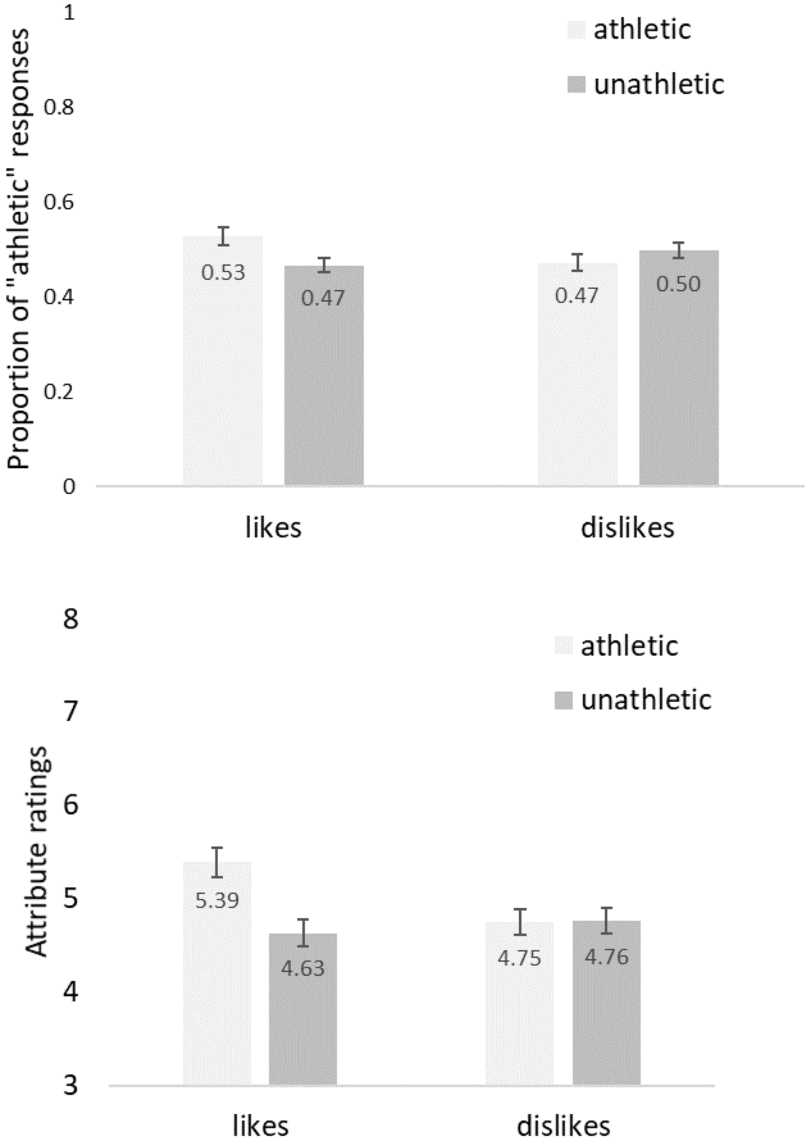


Figure 13. Experiment 4.2: Proportion of "athletic" responses in the semantic misattribution procedure (upper panel) and attribute ratings (lower panel) of CSs paired with an athletic or unathletic US and that either liked or disliked the US. Error bars represent the standard error of the mean.

Attribute ratings controlling for liking. Different from Experiment 4.1, the means of attribute and liking ratings corresponded. We thus conducted again a multi-level regression analysis predicting attribute ratings from liking ratings. Statistically, these residuals are corrected for variance in liking. The ANOVA of these residuals showed again a main effect of attribute. Participants rated CSs paired with athletic USs more athletic than CSs paired with unathletic USs, $F(1,100) = 5.09$, $p = .026$, $\eta_p^2 = 0.05$, *95%CI*[0.00, 0.17]. Controlling for likeability variance, we again found the interaction with relation, $F(1,100) = 4.63$, $p = .033$, $\eta_p^2 = 0.04$, *95%CI*[0.00, 0.14]. The effect main effect of relation was no longer significant, $F(1,100) = 2.75$, $p = .101$, $\eta_p^2 = 0.03$, *95%CI*[0.00, 0.04].

Discussion. Experiment 4.2 replicated Experiment 4.1. The CS-US relation moderated the AC effect and we also found this moderation effect on an indirect measure. As the interaction indicates, the AC effect was significantly different for CS that liked the US compared to CS that disliked the US. We observed this pattern on the proportion of “athletic” responses in the SMP, attribute ratings, and attribute ratings controlled for liking.

Experiment 4.2’s SMP measure also increases the confidence that the results are not due to participants strategic responding (i.e., demand effects); on average, participants responded within 1018.59 ms (*SD* = 1383.27). Within that time, it seems unlikely that participants recognized the CS face, remembered the paired US, the CS-US relation, and that the experiment seems to require a response contingent on US attribute and CS-US relation, despite being instructed to ignore the CS face in the first place. And similar to Experiment 4.1, the residual analysis suggests that the relational qualifiers’ influence does not, at least not fully, depend on the evaluative connotation of the attribute (i.e., a CS disliking something good).

However, both Experiment 4.1 and 4.2's learning phase highlights the anticipated possibility for the relational qualifier's moderating influence (see Figure 11): We presented pictures of male faces (CSs) together with pictures of men engaging in athletic or unathletic activities (USs). In between the pictures the word “likes” or “dislikes” appeared. Thus, participants may infer “This CS dislikes this athletic activity”. Concretely, the setup in Figure 11 might imply that “Neal dislikes cycling.” Experiment 4.3 will address this possibility. In addition, Experiments 4.3 and 4.4 substantially increase the sample size.

Experiment 4.3: Relational qualifiers influence AC effects in a second-order conditioning procedure

Experiment 4.3 tested if the relational qualifiers' influence depends on the relation between CS and US (i.e., “Neal dislikes Wade”) or on the relation between the CS and the activity or the CS and the attribute (i.e., “Neal dislikes cycling” or “Neal dislikes athletic activities”). To this end, we separated CS-US pairings and presentation of the relational qualifier in a second-order conditioning procedure. First, we paired CS_1s with athletic and unathletic USs in a standard AC procedure without relational qualifiers. Second, we paired these CS_1s, which are now effectively novel USs, with CS_2s in a second conditioning phase. These pairings now included the relational qualifiers of “likes” and “dislikes”. Importantly, the relevant CSs were now the CS_2s, which were never presented with the athletic activity. Observing a relational qualifier influence in this second-order conditioning procedure would suggest that the effect is not due to a perceived relation of CS_2 with the activity or the attribute itself.

Method.

Participants and design. Two hundred ninety-four people were recruited on Amazon Mechanical Turk and participated for a small monetary reward (mean age: 37.56, 123 female, 169 male, 1 other, 1 unspecified). We increased the sample size substantially in comparison to Experiment 4.1 and 4.2 because we anticipated smaller second-order conditioning effects and we aimed to increase confidence in the observed effects.

We manipulated whether CS_1s were paired with an athletic or unathletic USs and whether CS_2 and CS_1 liked or disliked each other within participants and measured athleticism and liking ratings of all CS_2s and only athleticism ratings of CS_1s.

Procedure and material. The SoSci Survey (Leiner, 2016) online study was similar to Experiment 4.1 with the following changes: First, participants saw pairings of four photos of male faces (CS_1s) and four photos of men engaging in athletic or unathletic activities (e.g. playing soccer, reading a book, USs). That is, the overall number of pairings was reduced to four and the US drawings were replaced by photos. CS_1s were shown alone on the screen for 1 s, then the US appeared to its right and they were shown together for another 3 s. The intertrial interval was 500 ms. Figure 14's upper panel shows an example trial. Every CS_1-US pair was shown six times resulting in 24 first-order conditioning trials. Then, as a manipulation check, we assessed attribute ratings of all CS_1s. Next, the second-order conditioning phase paired the four CS_1s with four CS_2s. In those pairings, the CS_2 either "liked" or "disliked" the CS_1. Figure 14's lower panel illustrates the second phase. CS_2s were presented alone for 1 s, then the relation appeared on its right side and they were shown together for another second. Finally, CS_1s appeared on the very right part of the screen and the three elements were shown together for 3 s. The intertrial interval was 500 ms. Finally, we assessed attribute ratings and liking rating of all CS_2s and the control questions at the end of the experiment. The order of presentation in the two conditioning phases and all measurement phases was random.

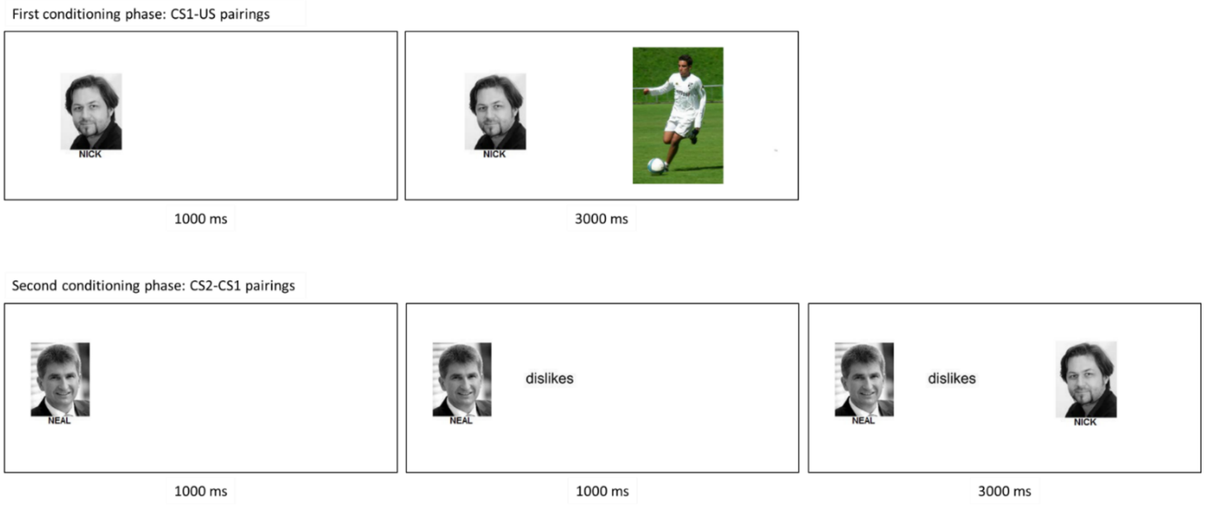

First conditioning phase: CS1-US pairings
NICK
NICK
1000 ms
3000 ms
Second conditioning phase: CS2-CS1 pairings
NEAL
NEAL
dislikes
NEAL
dislikes
NICK
1000 ms
1000 ms
3000 ms

Figure 14. Example trials from Experiment 4.3. In the first conditioning phase CS_1s are paired with athletic or unathletic USs; in the second conditioning phase CS_2s are paired with CS_1s and the two stimuli are either connected by a “like” or “dislike” relation.

Results.

Manipulation check. We analyzed participants’ CS_1 attribute ratings after the first conditioning phase in a *t*-test. As expected, participants rated athletic-paired CS_1s as more athletic (M = 83.09, SD = 17.61) than unathletic-paired CS_1s (M = 30.18, SD = 20.39), $t(293) = 28.34$, $p < .001$, $d_z = 1.65$.

Attribute ratings. Figure 15 shows attribute ratings of the CS_2s as a function of CS_1-paired attribute (athletic vs. unathletic) and CS_2-CS_1 relation (like vs. dislike). The respective ANOVA showed only an interaction, $F(1, 293) = 48.07$, $p < .001$, $\eta_p^2 = 0.14$, *95%CI*[0.07, 0.21]. The main effect of attribute was not significant ($F(1, 293) = 1.55$, $p = .214$, $\eta_p^2 < 0.01$, *95%CI*[0.00, 0.03]); the main effect of relation was also not significant. Descriptively, participants rated CS_2s liking CS_1s as more athletic (M = 52.87, SD = 28.17) than CS_2s disliking CS_1s (*M* = 50.59, *SD* = 27.75), *F*(1, 293) = 3.07, *p* = .081, η_p^2 = 0.01, *95%CI*[0.00, 0.04].

Liking ratings. The liking ratings again showed a main effect of relation. Participants liked CS_2s liking CS_1s more (*M* = 53.91, *SD* = 22.12) than CS_2s disliking CS_1s (*M* = 42.54, *SD* = 22.94), *F*(1,293) = 70.70, *p* < .001, η_p^2 = 0.22, *95%CI*[0.12, 0.27]. No other effects were significant, all *F*s < 0.14, all *p*s > .71, all η_p^2s < 0.01.

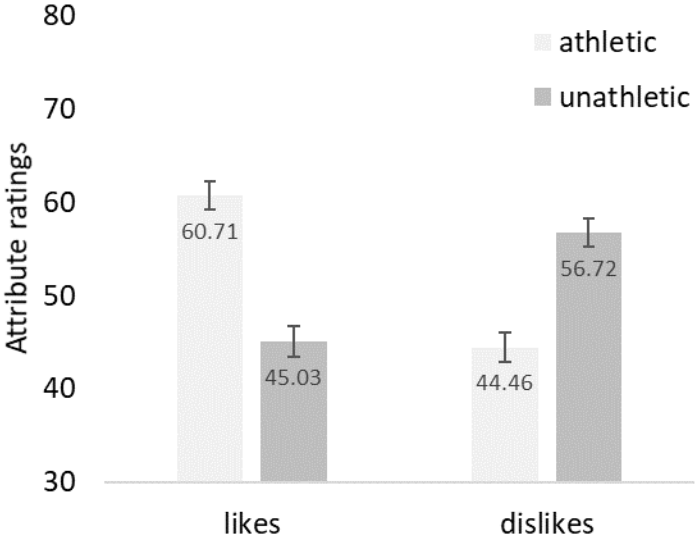


Figure 15. Experiment 4.3: Attribute ratings of CS_2s that were paired with a CS_1 that had previously either been paired with an athletic or unathletic US. CS_2 and CS_1 either liked or disliked each other. Error bars represent the standard error of the mean.

Attribute ratings controlling for liking. Similar to Experiment 4.1 and 4.2, we repeated the main analysis based on the residuals of the multi-level regression predicting attribute ratings from liking ratings. The respective ANOVA of the residuals replicated the interaction between attribute and relation, F(1,293) = 48.20, p < .001, η_p^2 = 0.14, *95%CI*[0.08, 0.21]. No other effects were significant, all Fs < 1.72, all ps > .19, all η_p^2s < 0.01.

Discussion. Experiment 4.3 replicated the relational qualifiers' influence on AC using a second-order conditioning setup. This setup has several methodological advantages. Foremost, the clear interaction suggests that the relational qualifiers exert their influence on the link between the paired stimuli, and not a potential link between the CS and the depicted activity or the relevant attribute (i.e., "Neal dislikes cycling"). Rather, the AC effect is moderated by the negative relation of CS and US, or here, between CS_2 and CS_1 (i.e., "Neal dislikes Wade"). This effect was also independent of CS_2 likeability; the effect sizes for the attribute ratings and the attribute ratings controlled for likeability were virtually identical.

Experiments 4.1-3 thereby showed the moderating influence of relational qualifiers on AC effects, while precluding our first and second potential explanation. That is, the effect is not based on generalized liking and not based on participants inferring that the CSs (dis)like (un)athletic activities. Yet, it would be premature to conclude that the proposed CS-US link is therefore propositional in nature. This reasoning is valid for EC effects, but less straightforward for AC effects. As we have argued above, participants disliking a CS that dislikes something good is justified, as the evaluative responses align (i.e., participants are justified to dislike Neal if the setup shows "Neal dislikes cute puppies", given they like puppies themselves). However, Neal disliking athletic Wade does not justify the rating that Neal is unathletic; a substantial number of additional assumptions would be necessary. For example, one might argue that a CS disliking someone athletic ("Neal dislikes athletic Wade") makes the CS also less likeable and therefore less athletic. Yet, this explanation is ruled out by the previous experiments.

Alternatively, our third possibility, outlined in the introduction, might provide an explanation: Participants might interpret the relational qualifier in a “is similar” vs. “is dissimilar” fashion; in other words, participants might read “Neal dislikes Wade” as “Neal is unlike Wade” Experiment 4.4 therefore measured the degree to which participants interpret (dis-)liking as (dis-)similarity.

Experiment 4.4: Does liking imply similarity?

Experiment 4.4 replicated Experiment 4.3 and additionally measured to what extent participants interpret the like/dislike relation as implying similarity between CS and US (here: between CS_2 and CS_1) and analyzed this as a mediating variable.

Method.

Participants and design. We aimed for a similar sample size as in Experiment 4.3. Three hundred and two people participated in Experiment 4.4 (female: 156, male: 143, other: 1, unspecified: 2, mean age: 33.90 years). We recruited the sample on Amazon Mechanical Turk.

The design was highly similar to Experiment 4.3, but included one additional measure: We measured the degree to which participants interpret the like/dislike relation as similarity between two stimuli.

Procedure and material. The program and stimuli were similar to Experiment 4.3 with one exception. Before participants saw pairings of CS_1s and the USs (i.e., first-order conditioning phase), they indicated in two trials how similar they thought two men were that liked versus disliked each other. A similarity rating trial looked as follows: A black-and-white photo of a male face was shown alone for 1 sec, then the relation appeared to its right and they were shown together for another second. Then another photo of a male face appeared on the right and all three elements were shown for 3 sec before the question “How similar do you think the two men above are?” appeared below. Participants indicated their similarity rating with a slider from “dissimilar” to “similar”. The stimuli were from the same pool as the CSs but were not used as CSs subsequently. Every relation was rated once, resulting in two similarity rating trials. Whether the like or dislike relation was rated first was determined randomly. Everything else was identical to Experiment 4.3.

Results.

Manipulation check. A *t*-test showed that participants rated CS_1s paired with athletic USs as more athletic (*M* = 81.62, *SD* = 15.70) than CS_1s paired with unathletic USs (*M* = 33.15, *SD* = 19.28), *t*(301) = 28.28, *p* < .001, d_z = 1.63.

Attribute ratings. Figure 16 shows CS_2 attribute ratings as a function of CS_1-paired attribute (athletic vs. unathletic) and CS_2-CS_1 relation (like vs. dislike). The respective ANOVA showed an attribute main effect. Participants rated CS_2s paired with athletic-paired CS_ 1s as more athletic (*M* = 52.42, *SD* = 26.81) than CS_2s paired with unathletic-paired CS_1s (*M* = 46.70, *SD* = 25.22), *F*(1,301) = 15.26, *p* < .001, η_p^2 = 0.05, *95%CI*[0.01, 0.10]. This main effect was qualified by an interaction with relation, *F*(1, 301) = 41.25, *p* < .001, η_p^2 = 0.12, *95%CI*[0.06, 0.19]. The main effect of relation in the ANOVA was not significant, *F*(1,301) = 2.25, *p* = .135, η_p^2 < 0.01, *95%CI*[0.00, 0.04].

Liking ratings. The same ANOVA for liking ratings showed the by now expected relation main effect. CS_2s that liked a CS_1 were rated as more positive (*M* = 51.48, *SD* = 22.27) than CS_2s that disliked a CS_1 (*M* = 40.97, *SD* = 21.61), *F*(1, 301) = 66.22, *p* < .001, η_p^2 = 0.18, *95%CI*[0.11, 0.26]. No other effects were significant, all *F*s < 2.05, all *p*s > .15, all η_p^2s < .01.

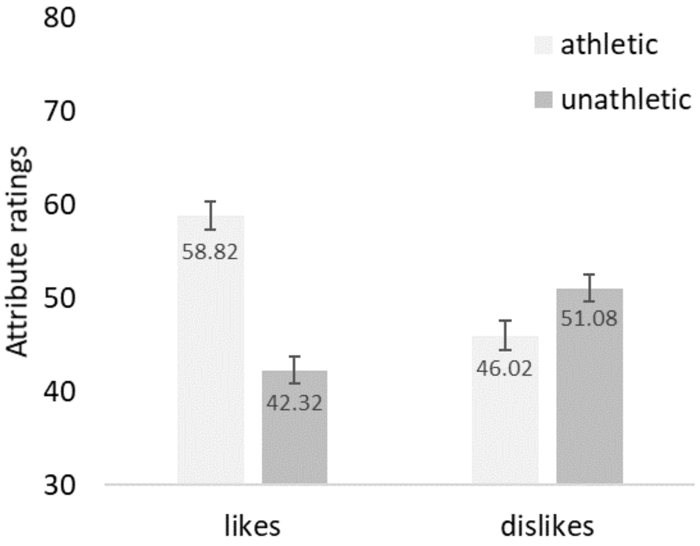


Figure 16. Experiment 4.4: Attribute ratings of CS_2s paired with a CS_1 that had previously either been paired with an athletic or unathletic US. CS_2 and CS_1 either liked or disliked each other. Error bars represent the standard error of the mean.

Attribute ratings controlling for liking. We again predicted attribute ratings from liking ratings in a multi-level regression analysis. The ANOVA of these residuals still showed the same attribute main effect: Participants rated CS_2s paired with athletic-paired CS_1s as more athletic, $F(1, 301) = 15.27$, $p < .001$, $\eta_p^2 = 0.05$, *95%CI*[0.01, 0.10]. We also found again the interaction with relation, $F(1, 301) = 39.39$, $p < .001$, $\eta_p^2 = 0.12$, *95%CI*[0.06, 0.19]. The main effect of relation in the ANOVA was not significant, $F(1, 301) = 0.15$, $p = .703$, $\eta_p^2 < 0.01$, *95%CI*[0.00, 0.02].

Similarity ratings. As described, participants' first two ratings indicated their perceived similarity of two targets that did not appear within the experiment. These ratings serve as an indicator to what extent they interpret a like/dislike relation as a similarity/dissimilarity relation. First, a *t*-test showed that participants saw two men as more similar ($M = 48.97$, $SD = 27.37$) when the first men liked the second man compared to when the first men disliked the second men (*M* = 37.66, *SD* = 26.11), *t*(301) = 5.98, *p* < .001, *d* = 0.34. This suggests that, on average, participants interpreted the like/dislike relation as implying similarity respectively dissimilarity between the two stimuli. To test if this interpretation may explain the relational qualifiers' influence on AC effects, we conducted a regression analysis predicting athleticism ratings from the categorical variables "attribute" and "relation" and the continuous variable "similarity". That is, the athleticism rating of a given CS_2 was predicted by the attribute paired with the CS_1, the CS_2 was paired with, the relation between CS_2 and CS_1 and the similarity rating. Specifically, if a given CS_2 liked a CS_1, the similarity rating from the trial in which two men liked each other was used as a predictor. When a given CS_2 disliked a CS_1, the similarity rating from the trial in which two men disliked each other was used.

Table 2 shows the beta weights for the predictors and their interactions. This analysis shows that both, the interaction of attribute and relation and the interaction of attribute and similarity contribute in predicting attribute ratings in the same direction. That is, if a given CS_2 disliked a CS_1 that was paired with an athletic US, the interaction of attribute and perceived similarity between CS_2 and CS_1 could account for substantial variance in the athleticism ratings of CS_2. But the interaction of attribute and relation shows that the relational qualifier influences AC effects beyond a perceived similarity of CS_2 and CS_1. Given the regression approach, these effects are the unique and independent contributions of these interactions to the overall AC effect.

Table 2. Statistics of the regression analysis predicting attribute ratings from the dichotomous variables attribute (athletic vs. unathletic) and relation (like vs. dislike) and the continuous variable similarity. Asterisks indicate effects that are significant on the standard alpha level.

Effect	**β**	**t**	**df**	**p**
intercept	48.42	34.70	1200	< .001*
attribute	0.87	0.63	1200	.532
relation	0.69	0.49	1200	.623
similarity	0.03	0.94	1200	.347
attribute*relation	-3.65	-2.62	1200	.009*
attribute*similarity	-0.08	-3.01	1200	.003*
relation*similarity	0.004	0.15	1200	.879
attribute*relation*similarity	-0.03	-1.07	1200	.283

Discussion. Experiment 4.4 replicated Experiment 4.3. Relational qualifiers significantly moderated the AC effect in a second-order conditioning setup and controlling for evaluative variance. Going beyond Experiment 4.3, we measured to what extent participants interpret the like/dislike relation as similarity/dissimilarity. This allowed testing our third potential explanation. On the level of means, we found evidence that participants interpreted the relation as similarity/dissimilarity. On the level of persons, this rating of the relation as similarity/dissimilarity, which was collected at the beginning of the experiment, significantly predicted a given CS_2's athleticism rating, depending on its pairing with an athletic-paired or unathletic-paired CS_1. This supports a propositional nature of the CS_2 – CS_1 link, or more generally, the CS-US link. However, even controlling for this similarity interpretation, the like/dislike relation still predicted CS_2 athleticism ratings as a function of CS_1 athleticism (i.e., athletic vs. unathletic). In other words, the interpretation of the relation does not fully explain the relational qualifiers' influence. We will address possible further explanations in the General Discussion.

General Discussion

On an effect level, pairing an initially neutral stimulus (CS) with a stimulus possessing a specific attribute (US) changes observers' rating of the neutral stimulus on this attribute. Concretely, if the initially neutral "Neal" appears together with athletic "Wade", people rate Neal more athletic compared to when "Wade" would be unathletic. We termed this effect *attribute conditioning* (AC) to set it apart from evaluative conditioning (EC; Förderer & Unkelbach, 2015, 2016; Unkelbach & Förderer, 2018). On a process level, Förderer and Unkelbach suggested that the AC effect is due to a link between the CS and the US; however, the nature of the link remains unspecified. It might be associative/referential, propositional, or both.

To investigate the link's nature, we used for the first time relational qualifiers in an AC paradigm. This strategy has been employed numerous times in EC research to investigate the nature of CS-US links (e.g., Hu et al., 2017). In EC, research on relational qualifiers showed that the mental representations of CS and US are propositionally related. We aimed to use this strategy to shed light on the role of propositional relations in AC. Concretely, we asked, how athletic do people rate Neal if he dislikes athletic Wade? As the example shows, the questions if and how relational qualifiers influence AC effects does not only inform mental process theories of AC but is also of applied interests.

Our main question if relational qualifiers influence AC effects has a clear answer. Across four experiments, we always found an interaction of the US attribute and the CS-US relation (that is, the CS_2-CS_1 relation in Experiments 4.3 and 4.4). Across all experiments in the series, we found the interaction five out of six times. Thus, athleticism ratings between athletic-paired and unathletic-paired CSs differed significantly depending on whether the CS liked or disliked the US (or CS_1, to be precise, in Experiments 4.3 and 4.4). In addition, Experiment 4.2 showed this interaction also on an indirect measure. The smallest effect size for the interaction we observed was η_p^2 = 0.04 in the attribute ratings controlling for liking in Experiment 4.2, the highest was η_p^2 = 0.14 in Experiment 4.4. Table 3 summarizes the effect sizes of the moderation in all experiments for all measures of athleticism.

Table 3. Effect sizes (η_p^2) and 95% confidence intervals for the interaction of attribute and relation (i.e., the size of the moderation effect) in all athleticism measures in all experiments.

	Attribute ratings	**Attribute ratings controlling for liking**	**Indirect measure**
Experiment 4.1	**0.07** [0.01, 0.18]	**0.07** [0.01, 0.17]	
Experiment 4.2	**0.06** [0.00, 0.17]	**0.04** [0.00, 0.14]	**0.05** [0.00, 0.15]
Experiment 4.3	**0.14** [0.07, 0.21]	**0.14** [0.08, 0.21]	
Experiment 4.4	**0.12** [0.06, 0.19]	**0.12** [0.06, 0.19]	

The strength of this influence is surprising. Although we used high-powered studies, the manipulations, in particular in Experiments 4.3 and 4.4 were rather subtle. To be precise, a given CS_1 was only together onscreen for 18 s in total with the US. The respective CS_2 was also only presented 18 s together with the relational qualifier and the CS_1. Given the full randomization of the trials and the ratings, the clear interaction pattern across four experiments provides strong support for the existence of the effect. Thus, on an effect level, there is clear evidence for the relational qualifiers' influence.

The question on the process level has a more differentiated answer. We proposed three potential explanations. First, an evaluative connotation of the attribute, which would be in analogy to relational qualifiers in EC research. Second, an impression formation effect in which participants infer a like/dislike relation of the CS with the activity or the attribute itself. And third, an interpretation of the relation as a similarity relation, which would justify propositional influences of the relational qualifiers on AC effects.

The first explanation received little support; particularly Experiments 4.3 and 4.4 showed almost no change in effect sizes when we controlled for participants' evaluative variance in the attribute ratings. We investigated the second explanation by using a second-order conditioning procedure (Experiments 4.3 and 4.4). Thereby, the target CS never appeared together with the attribute or the activity, but only with another stimulus that we previously conditioned to athletic or unathletic. If the second explanation would be valid, this setup should substantially reduce the relational qualifiers' influence. The opposite was the case. For the second order conditioning setup (Experiment 4.3 and 4.4), we observed numerically stronger effects in comparison to the direct pairings, in which target CS and the activity were presented together. The third explanation was partially supported in Experiment 4.4. Participants' perceived similarity of stimuli (dis)liking each other influenced the AC effect on athleticism ratings. This predictive effect is indicative of a propositional nature of the CS-US link in AC. It shows that reasoning along the lines of "Neal dislikes athletic Wade. People who dislike each other are usually dissimilar. Therefore, Neal is unathletic." contributes to the effect. However, a substantial variance part was still due to the effect of the relational qualifier on AC (s. Table 2).

Potential explanations for the residual effect. At this point, we may only speculate about the explanation for this remaining variance. One possibility is the triadic CS-predicate-US model suggested by Unkelbach and Fiedler (2016) for relational qualifiers in EC research. It explained reversal effects with a "functional predicate" that indicates the relation of CS and US. Importantly, this predicate is not propositional. It has, for example, been shown that animals who are unlikely to engage in propositional reasoning, are also sensitive to specific relations between CS and US. Flaherty and Rowan (1986) observed a phenomenon referred to as negative anticipatory contrast in rats. For negative anticipatory contrast, a liked CS (e.g., saccharine solution) is followed by an even more liked US (e.g., sugar solution). Afterwards the CS is preferred less, although the CS predicted the US and should, accordingly, be preferred more. This shows that the specific relation (e.g., X is better than Y) can affect conditioning in animals. Therefore, if animal learning may be sensitive to relational predicates it is not necessary to explain influences of relations between CS and US in propositional terms.

Another, theoretically interesting possibility is that referential links may be excitatory or inhibitory (Wagner, 1981; Wheeler, Sherwood, & Holland, 2008). That is, the activation of a CS may foster (excitatory) or inhibit (inhibitory) the US representation. By default, referential links from stimulus pairings would then be excitatory (Wagner, 1981), resulting in the typical AC effect. Learning phases in which CS and US are negatively related, however, may create inhibitory links. In the present case, excitatory links in the "like" condition and inhibitory links in the "dislike" condition may then lead to relational qualifier influences which are not mediated by propositions about the CS-US relation. However, at present these considerations are speculative and further experiments will be necessary to determine the exact nature of the CS-US link structure, besides participants perception of the like/dislike relation as an indicator of CS-US similarity or dissimilarity.

Spontaneous trait transference as an alternative explanation. AC has procedural similarities with the spontaneous trait transference (STT) paradigm (Skowronski, Carlston, Mae, & Crawford, 1998). In STT, a communicator describes another persons' behavior. The behavior implies a certain trait, for example, playing soccer implies that the target person is athletic. The trait is then not only ascribed to the target person exerting the behavior but also to the communicator. In STT, communicator and target person co-occur and the trait is implied. In AC, CS and US co-occur and the attribute is implied. Carlston and Skowronski (2005) showed that STT is unaffected by presentation time and that it is uncontrollable; that is, STT emerges even if participants are instructed not to apply the trait to the communicator. They concluded that STT is unlikely to be mediated by deliberate, "attributional" processes. Instead, they suggest that communicator and target are merely associated and that those associations influence subsequent judgments of the communicator. This explanation is highly similar to a referential link account of AC. Here, however, we show that relational qualifiers do moderate the AC effect, and that this moderation is partially due to participants' interpretation of the CS-US relation. The present results thereby differentiate the phenomena both on a functional and a process level. It is an intriguing question though, if STT effects might not also be subject to the influence of relational qualifiers, which would then indicate a shared basis for AC and STT effects.

Limitations and open questions. While the basic influence of the relational qualifiers on AC effects is intriguing, we must also concede several limitations that emerged because we remained within a single paradigm. Consequently, we have restrictions in terms of sampling CSs, USs, relations, and attributes. That is, we used a set of only 12 male CS in total, and 16 US (12 drawings in Experiment 4.1 and 4.2, and four photos in Experiment 4.3 and 4.4) in total. We used only one attribute, athleticism, and only one relational qualifier, namely likes and dislikes. In addition, we only controlled evaluative variance statistically, and not experimentally (see Förderer & Unkelbach, 2014, for an experimental approach). While we firmly believe that the observed effects are generalizable, it is a question for the future if one may observe the same pattern for, for example, female CSs that hate another intelligent female US.

The patterns observed for liking ratings differed between Experiment 4.1, 4.3 and 4.4, that were conducted online with an US American sample and Experiment 4.2 that was conducted in the laboratory with a German student sample. Participants in online experiments consistently showed a main effect of relation, that is, they liked CSs that liked others and disliked those that disliked others; this effect was already reported by Fiedler and Unkelbach (2011). Participants in the laboratory experiment, in contrast, showed that same pattern for liking ratings as for attribute ratings: a main effect of attribute and an interaction. It might be that laboratory participants' liking ratings were informed by the conditioned attribute because athleticism is considered positive. Our inferences, however, are unaffected by these differences, as the analysis of attribute ratings that control for liking precludes that attribute ratings were informed by conditioned liking.

Conclusions. Relational qualifiers moderate attribute conditioning: participants indeed rated target persons disliking an athletic person as less athletic (and vice versa). Theoretically, the present experiments partially explain this effect because people seem to use liking as an indicator of similarity regarding attributes between persons. Practically, this implies that if someone dislikes Serena Williams or Dirk Nowitzki, the person will be perceived as unathletic despite the fact that the person might dislike them for reasons that have nothing to do with athleticism.

Chapter 5: Similarity-based and rule-based generalization in the acquisition of attitudes via evaluative conditioning

Abstract

Generalization in learning means that learning with one particular stimulus influences responding to other novel stimuli. Such generalization effects have largely been overlooked within research on attitude acquisition via evaluative conditioning (i.e. EC effects). In five experiments, we investigated whether and when generalization of EC effects is based on similarity or on abstract rules. Experiments 5.1, 5.2a, 5.2b and 5.3 showed that participants who abstracted a rule during the learning phase used that rule for category judgments of novel stimuli. However, evaluative ratings of the same stimuli were unaffected by the learned rule but followed the similarity to learned stimuli. Experiment 5.4 showed that this similarity-based pattern of generalization is not specific to evaluative ratings. Rather, resemblance between judgment task and learning task seems to determine whether acquired rules are taken into account. We discuss how dual-process and single-process models of EC may account for the obtained generalization results.

Similarity-based and rule-based generalization in the acquisition of attitudes via evaluative conditioning

Attitudes are a central construct within psychological research. From simple food choices (e.g. a salad vs. a hotdog) to complex voting decisions (e.g. Republicans vs. Democrats), attitudes play an important role in psychological theorizing (e.g. Ajzen, 1991; see Vogel & Wänke, 2016, for an overview). Evaluative conditioning (EC) is one paradigm that studies attitude acquisition and change in these domains. Functionally, EC is people's change in evaluative responses towards previously neutral stimuli (CSs) after these stimuli co-occurred with a positive or negative stimulus (US; see De Houwer, 2007, for a theoretical overview). For example, people will evaluate a neutral face that co-occurred with a happy face more positively than a neutral face that co-occurred with an angry face. Such EC effects are found in many domains, including person perception (Hütter et al., 2012), advertising (Sweldens, Van Osselaer, & Janiszewski, 2010), or eating behavior (Baeyens, Vansteenwegen, De Houwer, & Crombez, 1996).

Here, we investigate how EC effects generalize (see Pearce, 1987; Spence, 1937, for the concept of generalization); that is, how does the changed evaluation of one stimulus (i.e. an EC effect) influence other, novel stimuli? Given the claimed far-reaching explanatory potential of EC, investigating how evaluations of individual persons translate into evaluations of groups or how preferences for a certain product become preferences for a whole brand seems a worthwhile topic (Hütter, Kutzner, & Fiedler, 2014). In the remainder, we introduce the distinction between similarity- based and rule-based generalization, review relevant research and proceed to explain our empirical approach.

Similarity-based and rule-based generalization. Generalizing from previous instances of learning is crucial to be able to behave adequately in novel situations. Generalization is often studied by testing stimuli that vary in their degree of similarity to the learned CS on a certain stimulus dimension (e.g. size). Experiments typically show that conditioned responses are stronger when generalization stimuli are similar to the CS (e.g. Pearce, 1987). Such similarity-based generalization is manifest in responses towards novel stimuli that follow a similarity gradient and are unaffected by more complex propositions. However, responses towards novel stimuli may also be based on inference rules. Such rule-based generalization is manifest in responses that follow rules that can be instructed or learned by experience and require propositional knowledge.

This distinction is not new in the literature. Boddez, Bennett, van Esch, and Beckers (2017), for example, showed both rule-based generalization and similarity-based generalization in fear conditioning. They tested novel stimuli varying in the shade of grey to the CS that was previously followed by a shock. In one condition they observed the typical pattern: A generalization gradient of responses that are maximal for the CS and decrease as similarity to the CS decreases. However, when participants were instructed that the more similar stimuli are to the CS the less likely they will be followed by a shock (i.e. reverse similarity relation), they observed a reversed generalization gradient (see also Wong & Lovibond, 2017). That is, in this condition, responses to the novel stimuli were based on the instructed rule. Interestingly, valence ratings of the generalization stimuli all followed the standard gradient shape. Thus, the instructed proposition was applied for shock expectancy ratings but not for judgments of liking of the generalization stimuli; in our terms, generalization of liking took place in a similarity-based manner.

The present research. We adapted a paradigm from category learning (Shanks & Darby, 1998) to study the distinction between similarity-based and rule-based generalization in EC. While the original experiments only studied category learning, our study additionally included the typical elements of EC, pairings with positive or negative stimuli and a measure of liking: In a learning phase, participants saw pairings of individual stimuli and compounds of those stimuli (i.e. two individual stimuli were combined to form a compound) with positive or negative pictures. Crucially, the pairings followed a rule: Compound stimuli were paired with the opposite valence than their individual component stimuli. That is, if individual stimuli A and B were both paired with positive pictures, their compound AB was paired with a negative picture. Conversely, if stimuli C and D were paired with negative pictures, their compound CD was paired with a positive picture. The learning phase also included individual stimuli without their respective compound (i.e. individual stimuli E and F paired with positive pictures and individual stimuli G and H with negative pictures). Thus, the learning phase was at the same time a category learning and an EC setup.

We classified participants as “rule learners” and “non-rule learners”, depending on whether they could verbalize the underlying rule. In a test phase, participants then responded towards individual and compound stimuli that they had encountered in the learning phase. Additionally, they responded to novel compound stimuli which had not appeared in the learning phase (e.g. EF), but whose individual component stimuli had (i.e. E and F).

In the test phase, we obtained two measures: We asked participants (a) to categorize stimuli into a positive or a negative category and (b) to evaluate them. We aimed to replicate the finding by Shanks and Darby (1998) that rule learners generalize the rule they abstracted in the learning phase to novel stimuli, namely that two positives form one negative and two negatives form one positive. Thus, rule learners should classify the compounds EF as negative and GH as positive. Non-rule learners, in contrast, should respond towards novel compound stimuli in a similarity-based manner; that is, based on their components' pairing with positive or negative stimuli. Thus, non-rule learners should classify the compounds EF as positive and GH as negative, because its elements were both paired with positive and negative pictures, respectively. We tested these predictions in five experiments.

Overview of the experiments. Experiment 5.1 tested generalization for categorization responses and evaluative ratings as a function of rule abstraction. Replicating Shanks and Darby (1998), categorization of generalization stimuli depended on participants' rule abstraction. Yet, liking of generalization stimuli did not depend on rule abstraction but followed similarity-based generalization. However, the number of rule learners was very low. Subsequent experiments, therefore successfully manipulated factors that increased the number of rule learners. Additionally, Experiments 5.2a and 5.2b explored the role of responses and feedback during the learning phase. Experiment 5.3 instructed the relevant rule. Regardless, however, Experiments 5.2a, 5.2b and 5.3 showed the same pattern as Experiment 5.1 – rule knowledge (if present) was used for categorizing but not for judgments of liking of novel stimuli. Experiment 5.4 then showed that this similarity-based generalization is not specific to liking but potentially a function of closeness of the generalization judgments to the learning task.

We report only significant effects and null-effects when they are theoretically relevant. We conducted an additional preliminary study that was similar to Experiment 5.1, but did not probe rule knowledge that we do not report here. The results are redundant with Experiment 5.1. We excluded participants from the data sets who indicated at the end of the experiment that they had not taken part seriously (Aust et al., 2013). Apart from this, we did not exclude any data and we report all manipulations and all measures in the studies. The data for all reported experiments is available at https://osf.io/x4cw9/?view_only=e2aa7d6c90fb4da1b3d19189cf71d1d5.

Experiment 5.1

Participants categorized stimuli as “mammals” (positive) or “reptiles” (negative). The learning phase paired target stimuli (CSs) with mammal or reptile pictures (USs). The task could be mastered by memorizing which stimuli co-occur with which pictures (non-rule learning) or by additionally abstracting the underlying pairing rule (rule learning). We investigated whether rule learners apply their rule knowledge when categorizing and evaluating novel compound stimuli.

Method.

Participants. Seventy-seven people (female: 63, male: 13, unspecified: 1; mean age: 22.83 years, excluding two participants who gave nonsensical age information) recruited on campus participated for course credit or a small monetary reward. Experimental sessions included up to six people. A sensitivity analysis with G*Power (Faul et al., 2007) showed that this sample size allows detecting at least medium-sized effects for both categorization and liking judgments (f = 0.23) with a power of .85 and α = .05 (within-between interaction in an ANOVA, correlation among repeated measures: r = 0.1, sphericity assumed, see Results section).[12]

Design. The design included 20 CSs: 16 individual stimuli (A to P) and four compound stimuli (AB, CD, EF, and GH). Eight individual CSs were paired with positive mammals. Participants should learn to categorize these CSs as mammals and like them. Eight individual CSs were paired with reptiles. Participants should learn to categorize these CSs as reptiles and

[12] As we were unsure about what correlation among repeated measures to assume, we tested values between r=0.1 and r=0.5 in steps of 0.1. These analyses showed that we could potentially detect effects as small as f=0.17 (Experiment 5.1, 5.2a, 5.2b, and 5.4) and f=0.14 (Experiment 5.3). In the text, we report only the most conservative estimates. Note also, that the correlation among repeated measures and the resulting sensitivity will most likely be higher for the categorization DV than for the evaluation DV because concerning the former, we collected five responses per stimulus per participant and only one per stimulus per participant for the latter.

dislike them. Please note that the conjunction "and" does not imply causality. The four compound stimuli were paired with an animal of the other category (and opposite valence) than their components; for example, if A and B were paired with positive mammals, AB was paired with a negative reptile.

Four additional compound stimuli IJ, KL, MN and OP did not appear during the learning phase but served as generalization stimuli in the test phase. Table 4 shows the full design. The underlying rule "a compound and its elements belong to opposite categories" makes IJ and MN negative reptiles (because I, J, M, N were paired with positive mammal stimuli; see Table 4). Conversely, the rule makes KL and OP positive mammals (because K, L, O, P were paired with negative reptile stimuli; see Table 4). The test phase measured participants' categorization (mammal vs. reptile) and evaluation (positive to negative) for all individual and compound stimuli in two separate blocks. We counterbalanced the order of categorization and evaluation blocks. Between the learning and the test phase, we asked participants to verbalize the rule underlying the stimulus pairings in the learning phase to classify them as rule learners or non-rule learners.

Table 4. Learning phase design of Experiment 5.1 and 5.3. Stimuli A-P were individual conditioned stimuli (CSs) that were either paired with a positive mammal category ("pos") or a negative reptile category ("neg"). The compounds AB-GH were paired with the opposite valence than their element stimuli. IJ-OP were not paired in the learning phase but served as generalization stimuli (GS). According to the rule underlying the other stimulus pairings, IJ and MN would belong to the negative, and KL and OP to the positive category.

Individual CS	A → pos B → pos	C → neg D → neg	E → pos F → pos	G → neg H → neg
Compound CS	AB → neg	CD → pos	EF → neg	GH → pos
Individual CS	I → pos J → pos	K → neg L → neg	M → pos N → pos	O → neg P → neg
Compound GS	**IJ?**	**KL?**	**MN?**	**OP?**

Material and procedure. Upon arrival, experimenters seated participants at individual cubicles with personal computers and started the experimental computer program (OpenSesame; Mathôt et al., 2012). The program randomly assigned 16 CSs to serve as stimuli A to P and thus whether it would be paired with a positive or negative US and whether it would also appear in a compound stimulus. We used geometric shapes with light yellow background as CSs (see Appendix B). Ten colored photos of clearly positive baby mammals and ten photos of clearly negative reptiles served as positive and negative USs.

After consenting to participate, participants read a cover story. Their task would be to classify cellular samples of mammals and reptiles. These samples may consist of one or two cells (i.e. individual and compound stimuli). If a sample would not appear together with the picture of a mammal or a reptile, they would need to classify the sample. They received feedback on these classification responses.

A learning trial started with a blank screen for 500 ms. Then, the CS appeared in the screen’s top half and after 1000 ms the US appeared below. CS and US were shown together for 1500 ms. Every tenth trial, a CS appeared alone and participants classified the shape as a reptile or a mammal cellular sample. Contingent upon their response, the program showed a feedback screen with “Correct!” or “False!” at the top, the CS below and the US at the bottom for 2500 ms. The 20 CS-US pairs (16 individual and four compound stimuli; see Table 4) were shown ten times each, resulting in 180 learning trials and 20 response trials. Within these constraints, stimulus presentation order was random.

Next, participants answered an open-ended question regarding any regularities they may have noticed: “Did you notice a difference between cellular samples of mammals and reptiles? If so, please describe the difference." Then they proceeded to the test phase that consisted of a categorization and an evaluation block.

A trial in the categorization block showed a blank screen for 500 ms. Then the program showed the question "reptile or mammal?" on the top, a CS below, and the answer keys denoting the mammal and reptile categories at the bottom of the screen. We counterbalanced the assignment of the keys ("a" and "l") to the answer categories. Participants categorized all 24 stimuli (20 CSs plus four generalization stimuli) five times resulting in 120 trials.

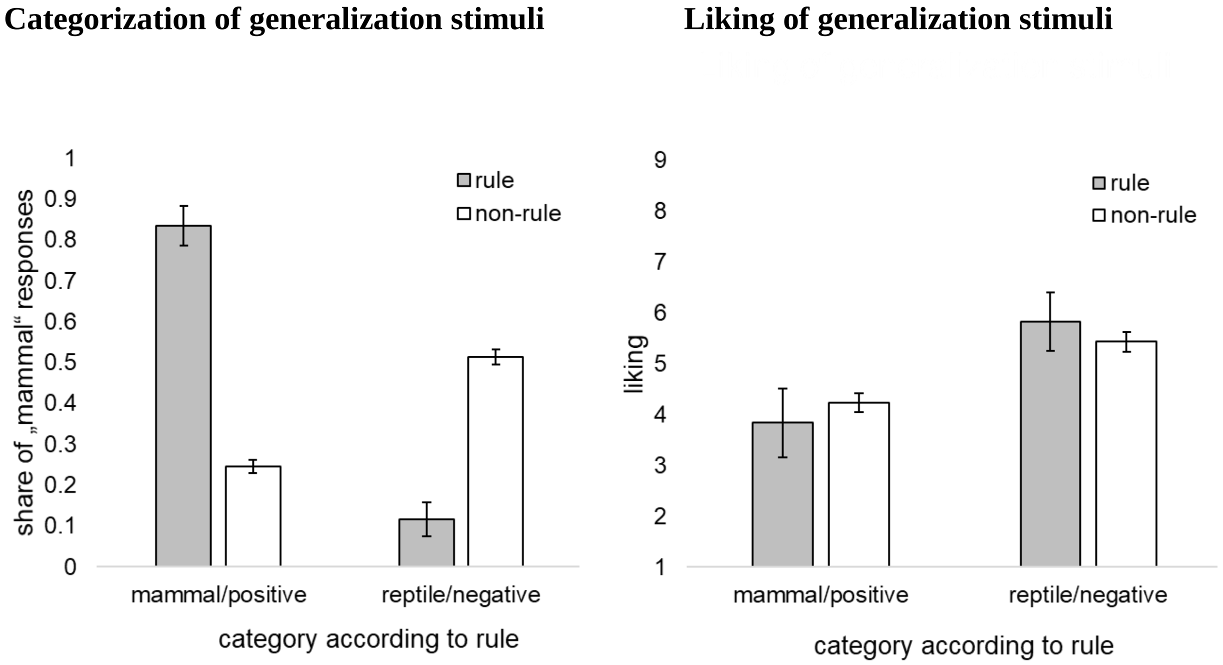
A trial in the evaluation block also showed a blank screen for 500 ms; yet, then the program asked participants to "evaluate the shape with the number keys" on the top of the screen and showed the CS and the scale (1, very negative, to 9, very positive) below. Participants evaluated all 24 stimuli once.

Stimulus presentation order within both blocks was random. Depending on the counterbalancing condition, participants completed the categorization or the evaluation block first. Upon completion of the test phase, experimenters thanked participants, informed them about the aim of the experiment and rewarded them.

Results. Based on their responses to the open-ended question, we classified six participants as "rule learners" and 71 participants as "non-rule learners". Appendix C details the classification procedure. The low number of rule learners is problematic but will be addressed in the following experiments which replicate Experiment 5.1's effect pattern. Participants responded above chance in the learning phase and rule learners descriptively performed better than non-rule learners (see also Appendix C).

Categorization. Figure 17's left panel shows the mean proportion of "mammal" classifications of generalization stimuli as a function of learner type (rule vs. non-rule learner) and the category according to the rule (mammal/positive vs. reptile/negative). The respective 2 (learner type) × 2 (category according to the rule) mixed ANOVA showed a category main effect contrary to the rule: Participants categorized rule-wise “mammal” compounds less often as mammals (*M* = 0.29, *SD* = 0.46) than rule-wise “reptile” compounds (*M* = 0.48, *SD* = 0.50), *F*(1,75) = 4.12, *p* = .046, η_p^2 = .05. Importantly, this effect was qualified by an interaction with learner type, *F*(1,75) = 19.78, *p* < .001, η_p^2 = .21. Separate ANOVAs for each learner type showed that non-rule learners classified novel compound stimuli contrary to the rule but in line with the pairings of their individual component stimuli. They categorized “mammal” compounds (i.e. comprised of two reptile-paired stimuli) less often as mammals (*M* = 0.25, *SD* = 0.43) than “reptile” compounds (i.e. comprised of two mammal-paired stimuli; *M* = 0.51, *SD* = 0.50), *F*(1,70) = 17.78, *p* < .001, η_p^2 = .20. Rule learners, on the other hand, classified compounds in line with the rule: They categorized “mammal” compounds as mammals more often (*M* = 0.83, *SD* = 0.38) than “reptile” compounds (*M* = 0.12, *SD* = 0.32), *F*(1,5) = 49.97, *p* < .001, η_p^2 = .91.

Evaluation. Figure 17’s right panel shows the mean evaluations of generalization stimuli. We used the same ANOVA as for the categorizations to analyze evaluations. Overall, participants evaluated compound generalization stimuli contrary to rule-based liking: Compounds consisting of two negatively paired stimuli and should –according to the rule – be evaluated positively, were evaluated more negatively (*M* = 4.20, *SD* = 2.16) than compounds consisting of two positively paired stimuli (*M* = 5.46, *SD* = 2.34), *F*(1,75) = 10.90 *p* = .001, η_p^2 = .13. Crucially, both rule and non-rule learners showed this pattern; the interaction between category according to the rule and learner type was not significant, *F*(1,75) = 0.69, *p* = .410, η_p^2 < .01.



Categorization of generalization stimuli

Liking of generalization stimuli

Figure 17. Data from Experiment 5.1: The left panel shows the proportion of ”mammal” categorization responses towards the four generalization stimuli, grouped by the correct category according to the rule as a function of learner type. Thus, high values for the mammal/positive category and low values for the reptile/negative category indicate good categorization performance. The right panel shows the evaluation of the generalization compound stimuli as a function of the correct category according to the rule and learner type. Error bars show the standard error of the mean.

Comparing categorization and evaluation. Above, we observed a significant interaction between correct category and learner type for categorization but not for evaluation. As the difference between significant and nonsignificant is not necessarily significant (Gelman & Stern, 2006), we z-standardized scores of both measures and submitted them to a 2 (learner type: rule vs. non-rule) x 2 (category according to the rule: mammal/positive vs. reptile/negative) x 2 (measure: categorization vs. evaluation) mixed ANOVA. If the interaction between category and learner type differ between the measures, we expect a three-way interaction.

We observed a two-way interaction of category and learner type ($F(1{,}75) = 6.08$, $p = .016$, $\eta_p^2 = .07$), a two-way interaction of category and measure ($F(1{,}75) = 27.18$, $p < .001$, $\eta_p^2 = .27$) and, most importantly, a three-way interaction of measure, category and learner type, $F(1{,}75) = 34.05$, $p < .001$, $\eta_p^2 = .31$. That is, categorization and evaluation significantly differed regarding the size of the interaction of category and learner type.

Discussion. Participants who could verbalize the underlying rule after the learning phase used this rule to categorize novel compound stimuli in the test phase. Participants who did not correctly verbalize the rule categorized novel compound stimuli based on their components' paired category. Experiment 5.1, thus, replicated the findings by Shanks and Darby (1998).

This pattern, however, did not emerge for the evaluation of novel compounds. Both rule and non-rule learners showed similarity-based generalization, basing their judgments on the components' paired valence. That is, even participants who understood that, according to the rule, a certain compound is a pleasant mammal stimulus (as attested by correct categorization), did not like it. Vice versa, even though they knew that a certain compound is supposed to be an aversive reptile stimulus, they nevertheless liked it. Importantly, this dissociation of measures did not occur for compounds that were paired in the learning phase. Appendix E reports detailed analyses of the responses for paired compounds and a comparison of paired and novel compounds. The conclusion is that in all experiments the observed pattern is specific for generalization.

While this is initially interesting, Experiment 5.1 does not allow strong conclusions due to the low number of rule learners. The open-ended question at the end of the learning phase was potentially not sensitive enough to detect rule knowledge. Furthermore, the diverging pattern between categorization and evaluation responses at test might follow because participants gave categorization but not liking responses during the learning phase. Potentially, participants' focus was on category, not evaluative information and therefore rule knowledge was only applied to categorization. We aimed to follow up on those shortcomings in Experiment 5.2a and 5.2b.

Experiment 5.2a and 5.2b

Foremost, Experiment 5.2a and 5.2b conceptually replicated Experiment 5.1 but aimed to increase the number of rule learners with two changes. First, we made the question used to classify participants as rule or non-rule learners more sensitive to rule knowledge. Second, we decreased the number of stimuli to facilitate rule abstraction. In addition, participants not only categorized CSs on every tenth trial but also evaluated them. In Experiment 5.2a, participants received feedback for their categorization responses like in Experiment 5.1. In Experiment 5.2b, they did not receive feedback for any response during the learning phase.[13]

Method.

Participants and design. Eighty-one people (female: 57, male: 24, mean age: 22.04 years) participated in Experiment 5.2a. Eighty people (female: 46, male: 33, unspecified: 1, mean age: 22.33 years) participated in Experiment 5.2b. Sensitivity was comparable to Experiment 5.1 (see Footnote 9). The design was highly similar to Experiment 5.1, but used only ten CSs: eight individual stimuli (A to H) and two compound stimuli (AB, CD). Four individual CSs and one compound CS were paired with positive USs and four individual CSs and one compound with negative USs.

Material and procedure. Experiment 5.2a and 5.2b's procedure followed Experiment 5.1 with two variations. In every tenth trial in the learning phase, the program asked participants to

[13] Experiment 5.2b was conducted last in the present series of experiments, as a result of the revision process of this paper.

categorize the CS. In Experiment 5.2a, participants received feedback for their response like in Experiment 5.1. In Experiment 5.2b, they did not receive feedback. In both experiments they were then also asked to evaluate the stimulus. They should judge whether the CS rather elicited a positive or negative feeling and to respond spontaneously. In Experiment 5.2b, the US appeared only after both responses had been given.

We changed the cover story to account for the liking judgments during the learning phase because “liking” of cellular samples is not meaningful. Instead, the cover story instructed participants that they had to learn an encrypted language with symbols that stand for mammals and reptiles.

As we used only 10 CSs, the learning phase contained 100 trials. The test phase included only 12 stimuli (ten CSs plus two novel generalization stimuli). After the learning phase, participants had the opportunity to verbalize the rule in two open-ended questions: “Did you detect a pattern in the symbolic language? Did you notice anything that you could use for the classification and the evaluation? If so, please describe the difference.” And after that: “Did you notice something that you could use for the classification and evaluation of the symbols that consisted of two figures? If so, please describe it.”.

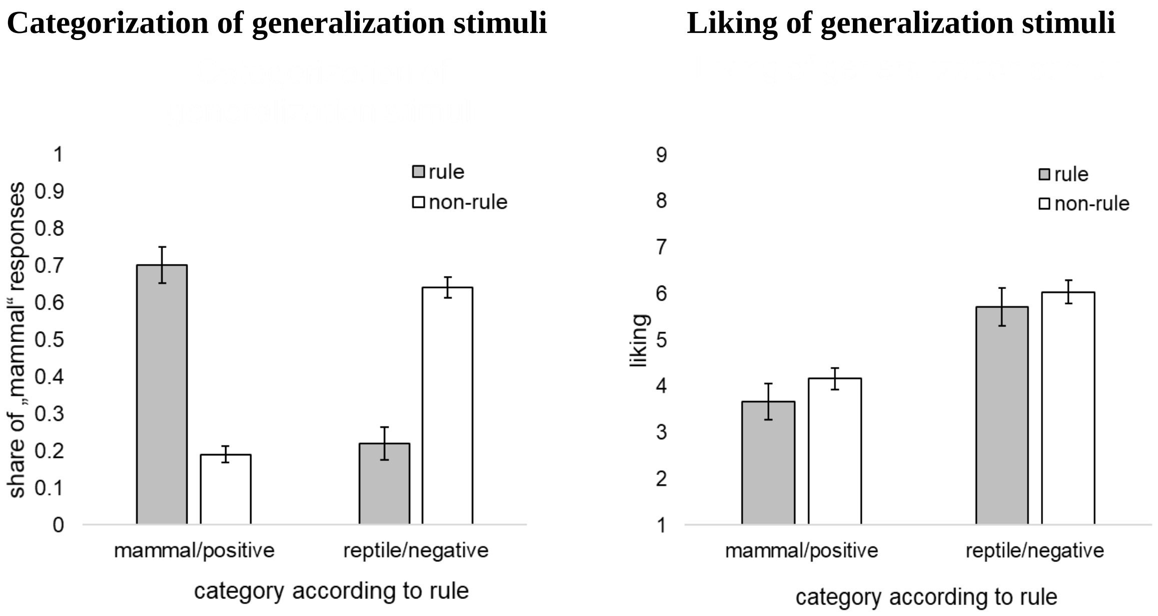
Then, participants proceeded to the test phase. The categorization block of the test phase consisted of 60 trials (twelve stimuli with five responses each) and the evaluation block consisted of twelve trials.

Results.

Experiment 5.2a. We classified 18 participants as rule learners and 63 as non-rule learners. Appendix C again reports the details. In the learning phase, participants responded above chance and rule learners descriptively performed better than non-rule learners (see Appendix C).

Categorization. Figure 18’s left panel shows the mean proportion of “mammal” classifications of generalization stimuli as a function of learner type and category according to the rule. As the pattern suggests, Experiment 5.2a replicated Experiment 5.1 for participants’ categorization responses. The respective ANOVA only showed an interaction between learner type and category according to the rule, $F(1,79) = 38.76$, $p < .001$, $\eta_p^2 = .33$. Follow-up ANOVAs showed that non-rule learners categorized “mammal” stimuli less often as mammals ($M = 0.19$, $SD = 0.40$) than “reptile” stimuli ($M = 0.63$, $SD = 0.48$), $F(1,62) = 45.45$, $p < .001$, $\eta_p^2 = .42$. That is, they based categorization of generalization stimuli on their elements’ categories. Rule learners, on the other hand, showed the opposite pattern. They categorized “mammal” symbols correctly as mammals more often ($M = 0.70$, $SD = 0.46$) than “reptile” symbols ($M = 0.22$, $SD = 0.42$), $F(1,17) = 9.48$, $p = .007$, $\eta_p^2 = .36$.

Evaluation. Figure 18’s right panel shows the mean evaluation of generalization stimuli as a function of learner type and category according to the rule. The respective ANOVA showed a category main effect contrary to the rule for evaluations. “Mammal“ stimuli were evaluated more negatively ($M = 4.05$, $SD = 1.80$) than “reptile” stimuli ($M = 5.96$, $SD = 1.93$), $F(1,79) = 25.63$, $p < .001$, $\eta_p^2 = .24$. That is, participants’ evaluations followed the components’ paired valence. There was no interaction with learner type, $F(1,79) = 0.06$, $p = .815$, $\eta_p^2 < .01$. Thus, even rule learners who categorized generalization stimuli based on the rule, did not use that rule knowledge to evaluate them.



Categorization of generalization stimuli

Liking of generalization stimuli

Figure 18. Data from Experiment 5.2a: The left panel shows the proportion of ”mammal” categorization responses towards the generalization stimuli, grouped by the correct category according to the rule as a function of learner type. Thus, high values for the mammal/positive category and low values for the reptile/negative category indicate good categorization performance. The right panel shows the evaluation of the generalization compound stimuli as a function of the correct category according to the rule and learner type. Error bars show the standard error of the mean.

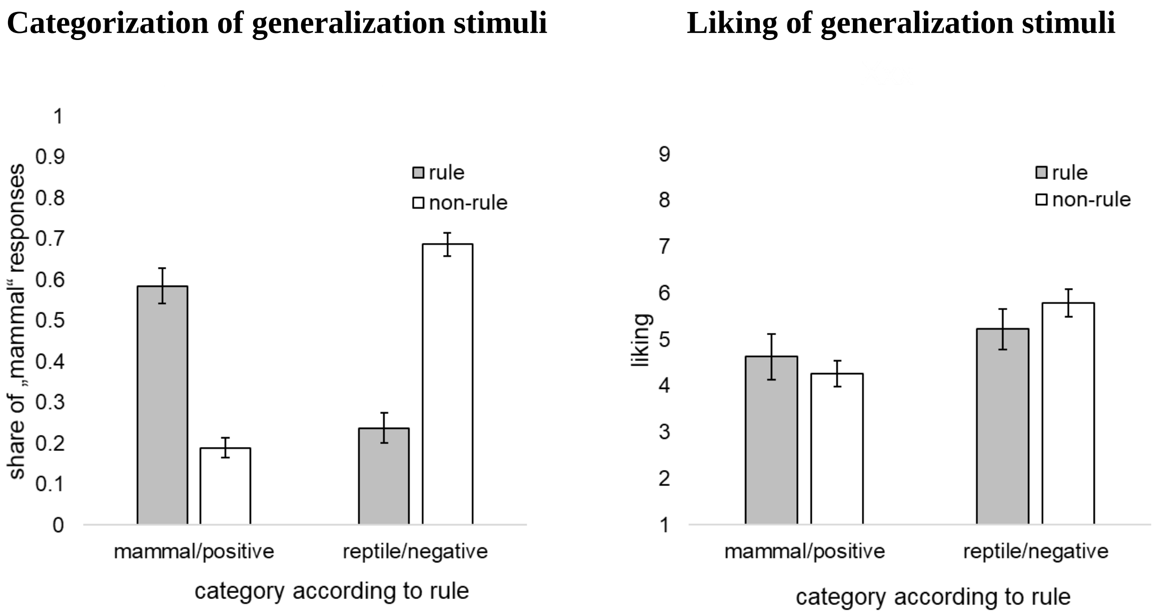
Comparing categorization and evaluation. We again tested whether the difference between measures was significant in a 2 (learner type: rule vs. non-rule) x 2 (category according to the rule: mammal/positive vs. reptile/negative) x 2 (measure: categorization vs. evaluation) mixed ANOVA with z-standardized scores as dependent variable and observed the predicted three-way interaction of category, learner type and measure, $F(1,79) = 34.77$, $p < .001$, $\eta_p^2 = .31$. Thus, the categorization and evaluation measure differed with regard to the interaction of correct category and learner type.

Experiment 5.2b. We classified 27 participants as rule-learners and 53 as non-rule learners. Participants responded above chance in the learning phase and rule learners performed better than non-rule learners. Appendix C again presents the details for the classification and the performance in the learning phase.

Categorization. Figure 19's left panel shows the mean proportion of "mammal" classifications and evaluations of generalization stimuli as a function of learner type and category according to the rule. The respective ANOVA showed only an interaction between correct category and learner type, $F(1,78) = 36.08$, $p < .001$, $\eta_p^2 = .32$. Follow-up ANOVAs showed that rule learners categorized novel compounds in line with the rule. "Mammal" compounds were classified as mammals more often ($M = 0.59$, $SD = 0.49$) than "reptile" compounds ($M = 0.24$, $SD = 0.43$), $F(1,26) = 6.54$, $p = .017$, $\eta_p^2 = .20$. Non-rule learners, in contrast, classified "mammal" compounds as mammals less often ($M = 0.19$, $SD = 0.39$) than "reptile" compounds ($M = 0.69$, $SD = 0.46$), $F(1,52) = 46.55$, $p < .001$, $\eta_p^2 = .47$.

Evaluation. Figure 19's right panel shows the mean evaluation of generalization stimuli as a function of learner type and category according to the rule. The respective ANOVA showed only a main effect of correct category which was contrary to what would be expected from rule-based generalization: "Mammal" compounds were evaluated more negatively ($M = 4.39$, $SD = 2.25$) than "reptile" compounds ($M = 5.60$, $SD = 2.21$), $F(1,78) = 5.82$, $p = .018$, $\eta_p^2 = .07$. No other effects were significant, all Fs < 1.14, all ps > .29, all η_p^2s < .02.

Comparing categorization and evaluation. As for Experiment 5.2a, the ANOVA with z-standardized scores of both measures showed a three-way interaction with measure, $F(1,78) = 17.20$, $p < .001$, $\eta_p^2 = .18$. Thus, categorization and evaluation differed with regard to the interaction of category and learner type.



Categorization of generalization stimuli

Liking of generalization stimuli

Figure 19. Data from Experiment 5.2b: The left panel shows the proportion of ”mammal” categorization responses towards the generalization stimuli, grouped by the correct category according to the rule as a function of learner type. Thus, high values for the mammal/positive category and low values for the reptile/negative category indicate good categorization performance. The right panel shows the evaluation of the generalization compound stimuli as a function of the correct category according to the rule and learner type. Error bars show the standard error of the mean.

Discussion. Experiment 5.2a and 5.2b replicated Experiment 5.1 with an increased number of rule learners. Rule learners categorized generalization stimuli according to the rule, but did not apply that rule for the evaluation of the same stimuli. Furthermore, the non-use of the rule for liking judgments cannot be explained by a sole focus on category information in the learning phase. While in Experiment 5.1, participants only categorized CSs, Experiment 5.2a included both categorization and liking responses into the learning phase and replicated Experiment 5.1’s findings. Yet, learning about categorization was more explicit in the learning phase because participants received feedback for categorization but not for liking judgments. Thus, learning about evaluative properties of the CSs took place in an incidental manner in Experiment 5.2a. In Experiment 5.2b, no feedback was given for categorization responses in the learning phase but we nevertheless replicated the pattern of rule-based generalization for categorization and similarity-based generalization for liking. Thus, explicit versus incidental learning can also not account for the findings.

Experiment 5.3

So far, the distinction between rule and non-rule learners was measured. Experiment 5.3 directly manipulated rule availability by instructing participants regarding the underlying rule. The setup was very similar to the previous experiments. Yet, participants in one condition explicitly received the rule underlying pairings in the learning phase.

Method.

Participants and design. One-hundred eighteen people (female: 77, male: 41, mean age: 22.48 years, excluding two participants who gave nonsensical age information) participated in Experiment 5.3. A sensitivity analysis showed that this sample allows detecting at least a small to medium-sized effect for both measures (f = 0.19) with a power of .85 and α = .05 (see previous experiments and Footnote 9).

The design was the same as in Experiment 5.1 using 20 CSs and the cover story of cellular samples (see Table 4). Additionally, we manipulated whether participants explicitly received the rule in the instructions or not.[14] To keep the experiments comparable, we still asked

[14] Due to a programming error, the conditions were not equal in size and they were not fully counterbalanced: We manipulated rule instruction and counterbalanced the order of DVs and assignment of response keys, resulting in eight counterbalancing conditions. We instructed the rule to five instead of four of those conditions. Thus, 74 participants received the rule, 44 did not. Also, among participants who did not receive the rule, we lacked a condition in which participants first underwent the evaluation block and then the categorization block of the generalization phase and responded with the key “a” for the reptile and “l” for the mammal category. That exact cell was twice as big among participants who received the rule. We consider the lack of full counterbalancing of minor importance as analyses from Experiment 5.1, 5.2a and 5.2b have not shown any relevant effects of the two variables we counterbalanced. Further, accidentally increasing the rule instruction condition lead to a higher number of rule learners in the sample. We consider this rather unproblematic for two reasons: First, rule learners are the subgroup of interest and more central to our arguments than non-rule learners. Second, non-rule learners constituted the vast majority of the samples in Experiments 5.1, 5.2a and 5.2b which gave ample opportunity to draw inferences about them.

participants to verbalize the rule. In the rule instruction condition they were asked to write down the rule directly after the instruction, while participants in the no rule instruction condition were asked to write it down after the learning phase.

Material and procedure. Stimuli and instructions were by and large the same as in Experiment 5.1. The rule was explicated in the following way in the rule instruction condition: “An experienced colleague gives you the following hint: When two cellular samples that consist of one cell (as depicted on the left) come from a mammal, then the cellular sample that contains both single cells together comes from a reptile. [example stimuli illustrating the rule were depicted]. The rule also applies the other way around: When two single cells come from a reptile, then the cellular sample that contains both single cells together comes from a mammal [example stimuli]. Do you understand the hint? If you have any questions, please contact the experimenters.” On the next page, participants were asked “To make sure that you internalized the colleague’s hint, please write down the aforementioned rule in the field below.”. That is, participants in the rule instruction condition, were asked to verbalize the rule directly after the instruction and therefore were not asked to verbalize it again after the learning phase. For participants in the no rule instruction condition, the procedure was the same as in Experiment 5.1 (i.e., rule verbalization after the learning phase).

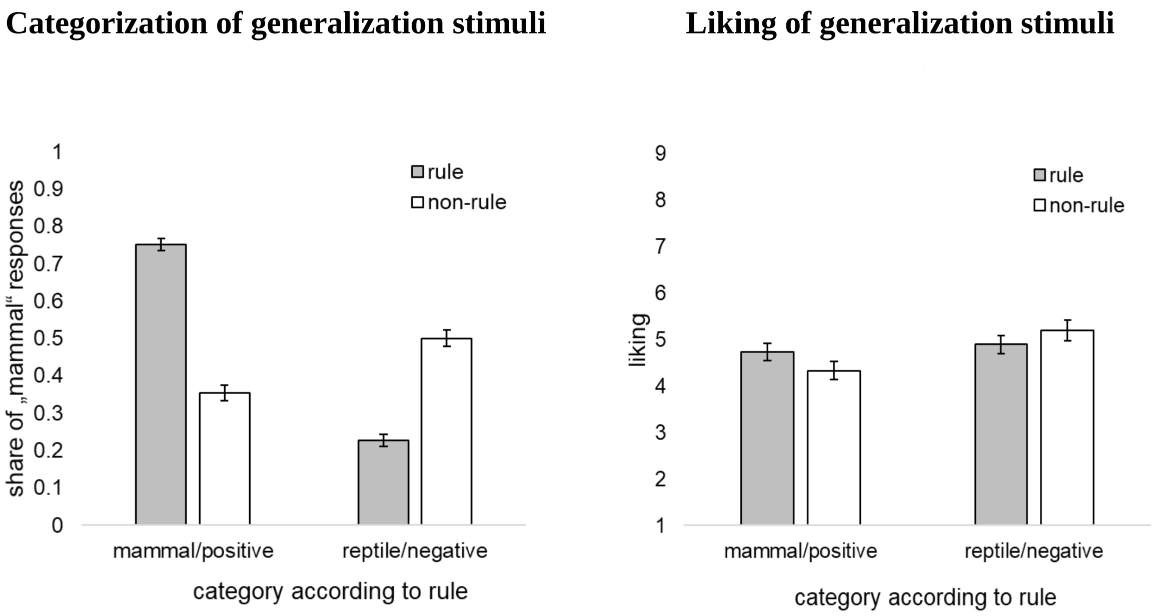
Results. We classified 68 participants as rule learners (62 in the rule instruction condition, 6 in the no rule instruction condition); 50 were classified as non-rule learners (12 in the rule instruction condition, 38 in the no rule instruction condition). Participants’ performance in the learning phase was above chance and rule learners descriptively outperformed non-rule learners (see Appendix C for details). For consistency, we report analyses with the measured factor “learner type” as opposed to the manipulated factor “instruction”. Appendix D presents the analyses with the manipulated factor "instruction". The conclusions are the same, as in the rule instruction condition 83.78% of participants were able to correctly verbalize the rule, while only 13.64% verbalized it in the no rule instruction condition.

Categorization. Figure 20's left panel again shows the categorization of generalization stimuli as a function of the learner type and category according to the rule. The respective ANOVA showed a main effect of category according to the rule. "Mammal" stimuli were more often classified as mammals (*M* = 0.58, *SD* = 0.49) than "reptile" stimuli (*M* = 0.34, *SD* = 0.47), *F*(1,116) = 18.05, *p* < .001, η_p^2 = .13. This effect was qualified by an interaction with learner type, *F*(1,116) = 56.78, *p* < .001, η_p^2 = .33. Non-rule learners based their categorization of the novel compounds on their elements' category: Participants classified "mammal" stimuli less often as mammals (*M* = 0.35, *SD* = 0.48) than "reptile" stimuli, (*M* = 0.50, *SD* = 0.50); a follow-up ANOVA showed that this difference was not significant on the standard alpha level, *F*(1,49) = 3.62, *p* = .063, η_p^2 = .07. Rule learners showed the reversed pattern. They categorized "mammal" stimuli more often as mammals (*M* = 0.75, *SD* = 0.43) than "reptile" stimuli (*M* = 0.23, *SD* = 0.42), *F*(1,67) = 104.56, *p* < .001, η_p^2 = .61. There was also a nonsignificant main effect of learner type, *F*(1,116) = 3.98, *p* = .051, η_p^2 = .03. Non-rule learners overall classified stimuli as mammals less often (*M* = 0.43, *SD* = 0.49) than rule learners (*M* = 0.49, *SD* = 0.50).

Evaluation. Figure 20's right panel shows the mean evaluation of generalization stimuli as a function of the learner type and category according to the rule. The respective ANOVA showed only a main effect of category for evaluations. Participants' evaluative responses were in line with the paired valence of the compounds' elements: Compounds consisting of two negative elements (and would, thus, rule-wise be positive) were evaluated more negatively (*M* = 4.56, *SD*

= 2.11) than compounds that consisted of two positive elements (M = 5.02, SD = 2.19), F(1,116) = 4.21, p = .042, η_p^2 = .04.

Comparing categorization and evaluation. The analysis with z-standardized scores of both measures showed the predicted three-way interaction, F(1,116) = 20.02, p < .001, η_p^2 = .15. Thus the two-way interaction of correct category and learner type, which we report for categorization, differed from the interaction in liking.



Categorization of generalization stimuli

Liking of generalization stimuli

Figure 20. Data from Experiment 5.3: The left panel shows the proportion of ”mammal” categorization responses towards the four generalization stimuli, grouped by the correct category according to the rule as a function of learner type. Thus, high values for the mammal/positive category and low values for the reptile/negative category indicate good categorization performance. The right panel shows the evaluation of the generalization compound stimuli as a function of the correct category according to the rule and learner type. Error bars show the standard error of the mean.

Discussion. Experiment 5.3 explicitly stated the underlying rule and thereby substantially increased the number of participants who could verbalize the rule. Still, Experiment 5.3 replicated Experiments 5.1, 5.2a and 5.2b and shows that liking judgments of novel stimuli were not sensitive to rule knowledge, although category membership judgments of the same stimuli were.

Factors that differed between the two dependent measures that might explain the use or non-use of an acquired rule at the test stage might, for example, be their dimensional (liking) versus categorical (categorization) nature. Experiment 5.4 investigated whether such extraneous factors influence the application of rule knowledge to evaluations.

Experiment 5.4

Experiment5. 4 investigated whether similarity-based generalization is specific to the difference between categorization and evaluation, or whether rule application may depend on other factors. For a strong test, Experiment 5.4 reversed the functional implementation of categorization and evaluation: We assessed evaluations categorically and categorizations by ratings; that is, participants categorized stimuli as positive or negative in the learning phase and in the test phase. Category information (i.e., mammal vs. reptile), which had not been relevant in the learning phase, was assessed via a rating measure in the test phase. Table 5 shows the generalization dimension (i.e., category vs. evaluative) and measurement type (i.e., categorical vs. ratings). This display illustrates that Experiments 5.1-3 may have confounded measurement with content dimension. Experiment 5.4 realizes the two missing cells in Table 5 and thereby de-confounds measurement and content. Additionally, we assessed liking via dimensional ratings.

Table 5. Overview of dependent measures in all experiments

		Type of measurement	
		Categorical	Rating
Type of information	Category information	Experiments **5.1,5.2a,5.2b,5.3**	Experiment **5.4**
	Evaluative information	Experiment **5.4**	Experiments **5.1,5.2a,5.2b,5.3,5.4**

Method.

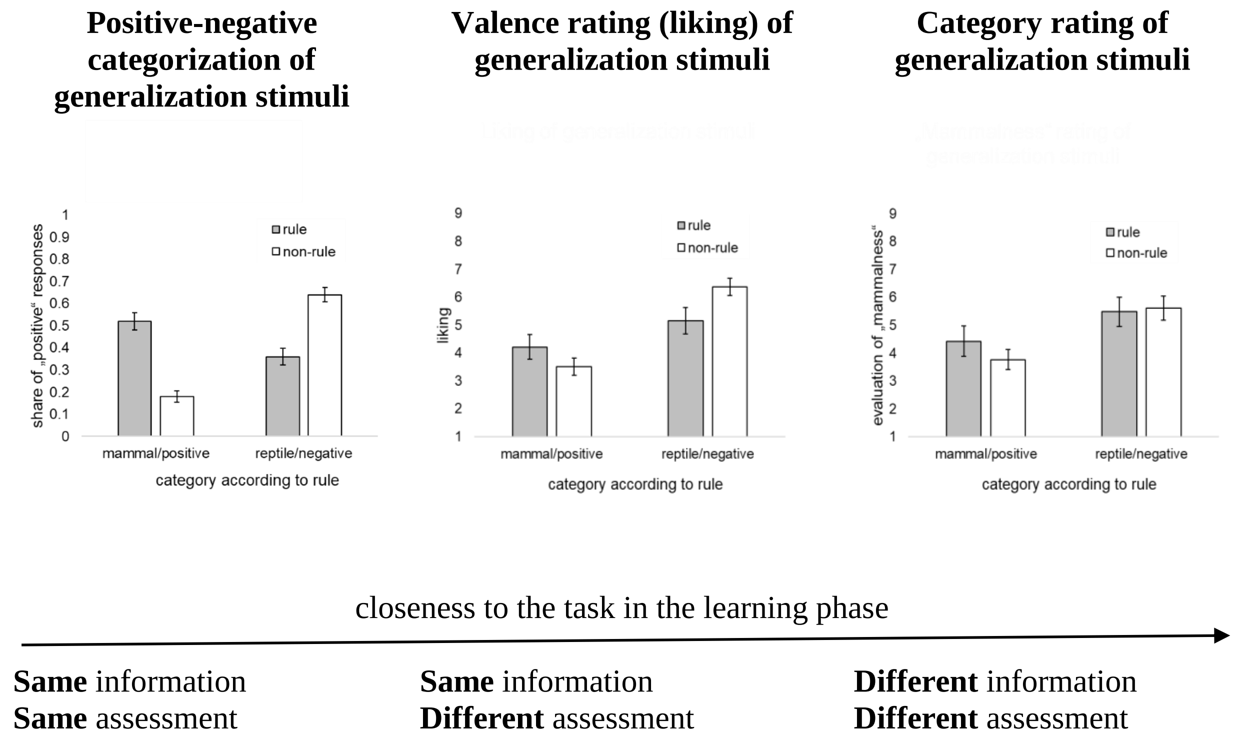
Participants and design. Seventy-nine people (female: 61, male: 18, mean age: 22.67 years) participated. The design used Experiment 5.2a's parameters and cover story. However, we administered as dependent measures a) a forced two-choice positive-negative categorization, b) a continuous "mammal – reptile" rating, subsequently referred to as category rating, and c) a continuous valence rating. We presented the measures in three separate blocks. We counterbalanced measurement order of a) and b); c) was always assessed last.

Material and procedure. Material and procedure were highly similar to Experiment 5.2a: We used the cover story about the encrypted language and used a reduced number of CSs in the learning phase. Crucially, participants categorized the CSs as positive or negative in the learning phase. Accordingly, the labels "mammals and reptiles" were replaced with "positive and negative". Thus, the USs remained the same as in previous experiments only their labels, and hence, the relevant categorization dimension, changed. Regarding the category ratings in the test phase, participants were asked to "[…] indicate whether you rather associate a symbol with a mammal or a reptile". The valence rating was assessed as in Experiment 5.1-3.

Results. We classified 33 participants as rule learners and 46 as non-rule learners. Participants' performance in the learning phase was above chance and rule learners performed descriptively better than non-rule learners (see Appendix C).

Positive-negative categorization. Figure 21's left panel shows the mean proportion of "positive" classifications generalization stimuli as a function of learner type and the category according to the rule. The respective ANOVA showed an overall effect of category according to the rule. Participants classified rule-wise "negative" stimuli more often as positive (M = 0.52, SD=0.50) than rule-wise "positive" stimuli (M = 0.32, SD = 0.47), F(1,77) = 4.54, p = .36, η_p^2 = .06. More relevant, an interaction with learner type showed that this main effect was mainly due to non-rule learners, F(1,77) = 19.14, p < .001, η_p^2 = .20. Follow-up ANOVAs showed that non-rule learners responded in the above described way ("negative": M = 0.64, SD = 0.48; "positive": M = 0.18, SD = 0.39), F(1,45) = 33.36, p < .001, η_p^2 = .43. Thus, non-rule learners responded towards novel compounds in line with their elements' category and contrary to the rule.

Rule learners, in contrast, descriptively responded in line with the rule. They classified "positive" stimuli as positive more often (M = 0.52, SD = 0.50) than "negative" stimuli (M = 0.36, SD = 0.48); this difference was not significant, though, F(1,32) = 1.62, p = .213, η_p^2 = .05.



Positive-negative categorization of generalization stimuli

Valence rating (liking) of generalization stimuli

Category rating of generalization stimuli

Figure 21. Data from Experiment 5.4: The left panel shows positive-negative categorization (i.e., proportion of “positive” responses) towards the generalization stimuli, grouped by the correct category according to the rule as a function of learner type. High values for the mammal/positive category and low values for the reptile/negative category indicate good categorization performance. The middle panel shows the liking ratings and the right panel shows ratings of category membership of the generalization compound stimuli as a function of the correct category according to the rule and learner type. The three dependent measures are arranged by their degree of closeness to the task during learning (similar to dissimilar from left to right). Error bars show the standard error of the mean.

Category rating. Figure 21’s right panel shows the category ratings of generalization stimuli. As these data show, participants’ rated generalization stimuli based on their elements’ paired category and thus contrary to the rule. Stimuli consisting of two mammal-paired elements (i.e., rule-wise “reptile” stimuli) were rated as more mammal-like (*M* = 5.56, *SD* = 2.98) than stimuli that consisted of two reptile-paired elements (i.e., rule-wise “mammal” stimuli; *M* = 4.04, *SD* = 2.76), *F*(1,77) = 7.50, *p* = .008, η_p^2 = .09. There was no interaction (*F*(1,77) = 0.55, *p* = .461, η_p^2 < .01) nor was there a main effect of learner type *F*(1,77) = 0.49, *p* = .488, η_p^2 < .01.

Valence rating. Figure 21's middle panel shows the mean valence ratings of generalization stimuli. Similar to the category ratings, we observed a main effect of paired valence. Participants based their evaluations of the novel compounds on the paired valence of their elements as opposed to the valence prompted by the rule. Participants rated compounds of two positive elements more positively (M = 5.86, SD = 2.44) than compounds of two negative elements (M = 3.80, SD = 2.29), F(1,77) = 18.58, p < .001, η_p^2 = .19. We also observed an interaction of paired valence with learner type. As Figure 21's middle panel shows, the effect of paired valence was more pronounced for non-rule learners. The liking difference between "positive" and "negative" compounds was Δpositive-negative = -2.87 compared to rule learners, Δpositive-negative = -0.94, F(1,77) = 4.77, p = .032, η_p^2 = .06. The learner type main effect was not significant, F(1,77) = 0.73, p = .396, η_p^2 < .01.

Comparison of measures. To test whether the interaction between type of learner and correct category differed between measures, we z-standardized all three measures and ran a 2 (learner type: rule vs. non-rule) x 2 (category according to the rule: mammal/positive vs. reptile/negative) x 3 (measure: positive-negative categorization vs. category rating vs. valence rating) mixed ANOVA. We observed the predicted three-way interaction of learner type, category and measure, F(1.95,150.08) = 5.57, p = .005, η_p^2 = .07, degrees of freedom were Greenhouse-Geisser corrected. That is, the two-way interaction of learner type and category differed between the three measures, although the exact nature of this difference is not clear due to the three levels.

To test between which measures the two-way interaction differs, we conducted three follow-up 2 x 2 x 2 ANOVAs. We observed the critical three-way interaction when comparing positive-negative categorization and category rating (F(1,77) = 9.86, p = .002, η_p^2 = .11) and when comparing positive-negative categorization and valence rating ($F(1,77) = 4.82$, $p = .031$, $\eta_p^2 = .06$) but not when comparing category rating and valence rating, $F(1,77) = 1.61$, $p = .209$, $\eta_p^2 = .20$.

Thus, the two-way interaction between correct category and learner type we observed for positive-negative categorization differs from both rating measures. Category rating and valence rating, however, do not differ significantly regarding the two-way interaction.

Discussion. Experiment 5.4 showed that similarity-based generalization is not specific to liking judgments and rule-based generalization is not specific for category judgments. When participants could verbalize the rule, they applied that rule when categorizing novel compounds into the categories “positive” and “negative”. Importantly, they did not generalize the rule to the same extent when rating their liking or the degree to which a compound belongs to a mammal versus a reptile category. Instead, rule learners’ response pattern especially for the latter measure showed that they based their judgments on the category the compounds’ elements were associated with in the learning phase. That is, as in Experiments 5.1, 5.2a, 5.2b and 5.3, we observed a use of the learned rule for one measure and neglect of the rule for another. Importantly, the information dimension of the measure (i.e., evaluative versus category) does not determine whether rule knowledge is applied or not. That is, the non-use of rules is not specific to measures of liking.

A possible plausible explanation might be that Experiment 5.4’s three measures can be conceptualized as ranging from very similar to dissimilar to the learning task. Figure 21 illustrates this similarity notion. The positive-negative categorization task was the same task as in the learning phase. The valence rating was only similar– it asked for the same information dimension (evaluative) but used a different measure (ratings as opposed to forced two-choice). The category rating was dissimilar – it asked for a different information dimension (category membership) and used a different measure.

The pattern suggests that the closer a generalization task is to the original learning task, the more likely propositional information acquired during learning will be applied in generalization: For positive-negative categorization we observed a clear interaction, indicating that rule learners showed rule-based generalization, whereas non-rule learners showed similarity-based generalization. For valence rating, we observed the pattern to be less pronounced. The interaction was smaller and both rule and non-rule learners showed similarity-based generalization (the former less clearly, though). For category ratings, we observed no interaction and thus clear similarity-based generalization for both types of learner.

General Discussion

Five experiments tested whether attitudes by EC generalize in a rule-based or a similarity-based manner. Similarity-based refers to responses towards novel stimuli that are only influenced by their similarity to learned stimuli. Rule-based generalization refers to responses that are influenced by acquired rules. Experiments 5.1, 5.2a, 5.2b and 5.3 showed that participants who acquired rule knowledge about stimulus relations (i.e., rule learners) applied this knowledge to categorical judgments of novel stimuli (i.e., rule-based generalization). Surprisingly, we observed liking judgments about those novel stimuli to be unaffected by learned rules. Rather, they were based on the feature similarity with the components of the novel stimuli (i.e., similarity-based generalization). However, Experiment 5.4 showed that similarity-based generalization is not specific to judgments of liking. Rather, the closeness of the generalization task to the learning task determined whether rule-based or similarity-based generalization occurred: The closer the generalization task was to the task during the learning phase, the more likely propositions acquired during learning were used in the generalization task. Further, our findings suggest that similarity-based generalization is the default mode of generalization which can, if certain conditions are met, be overridden by rule-based generalization. This notion is in line with the findings by Boddez and colleagues (2017) on similarity-based and rule-based generalization in fear conditioning.

The observed pattern of generalization might provide insights regarding the underlying learning processes. Two candidate processes in EC are associative and propositional learning. With regards to EC, associative learning means that spatiotemporal contiguity links stimuli (CS and US) in memory. Presenting one stimulus again (i.e., the CS), thereby activates the linked stimulus (i.e., the US), which influences evaluative responses. Propositional learning, on the other hand, means that statements about the relation of the co-occurring stimuli are stored in memory and retrieved upon presentation of one stimulus. These types of learning are assumed to be governed by different operating principles (activation of links versus validation of propositions; Gawronski & Bodenhausen, 2011). A purely associative perspective cannot account for the finding that category judgments of novel stimuli are sensitive to rule knowledge. Either, one has to assume an additional, propositional learning process that informs categorization and leads to rule-based generalization while liking is informed by the associative process that leads to similarity-based generalization. This conception would be in line with a dual-process perspective on the findings (Gawronski & Bodenhausen, 2006) but such an account has difficulties explaining the findings of Experiment 5.4 which showed that similarity-based generalization is not specific to liking. Furthermore, studies by Zanon, De Houwer and Gast (2012) suggest that rule-based generalization can emerge for implicit measures of liking. Although they did not assess whether participants abstracted the relevant rule in a very similar paradigm, they observed responses towards novel stimuli on three different implicit measures to be affected by the rule that governed pairings in the learning phase. Alternatively, one abandons the associative perspective and explains the findings in terms of a propositional or memory-based learning process (De Houwer, 2018; Gast, 2018; Stahl & Aust, 2018). Such a view would locate the origin of the dissociating measures not at the level of learning process but at the judgment stage: participants hold different response strategies for different measures (cf. Aust, Haaf, & Stahl, in press).

However, this analysis of theoretical accounts makes clear that, in their current state, theories of learning do not allow for clear predictions regarding generalization of learning in evaluative conditioning. An extension of scope is called for to account not only for responses towards paired but also towards novel stimuli.

Regarding real-world implications, our findings suggest that rule knowledge acquired in a certain situation will not readily generalize to novel situations. This might be highly relevant, for example, for interventions be they clinical (e.g., cognitive treatment of phobias) or societal (e.g., measures to reduce stereotypes towards ethnic groups). Our data suggest that one measure to facilitate generalization of rule knowledge could be to make the learning situation as close as possible to real-life situations.

Limitations and conclusion. The relevant rule that participants had to abstract in order to show rule-based generalization was rather abstract and specific to the paradigm. Therefore, participants might view the rule as very narrowly applicable. If the rule was more in line with real-life propositions, its application might be less restricted. A more intuitive rule, would, for example be that two tasty foods mixed together can be awful, like pizza and pineapples. It is an interesting question, whether such a rule would be more widely applicable and would thus, for example, also be applied to liking judgments.

A promising future avenue could be to link generalization to operating conditions of learning processes. One could design a learning phase (or a test phase) which prompts, for example, automatic versus controlled processing and test whether similarity-based or rule-based generalization emerges. In fact, our finding that rule learners (descriptively) outperform non-rule learners in the learning phase speaks to the idea that different levels of attention or controlled processing during learning might lead to different pattern of generalization. Also, the findings of Experiment 5.2a and 5.2b suggest that intentionality is not a necessary condition for rule-based generalization to emerge. Thus, generalization might serve as a domain to gain insight about the processes operating at the learning stage. The present introduction of rule-based and similarity-based generalization may serve as an empirical framework for testing predictions from prominent single- versus dual-process models of evaluative learning and attitude acquisition.

Chapter 6: Discussion

In my dissertation, I aimed to investigate the mental processes underlying evaluative conditioning (EC). While the majority of learning effects is assumed to be mediated by elaborate mental processes, EC has repeatedly been identified as an effect that might be (partly) mediated by more primitive processes. Therefore, studying the processes underlying EC is essential to our understanding of learning: Is there only one way in which we learn (single-process theory) or are there qualitatively different ways of learning (dual-process theory)?

While there is general agreement nowadays, that elaborate, “propositional” processes contribute to EC, research focuses on the question whether the assumption of a more primitive, “associative” process also contributing to EC, is justified. This associative process can be characterized in terms of operating conditions that refer to when the process operates and operating conditions that specify how a process operates.

A central operating condition of associative processes, as identified by theories on automaticity, is awareness of the pairings of CS and US during conditioning. It was extensively studied and thus strongly shaped the debate about learning processes. Chapter 3 reports a series of studies on the role of awareness in EC: My colleagues and I manipulated awareness via a technique referred to as Continuous Flash Suppression (CFS). We presented CSs as stationary black and white images to one eye while a flash of colored pixel masks and US photos was shown to the other eye. This simultaneous conflicting input made the eyes compete for visual awareness, that is, only input from one eye could be consciously perceived. Since the CSs were much less visually informative, they were suppressed from awareness and participants performed worse at reporting them than they did for control stimuli. Across four experiments we observed awareness of the CS-US pairings to be a necessary condition for EC effects to emerge. This observation is in line with a recent comprehensive investigation of subliminal EC (Stahl et al., 2016) and other approaches to demonstrate unaware EC like parafoveal CS presentation (Dedonder et al., 2014). This shows that the processes that operate during EC depend on awareness.

Regarding operating principles of associative processes, I studied a central principle that was put forward by (at least) two influential theories of EC, the referential account by Baeyens and colleagues (1992) and the APE model by Gawronski and Bodenhausen (2006). It states that associative processes do not incorporate propositional information about, for example, the relation between CS and US. My colleagues and I tested this prediction in an attribute conditioning (AC) paradigm which is similar to EC. In AC, CSs are paired with USs that have a certain attribute, for example athleticism. After CS-US pairings, the CS will also be ascribed that attribute. We introduced a positive or a negative relation between CS and US in the learning phase, that is, they liked or disliked each other. We observed in four experiments that AC effects were sensitive to those relations: A CS that liked an athletic US was rated as more athletic than a CS that liked an unathletic US. This effect was reversed, however, when CS and US disliked each other. This finding is in line with evidence from the EC paradigm that shows EC's sensitivity to relations between CS and US (e.g. Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2012; Moran & Bar-Anan, 2013; Unkelbach & Fiedler, 2016).

Importantly, however, there are studies that show different patterns of findings on direct and indirect measures of evaluation when relational information is introduced: Most prominently, Moran and Bar-Anan (2013) manipulated whether CSs started (i.e. positive relation) or ended (i.e. negative relation) US sounds. They observed that EC effects on a direct measure were sensitive to this relation but an indirect measure (IAT) was not. This and other studies that show similar dissociations between direct and indirect measures (e.g., Gawronski, Balas, & Creighton, 2014; Hu, Gawronski, & Balas, 2017a, 2017b) have been explained with the existence of an associative process in EC that informs the indirect measures. Thus, these findings have been invoked as support for dual-process models (especially the APE model).

The empirical studies reported in Chapter 5, however, show that dissociations between measures can be explained by other factors than two distinct learning processes: My colleagues and I studied similarity-based and rule-based generalization in EC. Similarity-based generalization means that responses to novel stimuli are based only on their feature similarity to learned CSs. Rule-based generalization means that responses to novel stimuli are based on rule knowledge. We observed a dissociation between a categorization measure and a liking measure in the first four experiments: Categorization of novel stimuli showed both types of generalization depending on whether participants abstracted the rule necessary for rule-based generalization. Liking, in contrast, showed only similarity-based generalization. Importantly, however, in the final experiment, we showed that it is plausible that this dissociation emerges because the two different measures encourage different response strategies. Specifically, we concluded that the closer a judgment task is to the task during learning, the more likely rules acquired during learning will be applied. Thus, the duality does not emerge at the level of learning processes but rather at the measurement level. Information provided by one single learning process can be used flexibly and can thus produce different pattern of results depending on what measures one uses. This provides an alternative explanation for the findings put forward to support the operating principles of associative processes suggested by the APE model.

Bading, Stahl, and Rothermund (in press) provided a direct test of whether response strategies can account for use and non-use of propositional information on direct and indirect measures, respectively. Specifically, they showed that the dissociation Moran and Bar-Anan observed in their influential paper from 2013, can be explained by specific response strategies for the indirect measure they used (IAT). When they altered the IAT to encourage different response strategies, it showed an effect of relational qualifiers and, thus, converged with the direct rating measure.

Relatedly, Aust, Haaf, and Stahl (in press) studied dissociations between an US expectancy rating and a liking rating after extinction: While US expectancy ratings are reduced after CS alone trials, liking ratings were unaffected (e.g., Hermans, Crombez, Vansteenwegen, Baeyens, & Eelen, 2002). Aust and colleagues argued that expectancy ratings are by default momentary judgments, drawing on recent learning trials. Hence, they reflect recent CS alone trials while liking ratings – by default integrative judgments that summarize many more learning trials – do not reflect them (to the same extent). They showed that the dissociation can be reversed (i.e. liking ratings were sensitive to extinction but US expectancy was not) when judgment strategies reversed (i.e. momentary liking rating, integrative expectancy rating). These findings are important in two ways: First, together with the studies by Bading and colleagues (in press), they provide an alternative, single-process explanation for the dissociation between two measures like the studies reported by Moran and Bar-Anan (2013). And second, they show that the studies reporting no effect of extinction in EC are not necessarily evidence for a different learning process underlying EC but are readily explained by different judgment strategy of liking ratings (see also Lipp, Oughton, & LeLievre, 2003; Lipp & Purkis, 2006). That is, the central original argument why EC should be mediated by different processes than other forms of learning loses its appeal.

Thus, taken together, my empirical studies do not provide evidence that make it necessary to assume that an automatic, associative learning process is involved in EC. Rather, the findings, that awareness is necessary for EC to emerge, that AC is sensitive to relational information, and that generalization of EC can be sensitive to rule knowledge are adequately explained by a single process underlying EC. In that process, all information provided in the learning phase is registered to the extent that capacity is available and to the extent that it is perceived as relevant for the task, and is stores in memory. My empirical work does not allow to differentiate and I, therefore, do not commit to a specific idea of memory representation as propositions or a distributed associative network. At the judgment stage, all encoded information is retrieved (unless it was forgotten) to the extent that capacity is available. Depending on the judgment task, different pieces of information from the learning phase might be taken into account depending on what is considered relevant. Figure 22 visualizes this idea and locates the empirical findings of Chapter 3, 4 and 5 within this framework.

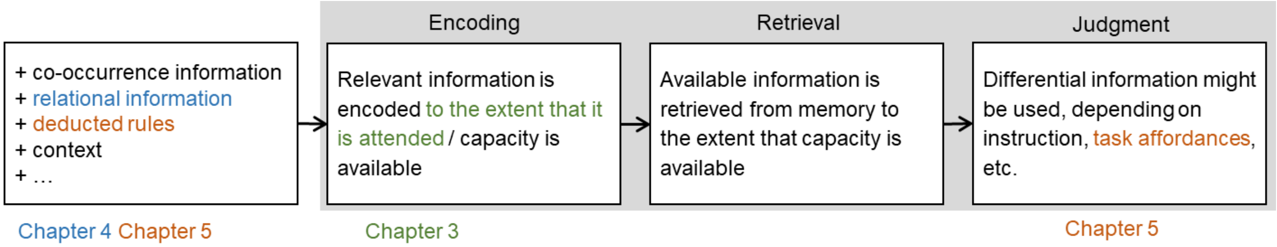

+ co-occurrence information
+ relational information
+ deducted rules
+ context
+ …
Chapter 4 Chapter 5
Encoding
Relevant information is encoded to the extent that it is attended / capacity is available
Chapter 3
Retrieval
Available information is retrieved from memory to the extent that capacity is available
Judgment
Differential information might be used, depending on instruction, task affordances, etc.
Chapter 5

Figure 22. Joint explanatory framework for the empirical findings of this thesis.

The box on the very left contains information that might be present in a learning phase. This is co-occurrence information of CS and US and potentially their relation, instructed or deducted rules, context stimuli and others. The subsequent boxes could be described as “hurdles” that can prevent certain pieces of information to influence evaluative judgment. First, a piece of information might not be registered by the learning process, for example because it was rendered invisible as in Chapter 3. Alternatively, it might not be considered relevant for the learning task and hence not be encoded. Then, it might not be retrieved at the judgment stage, for example because it was forgotten or because of capacity limitations. Even if it was encoded and retrieved it might not be taken into account for evaluative judgements, for example, because of task affordances. Chapter 4 and 5 show instances in which additional information to CS-US co-occurrence (Chapter 4: relational information, Chapter 5: deducted rules) is encoded and retrieved and is taken into account (Chapter 4, Chapter 5, Experiment 5.4) or not (Chapter 5, Experiment 5.1,5.2 and 5.3).[15]

This explanation of my empirical findings is akin to a declarative memory model of EC, recently proposed by Gast (2018) and it is also in line with a memory-based judgment account of EC as proposed by Stahl and Aust (2018) and the propositional account of EC (De Houwer, 2009, 2018). A central point of my explanation is that the information influencing evaluative judgments can vary in detail or quality because of the preceding “hurdles”. For example, time constraints in the measurement phase can lead to an imperfect retrieval of information. Hence, in an indirect measure which measures speeded responses, only information about CS-US co-occurrence but not about their specific relation might be retrieved. An observed insensitivity of indirect measures to relational information might, thus, look like the results of associative processes although it is the results of a memory-based process with reduced quality of retrieval. The point is that instead of invoking the dichotomy of associative and propositional, I propose a

[15] The figure is not exhaustive regarding the possible stages and processes in EC. It includes only concepts needed to explain the empirical findings of my thesis. Other conceptualizations have included a retention stage, for example (Gast, 2018).

continuum of varying quality. This is akin to the re-conceptualization of the automatic-controlled dichotomy as graded quality of representation by Moors (2016). Considering these aspects, I believe a single-process framework can go a long way in explaining findings in EC.

Limitations and open questions

My thesis mainly concerns one operating condition and one operating principle of associative processes in EC. As outlined in the introduction, however, many more conditions and principles have been put forward. Concerning operating conditions, some studies aimed to investigate whether EC effects can be observed under conditions of low attention, unintentionality and whether they can be uncontrolled. Studies aiming to show EC when participants' attentional resources are taxed, for example by a secondary task, have been reviewed in Chapter 3. In sum, many of those studies face methodological problems that arise from a between-participants manipulation of depletion and therefore, conclusive evidence for EC under depletion conditions is still pending. Another line of research, however, has repeatedly shown that EC can emerge under incidental learning conditions. This effect pertains both to the operating conditions of efficiency and unintentionality. Olson & Fazio (2001) developed the surveillance paradigm, in which participants view a stream of different stimuli and are asked to respond upon the presentation of a particular target item. Interspersed in this stream of stimuli are CS-US pairs. Although participants are instructed that these are distractors, EC effects could repeatedly be demonstrated in this incidental paradigm (e.g. Jones, Olson, & Fazio, 2010; Olson & Fazio, 2001, 2002; Stahl & Heycke, 2016). Thus, there is evidence that EC can be obtained even when probably little attention is devoted to CS-US pairings and when participants do not intend to learn about the evaluative characteristics of the CS. Incidental EC is not fundamentally at odds with the explanation in Figure 22. The additional load from the unrelated stimuli and the unrelated task should reduce the amount of CS-US pairs that are encoded or should reduce the quality of encoding. Accordingly, incidental EC effects should be smaller than those under intentional learning conditions which is in line with findings by Stahl and Heycke (2016).

Finally, Hütter and Sweldens (2018) addressed the last operating condition of associative processes. They used the rationale of the process dissociation procedure described in Chapter 3 to disentangle controllable and uncontrollable learning processes in EC. They instructed participants to either use (inclusion condition) or reverse (exclusion condition) the affective information of the US to form an impression of the CS. They showed that participants partly did not control (i.e. reverse) their evaluations of the CS. However, Stahl and Aust (2018) explained participants' failure to reverse some of the USs' valence in terms of capacity constraints at the learning stage which is compatible with the framework in Figure 22.

Reviewing evidence for and against all operating principles that were not empirically targeted in this thesis would go beyond the scope of this section and has comprehensively been done elsewhere (e.g. Corneille & Stahl, 2018; De Houwer et al., 2001; Hofmann et al., 2010). It can be summarized as follows: The principles suggested by older single-process associative theories (e.g., holistic and referential account) are inconsistent with a large body of robust effects in the EC paradigm, for example, its sensitivity to propositional information discussed in Chapter 4. The conceptual categorization account has been criticized among others by Baeyens and colleagues (1998) because it can hardly account for EC effects across modalities (Hofmann et al., 2010) and US revaluation effects (Walther et al., 2009). The implicit misattribution principle receives support from some studies showing S-R learning in EC (Baeyens, Vanhouche, Crombez, & Eelen, 1998; Gast & Rothermund, 2011a). Corneille and Stahl (2018), however, argued that S-R learning does not necessarily have to be an associative process. The "response", that is, the USs' valence, can be conceived of as information that is registered by a propositional or memory process and later influences liking judgments while the exact identity of the US might have been forgotten (cf. Stahl et al., 2009). Again, this conceptualization of S-R learning is compatible with the explanation in Figure 22.

Finally, the central limitation in all studies reported here is that it is logically impossible to prove the inexistence of an automatic, associative process. We aimed to address this issue in Chapter 3 by identifying optimal conditions for EC without awareness to occur, but nevertheless failed to show unaware EC. In Chapter 4 we studied AC as a variant of EC in which the application of a like/dislike relation is normatively questionable and hence a priori less likely (i.e., someone who dislikes an athletic person should not necessarily be unathletic). Still, we observed AC's sensitivity to this relation. The first four experiments reported in Chapter 5 showed similarity-based generalization of liking which is consistent with an associative view on EC. However, the final experiment showed that this pattern is adequately accounted for by judgment strategies at the measurement phase. Thus, while we cannot ultimately rule out the existence of an associative process, the present thesis shows that it is not necessary to assume such a process. "There is very little to be lost, and much to be gained, by the rejection of the dual-system approach" (Mitchell, Houwer, & Lovibond, 2009, p.185).

Conclusion

Summing up, while they might behave differently on a functional level, there is little compelling evidence that EC is mediated by primitive mechanisms of learning that are qualitatively different from the elaborate mechanisms underlying classical conditioning. Therefore, EC research, at the current state, is not a strong argument for the existence of a second route to learning in the broader debate on single versus dual processes in learning. Rather, the vast majority of effects in EC research are adequately accounted for by a propositional single-process account of EC and thus speak for a single process account of learning more broadly that is characterized by rather deliberate processing and reasoning.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alves, H., Koch, A., & Unkelbach, C. (2016). My friends are all alike - The relation between liking and perceived similarity in person perception. *Journal of Experimental Social Psychology*, *62*, 103–117. https://doi.org/10.1016/j.jesp.2015.10.011

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*, 527–535. https://doi.org/10.3758/s13428-012-0265-2

Aust, F., Haaf, J. M., & Stahl, C. (2019). A memory-based judgment account of expectancy-liking dissociations in evaluative conditioning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 417–439. https://doi.org/10.1037/xlm0000600

Bading, K., Stahl, C., & Rothermund, K. (in press). Why a standard IAT effect cannot provide evidence for association formation: The role of similarity construction. *Cognition and Emotion*. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/02699931.2019.1604322

Baeyens, F., Crombez, G., Hendrickx, H., & Eelen, P. (1995). Parameters of human evaluative flavor-flavor conditioning. *Learning and Motivation*, *26*, 141–160. https://doi.org/10.1016/0023-9690(95)90002-0

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy*, *10*, 179–199. https://doi.org/10.1016/0146-6402(88)90014-8

Baeyens, F., De Houwer, J., Vansteenwegen, D., & Eelen, P. (1998). Evaluative conditioning is a form of associative learning: On the artifactual nature of Field and Davey's (1997) artifactual account of evaluative learning. *Learning and Motivation*, *29*, 461–474. https://doi.org/10.1006/lmot.1998.1007

Baeyens, F., Eelen, P., Crombez, G., & van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy*, *30*, 133–142. https://doi.org/10.1016/0005-7967(92)90136-5

Baeyens, F., Eelen, P., van den Bergh, O., & Crombez, G. (1989). Acquired affective-evaluative value: Conservative but not unchangeable. *Behaviour Research and Therapy*, *27*, 279–287. https://doi.org/10.1016/0005-7967(89)90047-8

Baeyens, F., Hermans, D., & Eelen, P. (1993). The role of CS-US contingency in human evaluative conditioning. *Behaviour Research and Therapy*, *31*, 731–737. https://doi.org/10.1016/0005-7967(93)90003-D

Baeyens, F., Vanhouche, W., Crombez, G., & Eelen, P. (1998). Human evaluative flavor-flavor conditioning is not sensitive to post-acquisition US-inflation. *Psychologica Belgica*, *38*, 83–108.

Baeyens, F., Vansteenwegen, D., De Houwer, J., & Crombez, G. (1996). Observational conditioning of food valence in humans. *Appetite*, *27*, 235–250. https://doi.org/10.1006/appe.1996.0049

Bahrami, B., Carmel, D., Walsh, V., Rees, G., & Lavie, N. (2008). Spatial attention can modulate unconscious orientation processing. *Perception*, *37*, 1520–1528. https://doi.org/10.1068/p5999

Bahrami, B., Lavie, N., & Rees, G. (2007). Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Current Biology*, *17*, 509–513. https://doi.org/10.1016/j.cub.2007.01.070

Balas, R., & Gawronski, B. (2012). On the intentional control of conditioned evaluative responses. *Learning and Motivation*, *43*, 89–98. https://doi.org/10.1016/j.lmot.2012.06.003

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bar, M., & Biederman, I. (1999). Localizing the cortical region mediating visual awareness of object identity. *Proceedings of the National Academy of Sciences*, *96*, 1790–1793. https://doi.org/10.1073/pnas.96.4.1790

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *The American Journal of Psychology*, *105*, 181–199. http://dx.doi.org/10.2307/1423027

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In *Handbook of social cognition*. Hillsdale, NJ: Erlbaum.

Berscheid, E. (1985). Interpersonal attraction. In *Handbook of social psychology*. New York: Random House.

Blechert, J., Michael, T., Williams, S. L., Purkis, H. M., & Wilhelm, F. H. (2008). When two paradigms meet: Does evaluative learning extinguish in differential fear conditioning? *Learning and Motivation*, *39,* 58–70. https://doi.org/10.1016/j.lmot.2007.03.003

Boddez, Y., Bennett, M. P., van Esch, S., & Beckers, T. (2017). Bending rules: The shape of the perceptual generalisation gradient is sensitive to inference rules. *Cognition and Emotion*, *31*, 1444–1452. https://doi.org/10.1080/02699931.2016.1230541

Bonett, D. G. (2008). Confidence intervals for standardized linear contrasts of means. *Psychological Methods*, *13*, 99–109. https://doi.org/10.1037/1082-989X.13.2.99

Brewer, W. F. (1974). There is no convincing evidence for operant or classical conditioning in adult humans. In *Cognition and the symbolic processes*. Oxford, England: Lawrence Erlbaum.

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology*, *89*, 884–898. https://doi.org/10.1037/0022-3514.89.6.884

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.

Cook, S. A., & Harris, R. E. (1937). The verbal conditioning of the galvanic skin reflex. *Journal of Experimental Psychology*, *21*, 202–210. https://doi.org/10.1037/h0063197

Corneille, O., & Stahl, C. (2018). Associative attitude learning: A closer look at evidence and how it relates to attitude models. *Personality and Social Psychology Review*. https://doi.org/10.1177/1088868318763261

Davey, G. C. L. (1994). Is evaluative conditioning a qualitatively distinct form of classical conditioning? *Behaviour Research and Therapy*, *32*, 291–299. https://doi.org/10.1016/0005-7967(94)90124-4

Davey, G. C. L., & McKenna, I. (1983). The effects of postconditioning revaluation of CS1 and UCS following pavlovian second-order electrodermal conditioning in humans. *The Quarterly Journal of Experimental Psychology*, *35B*, 125–133. https://doi.org/10.1080/14640748308400899

Dawson, M. E., & Biferno, M. A. (1973). Concurrent measurement of awareness and electrodermal classical conditioning. *Journal of Experimental Psychology*, *101*, 55–62. https://doi.org/10.1037/h0035524

Dawson, M. E., & Shell, A. M. (1985). Information processing and human autonomic classical conditioning. In *Advances in psychophysiology*. Greenwich, CT: JAI Press.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*, 230–241.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning and Behavior*, *37*, 1–20. https://doi.org/10.3758/LB.37.1.1

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, *13*, e28046. https://doi.org/10.5964/spb.v13i3.28046

De Houwer, J., Baeyens, F., & Eelen, P. (1994). Verbal evaluative conditioning with undetected US presentations. *Behaviour Research and Therapy*, *32*, 629–633.

De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture-picture paradigm with random assignment of CSs to USs. *Journal of Experimental Psychology: Animal Behaviour Processes*, *26*, 237–242.

De Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin and Review*, *20*, 631–642. https://doi.org/10.3758/s13423-013-0386-3

De Houwer, J., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *The Quarterly Journal of Experimental Psychology*, *55A*, 965–985. https://doi.org/10.1080/02724980143000578

De Houwer, J., Hendrickx, H., & Baeyens, F. (1997). Evaluative learning with “subliminally” presented stimuli. *Consciousness and Cognition*, *6*, 87–107. https://doi.org/10.1006/ccog.1996.0281

De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon: On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. *Social Cognition*, *34*, 480–494. https://doi.org/10.1521/soco.2016.34.5.480

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*, 853–869. https://doi.org/10.1037/0033-2909.127.6.853

Dedonder, J., Corneille, O., Bertinchamps, D., & Yzerbyt, V. (2014). Overcoming correlational pitfalls: Experimental evidence suggests that evaluative conditioning occurs for explicit but not implicit encoding of CS-US pairings. *Social Psychological and Personality Science*, *5*, 250–257. https://doi.org/10.1177/1948550613490969

Dedonder, J., Corneille, O., Yzerbyt, V., & Kuppens, T. (2010). Evaluative conditioning of high-novelty stimuli does not seem to be based on an automatic form of associative learning. *Journal of Experimental Social Psychology*, *46*, 1118–1121. https://doi.org/10.1016/j.jesp.2010.06.004

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*, 204–211. https://doi.org/10.1016/j.tics.2006.03.007

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch’s t-test instead of Student’s t-test. *International Review of Social Psychology*, *30*, 92–101. https://doi.org/10.5334/irsp.82

Díaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition and Emotion, 19*, 245–268. https://doi.org/10.1080/02699930441000300

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59,* 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Faivre, N., Berthet, V., & Kouider, S. (2012). Nonconscious influences from emotional faces: A comparison of visual crowding, masking, and continuous flash suppression. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00129

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027. http://dx.doi.org/10.1037/0022-3514.69.6.1013

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition and Emotion*, *25*, 639–656. https://doi.org/10.1080/02699931.2010.513497

Field, A. P. (2000). I like it, but I'm not sure why: Can evaluative conditioning occur without conscious awareness? *Consciousness and Cognition, 9*, 13–36. https://doi.org/10.1006/ccog.1999.0402

Field, A. P., & Davey, G. C. L. (1999). Re-evaluating evaluative conditioning: A nonassociative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*, 211–224. http://dx.doi.org/10.1037/0097-7403.25.2.211

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1–74. http://dx.doi.org/10.1016/ S0065-2601(08)60317-2

Flaherty, C. F., & Rowan, G. A. (1986). Successive, simultaneous, and anticipatory contrast in the consumption of saccharin solutions. *Journal of Experimental Psychology: Animal Behavior Processes*, *12,* 381–393. https://doi.org/10.1037/0097-7403.12.4.381

Förderer, S., & Unkelbach, C. (2011). Beyond evaluative conditioning! Evidence for transfer of non-evaluative attributes. *Social Psychological and Personality Science*, *2,* 479–486. https://doi.org/10.1177/1948550611398413

Förderer, S., & Unkelbach, C. (2012). Hating the cute kitten or loving the aggressive pit-bull: EC effects depend on CS–US relations. *Cognition and Emotion*, *26*, 534–540. https://doi.org/10.1080/02699931.2011.588687

Förderer, S., & Unkelbach, C. (2013). On the stability of evaluative conditioning effects. *Social Psychology*, *44*, 380–389. https://doi.org/10.1027/1864-9335/a000150

Förderer, S., & Unkelbach, C. (2014). The moderating role of attribute accessibility in conditioning multiple specific attributes. *European Journal of Social Psychology*, *44,* 69–81. https://doi.org/10.1002/ejsp.1994

Förderer, S., & Unkelbach, C. (2015). Attribute conditioning: Changing attribute-assessments through mere pairings. *The Quarterly Journal of Experimental Psychology*, *68*, 144–164. https://doi.org/10.1080/17470218.2014.939667

Förderer, S., & Unkelbach, C. (2016). Changing US attributes after CS-US pairings changes CS-attribute-assessments: Evidence for CS-US associations in attribute conditioning. *Personality and Social Psychology Bulletin*, *42*, 350–365. https://doi.org/10.1177/0146167215626705

Fulcher, E. P., & Hammerl, M. (2001). When all is revealed: A dissociation between evaluative learning and contingency awareness. *Consciousness and Cognition*, *10*, 524–549. https://doi.org/10.1006/ccog.2001.0525

Gast, A. (2018). A declarative memory model of evaluative conditioning. *Social Psychological Bulletin*, *13*, e28590. https://doi.org/10.5964/spb.v13i3.28590

Gast, A., & Rothermund, K. (2011a). I like it because I said that I like it: Evaluative conditioning effects can be based on stimulus-response learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*, 466–476. https://doi.org/10.1037/a0023077

Gast, A., & Rothermund, K. (2011b). What you see is what will change: Evaluative conditioning effects depend on a focus on valence. *Cognition and Emotion*, *25*, 89–110. https://doi.org/10.1080/02699931003696380

Gast, A., Gawronski, B., & De Houwer, J. (2012). Evaluative conditioning: Recent developments and future directions. *Learning and Motivation*, *43*, 79–88. https://doi.org/10.1016/j.lmot.2012.06.004

Gawronski, B., Balas, R., & Creighton, L. A. (2014). Can the formation of conditioned attitudes be intentionally controlled? *Personality and Social Psychology Bulletin*, *40*, 419–432. https://doi.org/10.1177/0146167213513907

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2009). Operating principles versus operating conditions in the distinction between associative and propositional processes. *Behavioral and Brain Sciences*, *32*, 207–208. https://doi.org/10.1017/S0140525X09000958

Gawronski, B., & Bodenhausen, G. V. (2011). The associative - propositional evaluation model: Theory, evidence, and open questions. *Advances in experimental social psychology*, *44*, 59-127. https://doi.org/10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the associative-propositional evaluation model. *Social Psychological Bulletin*, *13*, e28024. https://doi.org/10.5964/spb.v13i3.28024

Gawronski, B., & Walther, E. (2012). What do memory data tell us about the role of contingency awareness in evaluative conditioning? *Journal of Experimental Social Psychology*, *48*, 617–623. https://doi.org/10.1016/j.jesp.2012.01.002

Gayet, S., Paffen, C. L. E., Belopolsky, A. V., Theeuwes, J., & Van der Stigchel, S. (2016). Visual input signaling threat gains preferential access to awareness in a breaking continuous flash suppression paradigm. *Cognition*, *149*, 77–83. https://doi.org/10.1016/j.cognition.2016.01.009

Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*, *60*, 328–331. https://doi.org/10.1198/000313006X152649

Gräf, M., & Unkelbach, C. (2016). Halo effects in trait assessment depend on information valence: Why being honest makes you industrious, but lying does not make you lazy. *Personality and Social Psychology Bulletin*, *42*, 290–310. https://doi.org/10.1177/0146167215627137

Gray, K. L. H., Adams, W. J., Hedger, N., Newton, K. E., & Garner, M. (2013). Faces and awareness: Low-level, not emotional factors determine perceptual dominance. *Emotion*, *13*, 537–544. https://doi.org/10.1037/a0031403

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

Hermans, D., Crombez, G., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Differential effects of extinction. In *Advances in psychology research*. Hauppauge, NY, US: Nova Science Publishers.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*, 390–421. https://doi.org/10.1037/a0018916

Hu, X., Gawronski, B., & Balas, R. (2017a). Propositional versus dual-process accounts of evaluative conditioning: I. The effects of co-occurrence and relational information on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *43*, 17–32. https://doi.org/10.1177/0146167216673351

Hu, X., Gawronski, B., & Balas, R. (2017b). Propositional versus dual-process accounts of evaluative conditioning: II. The effectiveness of counter-conditioning and counter-instructions in changing implicit and explicit evaluations. *Social Psychological and Personality Science*, *8*, 858–866. https://doi.org/10.1177/1948550617691094

Hughes, S., De Houwer, J., & Barnes-Holmes, D. (2016). The moderating impact of distal regularities on the effect of stimulus pairings: A novel perspective on evaluative conditioning. *Experimental Psychology*, *63*, 20–44. https://doi.org/10.1027/1618-3169/a000310

Hütter, M., Kutzner, F., & Fiedler, K. (2014). What is learned from repeated pairings? On the scope and generalizability of evaluative conditioning. *Journal of Experimental Psychology: General*, *143*, 631–643. https://doi.org/10.1037/a0033409

Hütter, M., & Sweldens, S. (2013). Implicit misattribution of evaluative responses: Contingency-unaware evaluative conditioning requires simultaneous stimulus presentations. *Journal of Experimental Psychology: General*, *142*, 638–643. https://doi.org/10.1037/a0029989

Hütter, M., & Sweldens, S. (2018). Dissociating controllable and uncontrollable effects of affective stimuli on attitudes and consumption. *Journal of Consumer Research*, *45*, 320–349. https://doi.org/10.1093/jcr/ucx124

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General*, *141*, 539–557. https://doi.org/10.1037/a0026477

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Jeffreys, S. H. (1998). *The theory of probability*. OUP Oxford.

Jiang, Y., & He, S. (2006). Cortical responses to invisible faces: Dissociating subsystems for facial-information processing. *Current Biology*, *16*, 2023–2029. https://doi.org/10.1016/j.cub.2006.08.084

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology*, *96*, 933–948. https://doi.org/10.1037/a0014747

Jones, C. R., Olson, M. A., & Fazio, R. H. (2010). Evaluative conditioning: The “how” question. *Advances in Experimental Social Psychology*, *43*, 205–255. https://doi.org/10.1016/S0065-2601(10)43005-1

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In *Punishment and aversive behavior*. New York: Appleton-Century-Crofts.

Kanai, R., Tsuchiya, N., & Verstraten, F. A. J. (2006). The scope and limits of top-down attention in unconscious visual processing. *Current Biology*, *16*, 2332–2336. https://doi.org/10.1016/j.cub.2006.10.001

Kaunitz, L., Fracasso, A., & Melcher, D. (2011). Unseen complex motion is modulated by attention and generates a visible aftereffect. *Journal of Vision*, *11*, 1–9. https://doi.org/10.1167/11.13.10

Kim, J., Allen, C. T., & Kardes, F. R. (1996). An investigation of the mediational mechanisms underlying attitudinal conditioning. *Journal of Marketing Research*, *33*, 318–328. https://doi.org/10.1177/002224379603300306

Köhler, W. (1925). *The mentality of apes*. https://doi.org/10.4324/9781351294966

Kruglanski, A. W., & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry*, *10*, 83–109. https://doi.org/10.1207/S15327965PL100201

Leiner, D. J. (2016). SoSci Survey (Version 2.6.00). Retrieved from soscisurvey.de

Levey, A. B., & Martin, I. (1987). Evaluative conditioning - A case for hedonic transfer. In *Theoretical foundations of behavior therapy*. https://doi.org/10.1007/978-1-4899-0827-8_5

Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human pavlovian conditioning: Extinct, but still there? *Learning and Motivation*, *34*, 219–239. https://doi.org/10.1016/S0023-9690(03)00011-0

Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple pavlovian conditioning. *Cognition and Emotion*, *19*, 269–282. https://doi.org/10.1080/02699930441000319

Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*, 431–440. https://doi.org/10.1037/0097-7403.32.4.431

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, *28*, 3–26. https://doi.org/10.1037/0097-7403.28.1.3

Maes, E., Boddez, Y., Alfei, J. M., Krypotos, A.-M., D’Hooge, R., De Houwer, J., & Beckers, T. (2016). The elusive nature of the blocking effect: 15 failures to replicate. *Journal of Experimental Psychology: General*, *145*, e49–e71.

March, D. S., Olson, M. A., & Fazio, R. H. (2018). The implicit misattribution model of evaluative conditioning. *Social Psychological Bulletin*, *13*, e27574. https://doi.org/10.5964/spb.v13i3.27574

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mausfeld, R. (2003). No psychology in - no psychology out. *Psychologische Rundschau*, *54*, 185–191. http://dx.doi.org/10.1026//0033-3042.54.3 .185

Mazur, J. E. (1994). *Learning and Behavior*. Englewood Cliffs, NJ: Prentice-Hall.

McClelland, J. L., McNaughton, B. L., & O’Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457. https://doi.org/10.1037/0033-295X.102.3.419

Mierop, A., Hütter, M., & Corneille, O. (2017). Resource availability and explicit memory largely determine evaluative conditioning effects in a paradigm claimed to be conducive to implicit attitude acquisition. *Social Psychological and Personality Science*, *8*, 758–767. https://doi.org/10.1177/1948550616687093

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183–198. https://doi.org/10.1017/S0140525X09000855

Mitchell, C. J., Wardle, S. G., Lovibond, P. F., Weidemann, G., & Chang, B. P. I. (2010). Do reaction times in the perruchet effect reflect variations in the strength of an associative link? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 567–572. https://doi.org/10.1037/a0018433

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*, *67*, 263–287. https://doi.org/10.1146/annurev-psych-122414-033550

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

Moran, T., & Bar-Anan, Y. (2013). The effect of object–valence relations on automatic evaluation. *Cognition and Emotion*, *27*, 743–752. https://doi.org/10.1080/02699931.2012.732040

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of bayes factors for common designs (Version R package version 0.9.12-2). Retrieved from https://CRAN.R-project.org/package=BayesFactor

Moutoussis, K., & Zeki, S. (2002). The relationship between cortical activation and perception investigated with invisible stimuli. *Proceedings of the National Academy of Sciences*, *99*, 9527–9532. https://doi.org/10.1073/pnas.142305699

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250–256. https://doi.org/10.1037/0022-3514.35.4.250

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*, 413–417. https://doi.org/10.1111/1467-9280.00376

Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition*, *20*, 89–104. https://doi.org/10.1521/soco.20.2.89.20992

Olson, M. A., Kendrick, R. V., & Fazio, R. H. (2009). Implicit learning of evaluative vs. non-evaluative covariations: The role of dimension accessibility. *Journal of Experimental Social Psychology*, *45*, 398–403. https://doi.org/10.1016/j.jesp.2008.10.007

Payne, B. K., Brown-Iannuzzi, J. L., & Loersch, C. (2016). Replicable effects of primes on human behavior. *Journal of Experimental Psychology: General*, *145*, 1269–1279. https://doi.org/10.1037/xge0000201

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293. https://doi.org/10.1037/0022-3514.89.3.277

Pearce, J. M. (1987). A model for stimulus generalization in pavlovian conditioning. *Psychological Review*, *94*, 61–73. https://doi.org/10.1037/0033-295X.94.1.61

Perruchet, P. (1985). A pitfall for the expectancy theory of human eyelid conditioning. *Pavlovian Journal of Biological Science*, *20*, 163–170. https://doi.org/10.1007/BF03003653

Perruchet, P. (2015). Dissociating conscious expectancies from automatic link formation in associative learning: A review on the so-called Perruchet effect. *Journal of Experimental Psychology: Animal Learning and Cognition*, *41*, 105–127.

Petty, R., & Cacioppo, J. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York, NY: Springer-Verlag.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 130–144. https://doi.org/10.1037/0278-7393.33.1.130

Pleyers, G., Corneille, O., Yzerbyt, V., & Luminet, O. (2009). Evaluative conditioning may incur attentional costs. *Journal of Experimental Psychology: Animal Behavior Processes*, *35,* 279–285. https://doi.org/10.1037/a0013429

Raio, C. M., Carmel, D., Carrasco, M., & Phelps, E. A. (2012). Nonconscious fear is quickly acquired but swiftly forgotten. *Current Biology*, *22,* R477–R479. https://doi.org/10.1016/j.cub.2012.04.023

Ratcliff, R., & McKoon, G. (1981). Does activation really spread? *Psychological Review*, *88*, 454–462. https://doi.org/10.1037/0033-295X.88.5.454

Rescorla, R. A. (1973). Effects of US habituation following conditioning. *Journal of Comparative and Physiological Psychology*, *82*, 137–143. https://doi.org/10.1037/h0033815

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning II: Current research and theory*. New York, US: Appleton-Century-Crofts.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*, 225–237. https://doi.org/10.3758/PBR.16.2.225

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, *17*, 688–690. https://doi.org/10.1093/beheco/ark016

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*, 995–1008. https://doi.org/10.1037/0022-3514.91.6.995

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science*, *17*, 954–958. https://doi.org/10.1111/j.1467-9280.2006.01811.x

Seitz, A. R., Kim, D., & Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron*, *61*, 700–707. https://doi.org/10.1016/j.neuron.2009.01.016

Shanks, D. R. (2005). Implicit learning. In *Handbook of cognition*. London, UK: Sage.

Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology*, *61*, 273–301. https://doi.org/10.1146/annurev.psych.093008.100519

Shanks, D. R., & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 405–415.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century.

Skinner, B. F. (1984). The evolution of behavior. *Journal of the Experimental Analysis of Behavior*, *41*, 217–221. https://doi.org/10.1901/jeab.1984.41-217

Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences*, *109*, 19614–19619. https://doi.org/10.1073/pnas.1211645109

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology*, *74*, 837–848. https://doi.org/10.1037/0022-3514.74.4.837

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22. http://dx.doi.org/10.1037/0033-2909 .119.1.3

Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, *44*, 430–444. https://doi.org/10.1037/h0062885

Staats, A. W., Staats, C. K., & Heard, W. G. (1959). Language conditioning of meaning using a semantic generalization paradigm. *Journal of Experimental Psychology*, *57*, 187–192. https://doi.org/10.1037/h0042274

Staats, C. K., & Staats, A. W. (1957). Meaning established by classical conditioning. *Journal of Experimental Psychology*, *54*, 74–80. https://doi.org/10.1037/h0047716

Stahl, C., & Aust, F. (2018). Evaluative conditioning as memory-based judgment. *Social Psychological Bulletin*, *13*, e28589. https://doi.org/10.5964/spb.v13i3.28589

Stahl, C., Haaf, J., & Corneille, O. (2016). Subliminal evaluative conditioning? Above-chance CS identification may be necessary and insufficient for attitude learning. *Journal of Experimental Psychology: General*, *145*, 1107–1131. https://doi.org/10.1037/xge0000191

Stahl, C., & Heycke, T. (2016). Evaluative conditioning with simultaneous and sequential pairings under incidental and intentional learning conditions. *Social Cognition*, *34*, 382–412. https://doi.org/10.1521/soco.2016.34.5.382

Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple unconditioned stimuli: The role of contingency awareness. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 286–291. https://doi.org/10.1037/a0013255

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, *97*, 404–420. https://doi.org/10.1037/a0016196

Stein, T., & Sterzer, P. (2011). High-level face shape adaptation depends on visual awareness: Evidence from continuous flash suppression. *Journal of Vision*, *11*, 1–14. https://doi.org/10.1167/11.8.5

Stein, T., & Sterzer, P. (2012). Not just another face in the crowd: Detecting emotional schematic faces during continuous flash suppression. *Emotion*, *12*, 988–996. https://doi.org/10.1037/a0026944

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The role of awareness in attitude formation through evaluative conditioning. *Personality and Social Psychology Review*, *18*, 187–209. https://doi.org/10.1177/1088868314527832

Sweldens, S., Van Osselaer, S. M. J., & Janiszewski, C. (2010). Evaluative conditioning procedures and the resilience of conditioned brand attitudes. *Journal of Consumer Research*, *37,* 473–489. https://doi.org/10.1086/653656

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, *2*, i–109. https://doi.org/10.1037/h0092987

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8,* 1096–1101. https://doi.org/10.1038/nn1500

Unkelbach, C., & Fiedler, K. (2016). Contrastive CS-US relations reverse evaluative conditioning effects. *Social Cognition*, *34*, 413–434. https://doi.org/10.1521/soco.2016.34.5.413

Unkelbach, C., & Förderer, S. (2018). A model of attribute conditioning. *Social Psychological Bulletin*, *13*, e28568. https://doi.org/10.5964/spb.v13i3.28568

Unkelbach, C., & Högden, F. (in press). Why does George Clooney make coffee sexy? The case for attribute conditioning. *Current Directions in Psychological Science.*

Unkelbach, C., Stahl, C., & Förderer, S. (2012). Changing CS features alters evaluative responses in evaluative conditioning. *Learning and Motivation*, *43*, 127–134. https://doi.org/10.1016/j.lmot.2012.04.003

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, 1088868318795730. https://doi.org/10.1177/1088868318795730

Vizueta, N., Patrick, C. J., Jiang, Y., Thomas, K. M., & He, S. (2012). Dispositional fear, negative affectivity, and neuroimaging response to visually suppressed emotional faces. *NeuroImage, 59*, 761–771. https://doi.org/10.1016/j.neuroimage.2011.07.015

Vogel, T., & Wänke, M. (2016). *Attitudes and attitude change*. Psychology Press.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin and Review, 14*, 779–804. https://doi.org/10.3758/BF03194105

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In *Information processing in animals: Memory mechanisms.* Hillsdale, NJ: Erlbaum.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the back door: The US-revaluation effect. *Cognition and Emotion*, *23*, 889–917. https://doi.org/10.1080/02699930802212423

Walther, E., Langer, T., Weil, R., & Komischke, M. (2011). Preferences surf on the currents of words: Implicit verb causality influences evaluative conditioning. *European Journal of Social Psychology*, *41,* 17–22. https://doi.org/10.1002/ejsp.785

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology: Animal Behavior Processes*, *32,* 454–459. https://doi.org/10.1037/0097-7403.32.4.454

Wheeler, D. S., Sherwood, A., & Holland, P. C. (2008). Excitatory and inhibitory learning with absent stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *34,* 247–255. https://doi.org/10.1037/0097-7403.34.2.247

White, K., & Davey, G. C. L. (1989). Sensory preconditioning and UCS inflation in human ‘fear’ conditioning. *Behaviour Research and Therapy*, *27*, 161–166. https://doi.org/10.1016/0005-7967(89)90074-0

Wong, A. H. K., & Lovibond, P. F. (2017). Rule-based generalisation in single-cue and differential fear conditioning in humans. *Biological Psychology*, *129*, 111–120. https://doi.org/10.1016/j.biopsycho.2017.08.056

Yang, E., Brascamp, J., Kang, M.-S., & Blake, R. (2014). On the use of continuous flash suppression for the study of visual processing outside of awareness. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00724

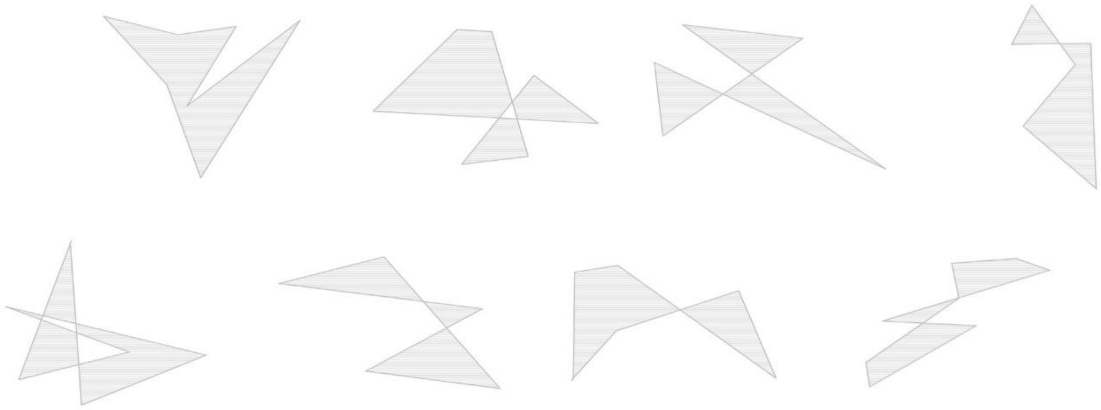
Yang, E., Hong, S.-W., & Blake, R. (2010). Adaptation aftereffects to facial expressions suppressed from visual awareness. *Journal of Vision*, *10*, 1–13. https://doi.org/10.1167/10.12.24

Yang, E., Zald, D. H., & Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion*, *7*, 882–886. https://doi.org/10.1037/1528-3542.7.4.882

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation*, *43*, 155–165. https://doi.org/10.1016/j.lmot.2012.02.003

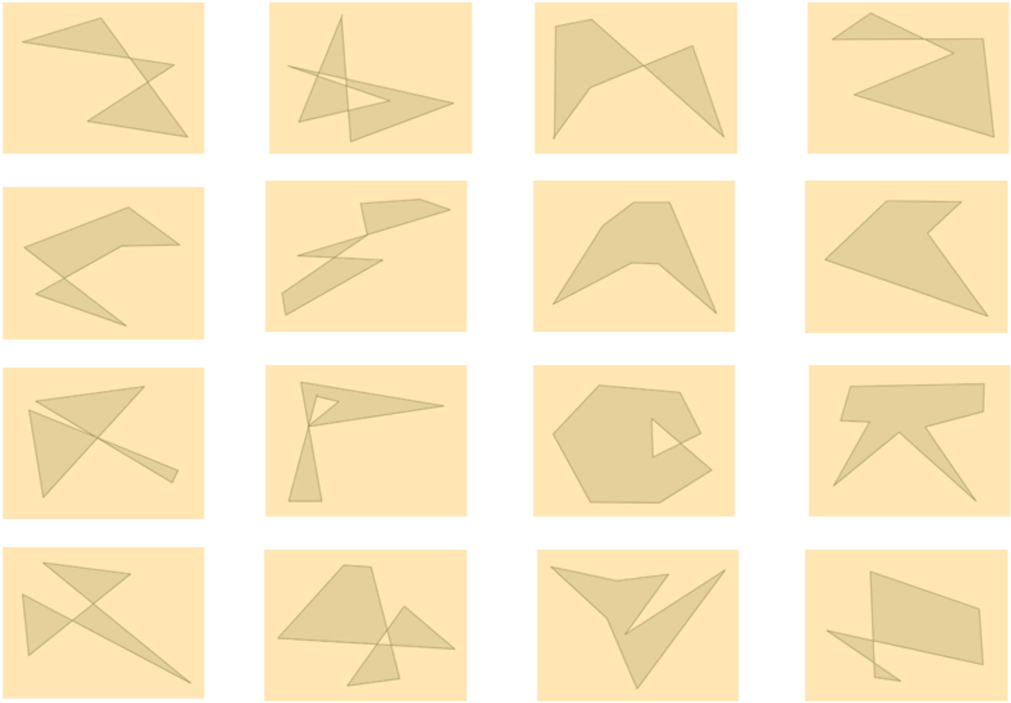
Appendix A

Eight CSs used in Experiment 3.1



Appendix B

16 CSs used in Experiment 5.1



Appendix C

Details regarding the coding of rule versus non-rule learners and analyses of participants' performance in the learning phase in all experiments

In all experiments, two independent coders classified participants' responses in the open-ended question after the learning phase. In Experiment 5.1, they disagreed in six cases, which were resolved by discussion. Only six participants were classified as having inferred the underlying rule (i.e., "rule learners"), while 71 participants did not correctly report the rule (i.e., "non-rule learners"). Overall, participants responded correctly in 13.87 (*SD* = 2.55) of the twenty trials, which is significantly different from a 10 correct responses guessing threshold, $t(76) = 13.31$, $p < .001$, $d = 1.52$; thus, participants paid attention during the learning phase.

To compare performance depending on the rule inference, we computed a Welch's t-test (Delacre, Lakens, & Leys, 2017; Ruxton, 2006). Rule learners provided relatively more correct responses (*M* = 15.50, *SD* = 1.87) than non-rule learners (*M* = 13.73, *SD* = 2.56), although this difference was not significant on a standard alpha level, $t(6.70) = 2.15$, $p = .070$, standardized mean difference=0.79. The mean difference was standardized by the average standard deviation of the two groups. That is, $standardized\ mean\ difference = \frac{x_1 - x_2}{mean\ SD}$, and $mean\ SD = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$ (Bonett, 2008).

In Experiment 5.2a, the coders disagreed in six cases which were resolved by a third coder. Eighteen participants were classified as rule learners; 63 were classified as non-rule learners. Overall, participants responded correctly on 7.72 (*SD* = 1.63) of the ten trials, which was better than the chance threshold of five, $t(80) = 15.00$, $p < .001$, $d = 1.67$. Although descriptively, rule learners performed better (*M* = 8.22, *SD* = 1.22) than non-rule learners (*M* =

7.57, *SD* = 1.71), this comparison did not reach significance, *t*(38.324) = 1.82, *p* = .077, *standardized mean difference* = 0.44.

Concerning evaluation responses of the stimuli in the learning phase of Experiment 5.2a, we conducted a 2 (learner type: non-rule vs. rule learner; between participants) x 2 (paired valence: positive vs. negative; within participants) mixed ANOVA. We observed a standard EC effect: Participants evaluated CSs paired with positive USs more positively (*M* = 6.36, *SD* = 1.82) than those paired with negative USs (*M* = 4.04, *SD* = 1.65), *F*(1,79) = 78.34, *p* < .001, η_p^2 = .50. No other effects were significant, all *F*s < 1.9, all *p*s > .17, all η_p^2s < .03.

In Experiment 5.2b, the coders disagreed in six cases which were resolved by a third coder. Twenty-seven participants were classified as rule-learners; 53 were classified as non-rule learners. Overall, they responded correctly in 7.48 of the ten trials (*SD* = 1.59). This was better than chance level, *t*(79) = 13.92, *p* < .001, *d* = 1.56. Rule-learners performed better (*M* = 8.15, *SD* = 1.32) than non-rule learners (*M* = 7.13, *SD* = 1.62), *t*(62.60) = 3.01, *p* = .004, *standardized mean difference* = -0.69.

A 2 (learner type: non-rule associative vs. rule learner; between participants) x 2 (paired valence: positive vs. negative; within participants) mixed ANOVA with participants' liking judgments from the learning phase as dependent variable showed an EC effect. CSs paired with positive USs were evaluated as more positive (*M* = 6.12, *SD* = 1.93) than those paired with negative USs (*M* = 4.08, *SD* = 2.08), *F*(1,78) = 87.73, *p* < .001, η_p^2 = .53. There was also a non-significant interaction with learner type, *F*(1,78) = 3.90, *p* = .052, η_p^2 = .05. The main effect of learner type was not significant, *F*(1,78) = 0.62, *p* = .434, η_p^2 < .01.

In Experiment 5.3, the coders who coded participants' responses on the open-ended questions agreed in all cases. Sixty-eight participants were classified as rule learners (62 in the rule instruction condition and 6 in the no rule instruction condition); 50 were classified as non-rule learners (12 in the rule instruction condition and 38 in the no rule instruction condition). Overall, participants responded correctly in 13.74 (*SD* = 2.65) of the 20 learning trials in which they provided categorization responses, which was different from guessing, *t*(117) = 15.31, *p* < .001, *d* = 1.41. Descriptively, participants classified as rule learners performed better (*M* = 14.06, *SD* = 2.79) than those classified as non-rule learners (*M* = 13.30, *SD* = 2.42); this difference was not significant, though, *t*(112.76) = 1.58, *p* = .117, *standardized mean difference* = 0.29.

In Experiment 5.4, the coders disagreed in four cases that were resolved by a third coder. Thirty-three participants were classified as rule learners; 46 as non-rule learners. Overall, participants correctly classified stimuli as positive or negative in 7.39 (*SD* = 1.40) of the ten response trials, which differed from the guessing threshold of five, *t*(78) = 15.19, *p* < .001, *d* = 1.71. Furthermore, rule learners performed better (*M* = 7.85, *SD* = 1.44) than non-rule learners, (*M* = 7.07, *SD* = 1.29), *t*(64.31) = 2.49, *p* = .015, *standardized mean difference* = 0.57.

Appendix D

Analyses data from Experiment 5.3 with manipulated factor “rule instruction” (rule instructed vs. not instructed) as opposed to the measured factor “learner type” (rule learners vs. non-rule learner)

Learning phase. A t-test showed that participants who were instructed the rule descriptively performed slightly better (M = 13.92, SD = 2.76) than participants who were not instructed the rule (M = 13.43, SD = 2.46), t(98.725) = 0.99, p = .323, *standardized mean difference* = -0.19.

Categorization. There was a main effect of category according to the rule. “Mammal” stimuli were more often classified as mammals (M = 0.58, SD = 0.49) than “reptile” stimuli (M = 0.34, SD = 0.47), F(1,116) = 11.76, p = .001, η_p^2 = .09. This effect was qualified by an interaction with rule instruction, F(1,116) = 36.15, p < .001, η_p^2 = .24. Follow-up ANOVAs showed that participants that were not instructed the rule classified “mammal” stimuli less often as mammals (M = 0.38, SD = 0.48) than “reptile” stimuli (M = 0.50, SD = 0.50). This effect was not significant, though, F(1,43) = 2.05, p = .160, η_p^2 = .05. Participants who were instructed the rule showed the reversed pattern. They categorized novel compounds in line with the rule: Stimuli belonging to the mammal category were correctly classified as mammals more often (M = 0.71, SD = 0.46) than stimuli belonging to the reptile category (M = 0.25, SD = 0.43), F(1,73) = 72.61, p < .001, η_p^2 = .50. The main effect of rule instruction in the overall ANOVA was not significant, F(1,116) = 1.55, p = .215, η_p^2 = .01.

Evaluation. We observed a main effect of category according to the rule for the evaluation of the novel stimuli. Participants’ liking was in line with the paired valence of the compounds' elements: "Positive" compounds (that consisted of two negative elements) were evaluated more negatively (M = 4.56, SD = 2.11) than "negative" compounds (that consisted of two positive elements; M = 5.02, SD = 2.19), F(1,116) = 4.14, p = .044, $\mathrm{\eta}_{\mathrm{p}}^{2}$ = .03. All other effects were not significant, all Fs < 0.95, all ps > .33, all $\mathrm{\eta}_{\mathrm{p}}^{2}$s < .01.

Comparing categorization and evaluation. To test whether the categorization and evaluation differ regarding the size of the interaction between category and rule instruction, we z-standardized scores of both measures and submitted them to a 2 (rule instruction: yes vs. no) x 2 (category according to the rule: mammal/positive vs. reptile/negative) x 2 (measure: categorization vs. evaluation) mixed ANOVA. We observed a two-way interaction of rule instruction and category (F(1,116) = 22.23, p < .001, η_p^2 = .16), a two-way interaction of measure and category (F(1,116) = 17.16, p < .001, η_p^2 = .13) and, most importantly, a three-way interaction showing that the measures differ regarding the two-way interaction of rule instruction and correct category, F(1,116) = 15.29, p < .001, η_p^2 = .12. All other effects: all Fs < 2.52, all ps > .11, all $\eta_p^2 s$ < .03.

Appendix E

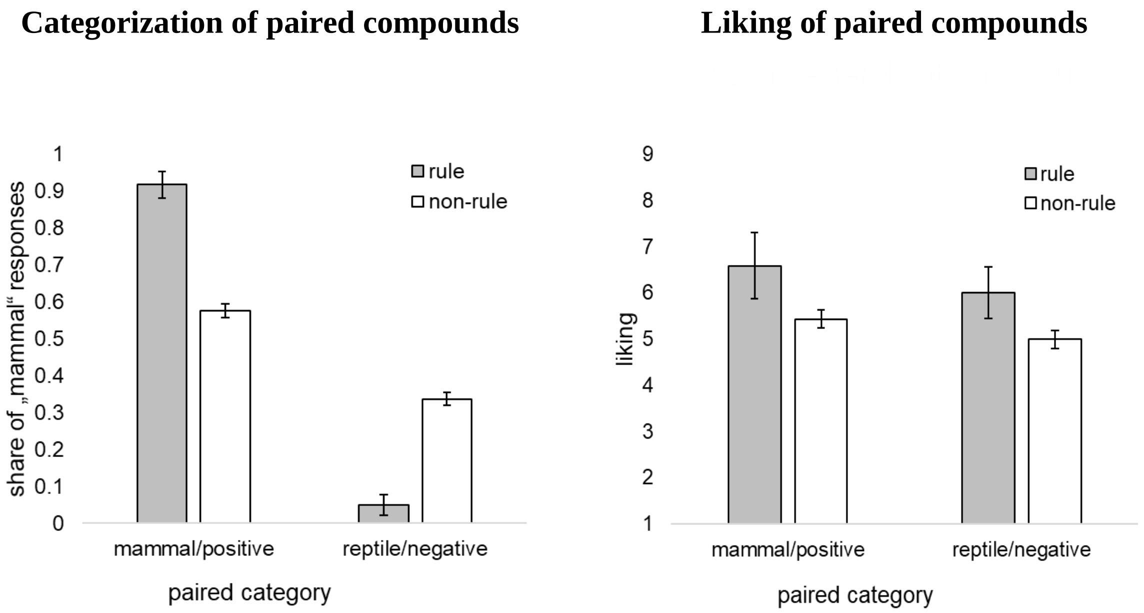
Analyses of paired compounds

To test whether the observed results are specific to generalization, we conducted the same analyses for compound stimuli that were paired in the learning phase.

Experiment 5.1.

Categorization. The ANOVA showed a category main effect: Mammal-paired stimuli were correctly categorized as mammals more often (*M* = 0.60, *SD* = 0.49) than reptile-paired stimuli (*M* = 0.31, *SD* = 0.46), *F*(1,75) = 34.62, *p* < .001, η_p^2 = .32. There was also an interaction between category and learner type, *F*(1,75) = 11.13, *p* = .001, η_p^2 = .13. Follow-up ANOVAs showed that both rule learners and non-rule learners correctly categorized the paired compounds (rule learners: mammal-paired: *M* = 0.92, *SD* = 0.28, reptile-paired: *M* = 0.05, *SD* = 0.22, *F*(1,5) = 307.27, *p* < .001, η_p^2 = .98; non-rule learners: mammal-paired: *M* = 0.58, *SD* = 0.49, reptile-paired: *M* = 0.34, *SD* = 0.47, *F*(1,70) = 19.53, *p* < .001, η_p^2 = .22). Thus, in contrast to generalization compounds, both types of learner correctly classified the paired compounds. The main effect of learner type in the overall ANOVA was not significant, *F*(1,75) = 0.10, *p* = .755, η_p^2 < .01.

Evaluation. The ANOVA showed only a main effect of learner type. Rule users evaluated the compounds more positively (*M* = 6.29, *SD* = 2.20), than non-rule learners (*M* = 5.21, *SD* = 2.32), *F*(1,75) = 4.67, *p* = .034, η_p^2 = .06. No other effects were significant, all *F*s < 0.84, all *p*s > .36, all η_p^2s < .02. For generalization stimuli, in contrast, we observed a main effect of category. The difference between paired and generalization compounds can be explained by the fact that for paired compounds there were two conflicting sources informing their liking judgments. The elements' pairings drove the evaluation into the opposite direction than the compound's pairing. If, for example, A and B were paired with positive USs and AB with a negative USs, upon evaluation of AB, not only AB's pairing will inform the judgment into a negative direction but A and B's pairing will also influence the judgment into a positive direction (i.e., they cancel each other out). This can explain the absence of an effect of paired valence (i.e., EC effect).


Categorization of paired compounds

Liking of paired compounds

Figure A1. Data from Experiment 5.1: The left panel shows the proportion of "mammal" categorization responses towards the paired compound stimuli, grouped by the paired category as a function of learner type. Thus, high values for the mammal/positive category and low values for the reptile/negative category indicate good categorization performance. The right panel shows the evaluation of the paired compound stimuli as a function of the paired category and learner type. Error bars show the standard error of the mean.

Paired versus generalization compounds. While non-rule learners show the same pattern of responding in both measures, rule learners show different pattern of responding for categorization and evaluation. To test whether this observed dissociation in rule learners is specific to generalization stimuli we pooled rule learners' responses for paired and generalization compounds and included stimulus type as a factor into the ANOVA.

A 2 (category according to the rule: mammal/positive vs. reptile/negative) x 2 (stimulus type: paired compound vs. generalization compound) mixed ANOVA for rule learners' categorization responses showed only a main effect of correct category. "Mammal" compounds (i.e., among paired compounds, those that were paired with the mammal category in the learning phase, and among generalization stimuli, those that would rule-wise belong to the mammal category) were categorized as mammals more often (*M* = 0.88, *SD* = 0.33) than "reptile" compounds (*M* = 0.08*, SD* = 0.28), *F*(1,5) = 204.19, *p* < .001, η_p^2 = .98. The main effect of stimulus type was not significant (*F*(1,5) = 0.02, *p* = .905, η_p^2 < .01), and, more importantly, the interaction was also not significant, *F*(1,5) = 1.71, *p* = .248, η_p^2 = .25. Thus, both for paired and for generalization stimuli, rule learners categorized correctly.

The same ANOVA with liking responses showed a main effect of stimulus type. Paired stimuli were evaluated more positively (*M* = 6.29, *SD* = 2.20) than generalization stimuli (*M* = 4.83, *SD* = 2.37), *F*(1,5) = 12.63, *p* = .016, η_p^2 = .72. The interaction between correct category and stimulus type was not significant on a standard alpha level, *F*(1,5) = 5.11, *p* = .073, η_p^2 = .51.

These analyses do not allow for strong conclusions whether the dissociation between categorization and liking judgements is specific to generalization. The low number of rule learners made it difficult to show that they use their rule knowledge (at least to a greater extent) for both measures within paired compounds. Experiments 5.2a and 5.2b, however, resolve this issue of statistical power and clearly show that our findings are specific for generalization.

Experiment 5.2a.

Categorization. The ANOVA showed a main effect of category according to the rule. Mammal-paired compounds were correctly categorized as mammals more often (*M* = 0.61, *SD* = 0.49) than reptile-paired compounds (*M* = 0.37, *SD* = 0.48), *F*(1,79) = 29.25, *p* < .001, η_p^2 = .27. We also observed an interaction between category and learner type, *F*(1,79) = 17.75, *p* < .001, η_p^2 = .18. Follow-up ANOVAs showed that rule learners categorized correctly: Mammal-paired: *M* = 0.84, *SD* = 0.36, reptile-paired: *M* = 0.08, *SD* = 0.27, *F*(1,17) = 63.78 *p* < .001, η_p^2 = .79. Non-rule learners descriptively also categorized paired compounds correctly (mammal-paired: *M* = 0.55, *SD* = 0.50, reptile-paired: *M* = 0.45, *SD* = 0.50) but this difference was not significant, *F*(1,62) = 1.40, *p* = .241, η_p^2 = .02. Like in Experiment 5.1, this suggests that, both types of learner categorized paired compounds rather correctly. Generalization stimuli, in contrast, were clearly only categorized correctly by rule learners. The main effect of learner type in the overall ANOVA was not significant, *F*(1,79) = 0.45, *p* = .506, η_p^2 < .01.

Evaluation. The ANOVA showed no significant effects, all *F*s < 3.3, all *p*s > .07, all η_p^2s < .04. Like in Experiment 5.1, we believe, influence of the compound's pairing and its elements' pairings cancelled each other out.

Paired versus generalization compounds. Analyzing rule learners' categorization responses for paired and generalization stimuli, we observed a main effect of correct category. "Mammal" compounds were categorized as mammals more often (*M* = 0.77, *SD* = 0.42) than "reptile" compounds (*M* = 0.15, *SD* = 0.36), *F*(1,17) = 37.84, *p* < .001, η_p^2 = .69. The main effect of stimulus type was not significant (*F*(1,17) < 0.01, *p* > .999, η_p^2 < .01) and the interaction did also not reach significant, *F*(1,17) = 3.25, *p* = .089, η_p^2 = .16.

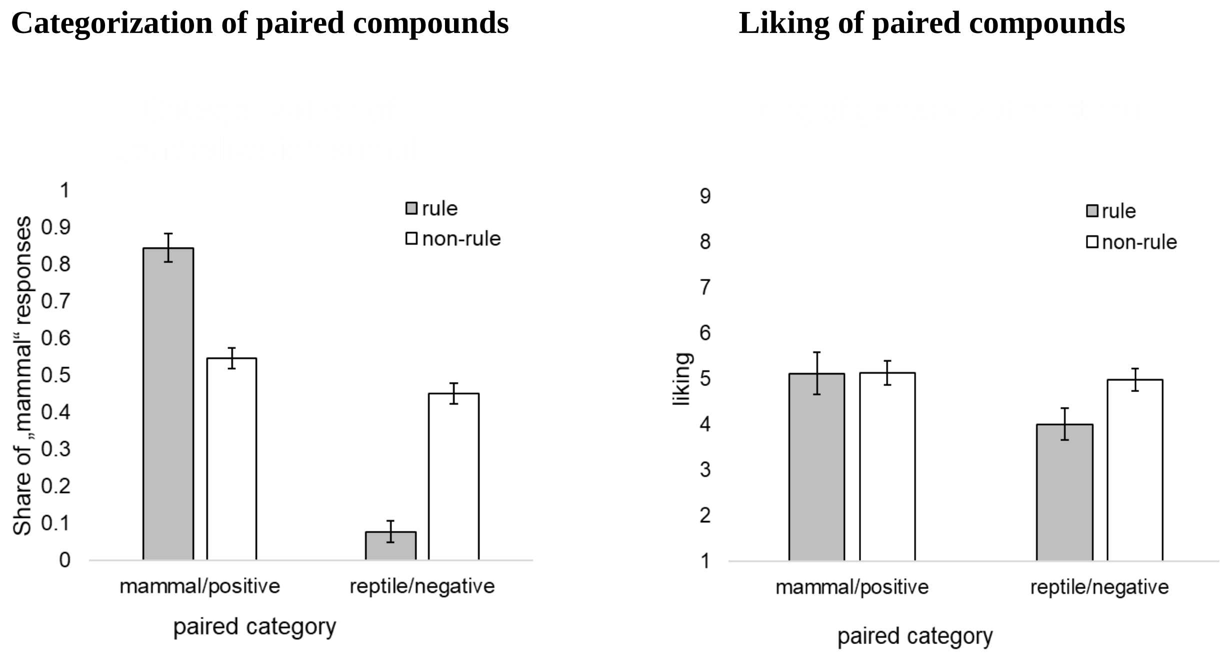


Figure A2. Data from Experiment 5.2a: The left panel shows the proportion of ”mammal” categorization responses towards the paired compound stimuli, grouped by the paired category as a function of learner type. Thus, high values for the mammal/positive category and low values for the reptile/negative category indicate good categorization performance. The right panel shows the evaluation of the paired compound stimuli as a function of the paired category and learner type. Error bars show the standard error of the mean.

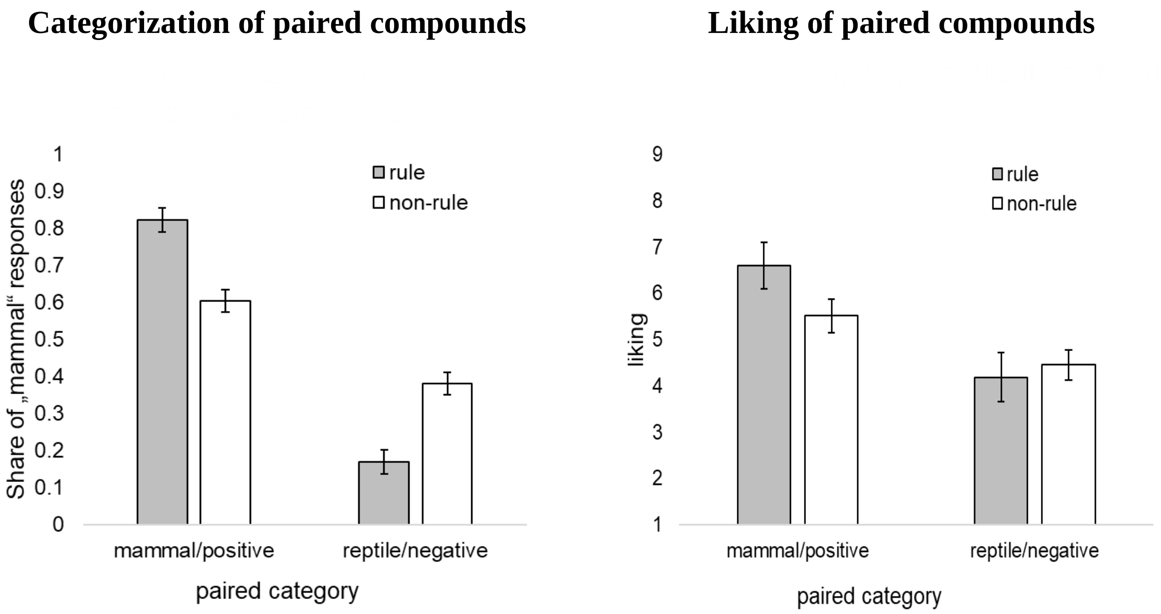
The ANOVA for rule learners’ liking showed only an interaction between correct category and stimulus type, $F(1,17) = 15.30$, $p = .001$, $\eta_p^2 = .47$. A follow-up ANOVA for paired stimuli showed that they were evaluated “in line with the rule”. Rule learners evaluated positively paired compounds more positively ($M = 5.11$, $SD = 1.97$) than negatively paired compounds ($M = 4.00$, $SD = 1.46$), $F(1,17) = 5.12$, $p = .037$, $\eta_p^2 = .23$. Generalization stimuli, in contrast, were evaluated “contrary to the rule” (reported in the paper). No other effect in the overall ANOVA for liking were significant, all Fs < 0.99, ps > .33, η_p^2s < .06.

Thus, the use and non-use of rules when categorizing respectively evaluating novel generalization stimuli is not present for paired compounds. Rather, rules are applied for both measures.

Experiment 5.2b.

Categorization. There was a main effect of correct category in the ANOVA. Participants classified mammal-paired compounds as mammals more often (M = 0.68, SD = 0.47) than reptile-paired compounds (M = 0.31, SD = 0.46), F(1,78) = 38.41, p < .001, η_p^2 = .33. An interaction with learner type qualified this effect, F(1,78) = 9.25, p = .003, η_p^2 = .11. Follow-up ANOVAs for each learner type showed that both rule and non-rule learners showed the effect in the above described direction. The effect was less pronounced for the former, though; rule learners: F(1,26) = 44.08, p < .001, η_p^2 = .63, non-rule learners: F(1,52) = 6.50, p = .014, η_p^2 = .11. Thus, like in previous experiments, both learner types classified paired compounds correctly, while generalization compounds were only categorized correctly by rule learners. The main effect of learner type was not significant in the overall ANOVA, F(1,78) < 0.01, p = .939, η_p^2 < .01.

Evaluation. The ANOVA showed only a main effect of correct category. Positively paired compounds were evaluated as more positive (M = 5.88, SD = 2.65) than negatively paired compounds (M = 4.36, SD = 2.47), F(1,78) = 12.96, p < .001, η_p^2 = .14. This constitutes a standard EC effect. Generalization compounds, in contrast, show the opposite pattern. No other effects were significant, all Fs < 1.98, all ps > .16, all η_p^2s < .03.



Categorization of paired compounds

Liking of paired compounds

Figure A3. Data from Experiment 5.2b: The left panel shows the proportion of ”mammal” categorization responses towards the paired compound stimuli, grouped by the paired category as a function of learner type. Thus, high values for the mammal/positive category and low values for the reptile/negative category indicate good categorization performance. The right panel shows the evaluation of the paired compound stimuli as a function of the paired category and learner type. Error bars show the standard error of the mean.

Paired versus generalization compounds. We analyzed rule learners' categorization responses of paired and generalization stimuli and observed a main effect of correct category. "Mammals" were categorized as mammals more often (*M* = 0.70, *SD* = 0.46) than "reptiles" (*M* = 0.20, *SD* = 0.40), *F*(1,26) = 21.67, *p* < .001, η_p^2 = .45. There was also an interaction with stimulus type, *F*(1,26) = 9.06, *p* = .005, η_p^2 = .26. Rule learners classified generalization stimuli correctly (reported in the paper) and, as a follow-up ANOVA showed, also paired compounds were classified correctly, ("mammals": *M* = 0.82, *SD* = 0.38; "reptiles": *M* = 0.17, *SD* = 0.38), *F*(1,26) = 44.08, *p* < .001, η_p^2 = .63. The main effect of stimulus type was not significant, *F*(1,26) = 3.32, *p* = .080, η_p^2 = .11.

The analysis for rule learners' liking of paired and generalization compounds showed only an interaction between correct category and stimulus type, *F*(1,26) = 12.54, *p* = .002, η_p^2 = .33. A follow-up ANOVA for paired stimuli showed that rule learners evaluated positively paired compounds more positive (*M* = 6.59, *SD* = 2.58) than negatively paired compounds (*M* = 4.19, *SD* = 2.72), *F*(1,26) = 7.64, *p* = .010, η_p^2 = .23. This standard EC effect we observe for paired compounds is "in line with the rule". For rule learners' liking of generalization stimuli, however, it was reversed (reported in the paper). The main effects in the overall ANOVA were not significant, all *F*s < 1.86, all *p*s > .18, all η_p^2s < .07.

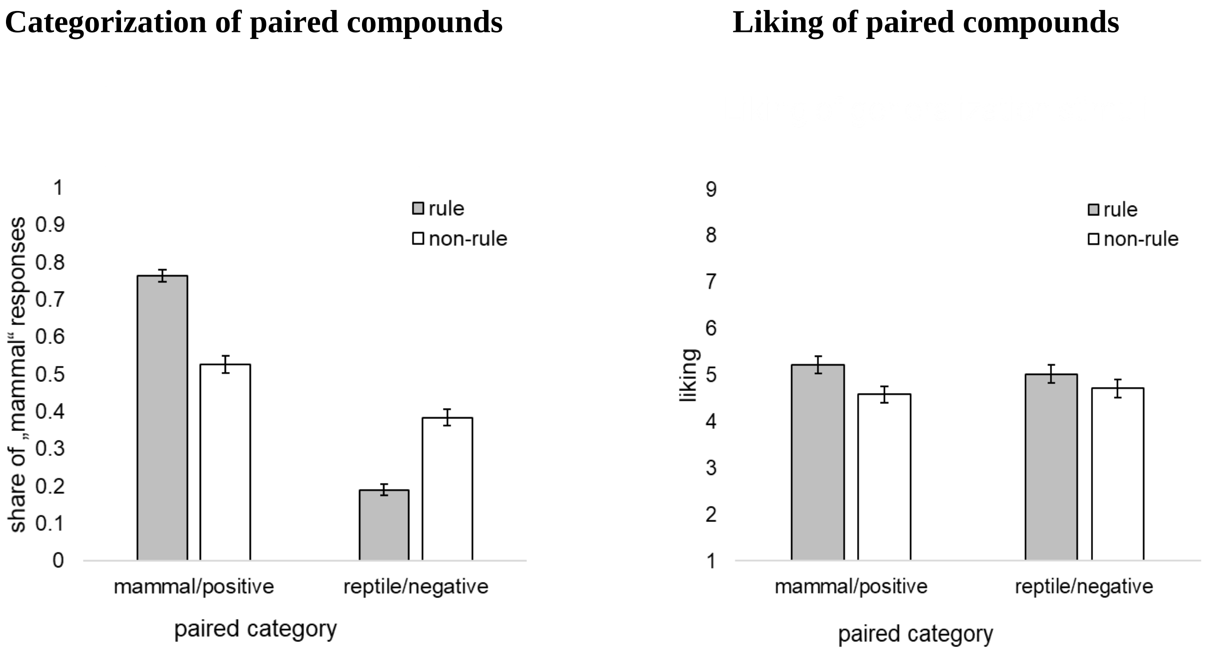
Thus, the observed dissociation for rule learners' category and liking judgments is specific to generalization compounds because paired compounds show the same pattern (that follows the rule) for both measures.

Experiment 5.3.

Categorization. The ANOVA showed a category main effect. Mammal-paired compounds were categorized as mammals more often (*M* = 0.66, *SD* = 0.47) than reptile-paired ones (*M* = 0.27, *SD* = 0.45), *F*(1,116) = 85.76, *p* < .001, η_p^2 = .43. Again, there was an interaction between category and learner type, *F*(1,116) = 31.19, *p* < .001, η_p^2 = .21. Follow-up ANOVAs showed that both types of learner showed the above described pattern (rule learners: mammal-paired: *M* = 0.76, *SD* = 0.43, reptile-paired: *M* = 0.19, *SD* = 0.39, *F*(1,67) = 142.83 *p* < .001, η_p^2 = .68; non-rule learners: mammal-paired: *M* = 0.53, *SD* = 0.50, reptile-paired: *M* = 0.38, *SD* = 0.49, *F*(1,49) = 5.22 *p* = .027, η_p^2 = .10). Like in Experiment 5.1, 5.2a and 5.2b, both learner types classified paired compounds correctly while generalization compounds were only categorized correctly by rule learners. The main effect of learner type in the overall ANOVA was not significant, *F*(1,116) = 0.50, *p* = .482, η_p^2 < .01.

Evaluation. There was a main effect of learner type in the ANOVA: Rule learners overall evaluated the compounds more positively (M = 5.12, SD = 2.18) than non-rule learners (M = 4.65, SD = 2.30), F(1,116) = 4.07 p = .046, η_p^2 = .03. All other effects were nonsignificant, all Fs < 0.62, all ps > .43, all η_p^2s < .01.

We explain the absence of an EC effect like in Experiment 5.1 and 5.2a: Compound's and elements' pairings cancelling each other out upon evaluation of the compounds.



Categorization of paired compounds

Liking of paired compounds

Figure A4. Data from Experiment 5.3: The left panel shows the proportion of "mammal" categorization responses towards the paired compound stimuli, grouped by the paired category as a function of learner type. Thus, high values for the mammal/positive category and low values for the reptile/negative category indicate good categorization performance. The right panel shows the evaluation of the paired compound stimuli as a function of the paired category and learner type. Error bars show the standard error of the mean.

Paired versus generalization compounds. The ANOVA for rule learners' categorization of paired and generalization compounds showed a main effect of correct category. "Mammal" compounds were categorized as mammals more often (*M* = 0.76, *SD* = 0.43) than "reptile" compounds (*M* = 0.21, *SD* = 0.41), *F*(1,67) = 137.09, *p* < .001, η_p^2 = .67. The main effect of stimulus type was not significant (*F*(1,67) = 0.48, *p* = .490, η_p^2 < .01) and, more importantly, the interaction was also not significant, *F*(1,67) = 2.34, *p* = .131, η_p^2 = .03.

The same analysis for liking yielded only a nonsignificant main effect of stimulus type. Rule learners tended to evaluate paired compounds more positively (*M* = 5.12, *SD* = 2.18) than generalization compounds (*M* = 4.81, *SD* = 2.19), *F*(1,67) = 3.00, *p* = .088, η_p^2 = .04. No other effects were significant, all *F*s < 0.92, all *p*s > .34, all η_p^2s < .02.

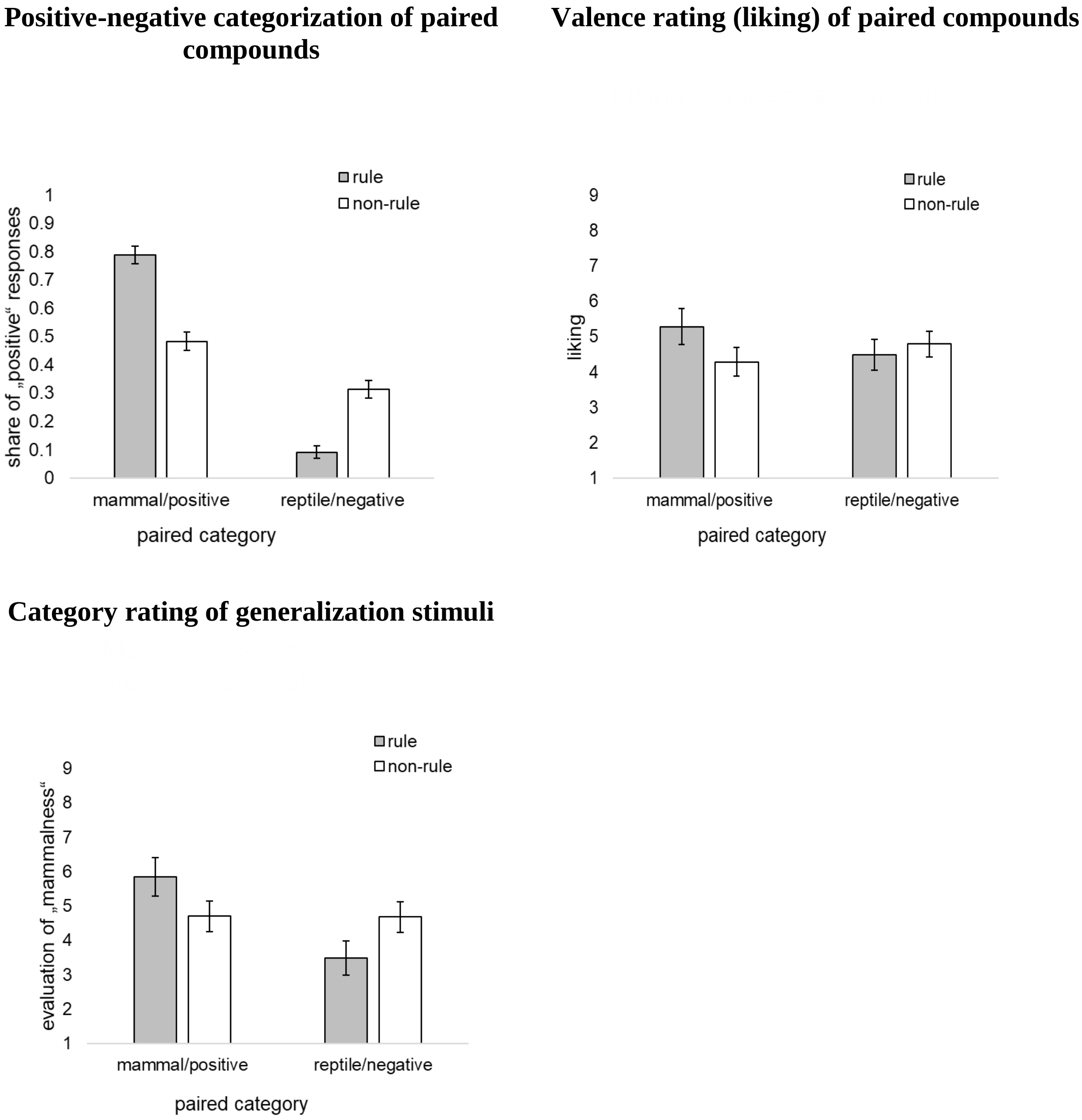
Although the findings are less clear-cut than in Experiments 5.2a and 5.2b, we can conclude the following: Jointly analyzing rule learners' categorization of paired and generalization stimuli yields a main effect of category and no interaction which shows that both types of stimuli were categorized in line with the rule. For liking, in contrast, we do not even observe a main effect of category which is present, however, when analyzing generalization stimuli alone (reported in the paper). This might suggest that responses for paired stimuli do not follow the same pattern as responses for generalization stimuli.

Experiment 5.4.

Positive-negative categorization. We observed a main effect of category according to the rule in the ANOVA. Positively paired stimuli were classified as positive more often (*M* = 0.61, *SD* = 0.49) than negatively stimuli (*M* = 0.22, *SD* = 0.41), *F*(1,77) = 52.99, *p* < .001, η_p^2 = .41. Further, there was an interaction between category and learner type, *F*(1,77) = 19.63, *p* < .001, η_p^2 = .20. Follow-up ANOVAs showed that rule learners correctly categorized the stimuli, non-rule learners descriptively also categorized correctly but this analysis was not significant (rule learners: positive: *M* = 0.79, *SD* = 0.41, negative: *M* = 0.09, *SD* = 0.29, *F*(1,32) = 101.58, $p < .001$, $\eta_p^2 = .76$; non-rule learners: positive: *M* = 0.48, *SD* = 0.50, negative: *M* = 0.31, *SD* = 0.46, *F*(1,45) = 3.74, *p* = .059, $\eta_p^2 = .08$). The pattern shows that, for paired compounds, both types of learner categorized rather correctly. For generalization compounds, in contrast, none of the both learner types (rule learners at most) categorized correctly. The main effect of learner type was not significant in the overall ANOVA, *F*(1,77) = 0.64, *p* = .427, $\eta_p^2 < .01$.

Category rating. The ANOVA showed a category main effect: Participants rated mammal-paired compounds as more mammal-like (*M* = 5.18, *SD* = 3.14), than reptile-paired compounds (*M* = 4.18, *SD* = 2.96), *F*(1,77) = 5.14, *p* = .026, $\eta_p^2 = .06$. There was also an interaction between ategory and learner type, *F*(1,77) = 4.96, *p* = .029, $\eta_p^2 = .06$. Follow-up ANOVAs showed that rule learners rated the compounds in line with their paired category (mammal-paired: *M* = 5.84, *SD* = 3.18, reptile-paired: *M* = 3.48, *SD* = 2.83), *F*(1,32) = 8.27 *p* = .007, $\eta_p^2 = .21$. Non-rule learners did not show an effect, *F*(1,45) = 10.26 *p* = .974, $\eta_p^2 < .01$. This shows that participants category ratings of paired compounds rather followed the compounds' paired categories (and thus, the rule) than their elements' paired category. That is, if it was an objective task they would have performed mostly "correctly". Their judgments of generalization stimuli, in contrast, were "incorrect", that is, not in line with the rule. The main effect of learner type was not significant in the overall ANOVA, *F*(1,77) < 0.01, *p* = .968, $\eta_p^2 < .01$.

Valence rating. We found no significant effects in the ANOVA, all *F*s < 2.23, all *p*s > .13, all η_p^2s < .03. Like Experiment 5.1, 5.2a and 5.3, we believe, when judging a paired compound, the conflicting influence of the compound's pairing and its elements' pairings cancel each other out and lead to the observed null effect.



Positive-negative categorization of paired compounds

Valence rating (liking) of paired compounds

Category rating of generalization stimuli

Figure A5. Data from Experiment 5.4: The left panel shows positive-negative categorization (i.e., proportion of "positive" responses) towards the paired compound stimuli, grouped by the paired category as a function of learner type. High values for the mammal/positive category and low values for the reptile/negative category indicate good categorization performance. The middle panel shows the liking ratings and the right panel shows ratings of category membership of the paired compound stimuli as a function of the paired category and learner type. Error bars show the standard error of the mean.

Paired versus generalization compounds. The ANOVA analyzing rule learners' positive-negative categorization for paired and generalization compounds showed a main effect of correct category. "Positive" compounds (i.e., among paired compounds, those that were paired with the positive category in the learning phase, and among generalization stimuli, those that would rule-wise belong to the positive category) were categorized as positive more often (*M* = 0.65, *SD* = 0.48) than "negative" compounds (*M* = 0.22, *SD* = 0.42), *F*(1,32) = 26.90, *p* < .001, η_p^2 = .46. There was also an interaction between category and stimulus type, *F*(1,32) = 22.14, *p* < .001, η_p^2 = .41. A follow-up ANOVA for paired stimuli showed that rule learners categorized them correctly ("positive": *M* = 0.79, *SD* = 0.41, "negative": *M* = 0.09, *SD* = 0.29, *F*(1,32) = 101.58, *p* < .001, η_p^2 = .76) just like generalization stimuli (reported in the paper). The main effect of stimulus type in the overall ANOVA was not significant, *F*(1,32) < 0.01, *p* = .942, η_p^2 < .01.

The same analysis for category ratings showed only an interaction between correct category and stimulus type, *F*(1,32) = 13.01, *p* = .001, η_p^2 = .29. A follow-up ANOVA for paired stimuli showed that rule learners evaluated mammal-paired compounds as more mammal-like (*M* = 5.85, *SD* = 3.18) than reptile-paired compounds (*M* = 3.48, *SD* = 2.83), *F*(1,32) = 8.27, *p* = .007, η_p^2 = .21. That is the opposite pattern as for generalization stimuli (reported in the paper). The other effects in the overall ANOVA were not significant, all *F*s < 0.79, all *p*s > .38, all η_p^2s < .03.

The ANOVA for valence ratings also only showed an interaction, *F*(1,32) = 4.71, *p* = .037, η_p^2 = .13. For paired compounds, rule learners descriptively evaluated "in line with the rule" (positively-paired: *M* = 5.27, *SD* = 2.91, negatively-paired: *M* = 4.48, *SD* = 2.48); this effect was not significant, however, *F*(1,32) = 1.06, *p* = .311, η_p^2 = .03. Generalization stimuli, in contrast, are descriptively evaluated "contrary to the rule" (reported in the paper). All other effects in the overall ANOVA: all *F*s < 0.35, all *p*s > .56, all η_p^2s < .02.

The clear dissociation between positive-negative categorization and category rating that we observed for generalization stimuli among rule learners is not present for paired compounds. Instead, rule learners' responses on both measures are in line with the rule. The pattern for valence ratings is less clear-cut but the observed interaction shows that rule learners responded differently towards paired than towards generalization stimuli.

Conclusions. Comparing categorization performance for paired versus generalization compounds in Experiment 5.1-3 shows that they substantially differ: While correct categorization of paired compounds is achieved both by rule and non-rule learners, only rule learners manage to correctly categorize novel generalization compounds. Positive-negative categorization and category ratings in Experiment 5.4 suggest very similar conclusions.

Comparing liking judgments for paired and generalization compounds in Experiment 5.1-4 also shows that they follow different patterns: Judgments of paired compounds are informed both by the compound's pairing and its elements' pairings in the learning phase which cancel each other out. Thus, there was no effect of paired category for paired compounds in most experiments (except for Experiment 5.2b). Judgments of generalization compounds, in contrast, are only informed by the elements' pairings since the generalization compounds were not paired in the learning phase. Thus, liking of generalization compounds reflects their elements' pairing in the learning phase.