Does association formation contribute to evaluative conditioning? A review of current methodologies and their findings

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

nach der Prüfungsordnung vom 18. Dezember 2018

vorgelegt von

Karoline Corinna Bading

aus

Hamburg

Tag der Abgabe: 12. Juli 2021

Erklärung über Eigenanteil

Chapter 2 basiert auf dem Manuskript:

Stahl, C., & Bading, K.C. (2020). Evaluative Conditioning of Pattern-Masked Nonwords Requires Perceptual Awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 822-850.

Die zugrundeliegende Idee sowie die theoretische Herleitung stammen von Christoph Stahl. Ich habe zur theoretischen Herleitung beigetragen. Das Studiendesign wurde von Christoph Stahl und mir entwickelt. Die Datenerhebung wurde von mir umgesetzt. Die Datenauswertung wurde von Christoph Stahl und mir durchgeführt. Das Manuskript wurde von Christoph Stahl und mir geschrieben.

Chapter 3 basiert auf dem Manuskript:

Bading, K., Stahl, C., & Rothermund, K. (2020). Why a standard IAT effect cannot provide evidence for association formation: The role of similarity construction. *Cognition & Emotion*, *34*(1), 128-143. doi: 10.1080/02699931.2019.1604322

Die zugrundeliegende Idee sowie die theoretische Herleitung stammen von mir. Das Studiendesign wurde von mir in Zusammenarbeit mit Klaus Rothermund und Christoph Stahl entwickelt. Die Datenerhebung und -auswertung wurden von mir umgesetzt. Das Manuskript wurde von mir in Zusammenarbeit mit Klaus Rothermund geschrieben. Christoph Stahl hat an der Manuskriptgestaltung mitgewirkt.

Chapter 4 basiert auf dem Manuskript:

Bading, K.C., Barth, M., Rothermund, K., Aust, F., & Stahl, C. (unpublished manuscript). Does mere co-occurrence affect evaluation? MPT modelling of relational evaluative conditioning cannot (yet) tell.

Die zugrundeliegende Idee sowie das Studiendesign stammen von mir. Die Datensimulation und -auswertung wurden von mir in enger Zusammenarbeit mit Marius Barth umgesetzt. Das Manuskript wurde von mir geschrieben. Marius Barth, Klaus Rothermund, Frederik Aust und Christoph Stahl haben an der Manuskriptgestaltung mitgewirkt.

Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im November 2021 angenommen.

Contents

Preface 11

1 Evaluative Conditioning: Theoretical and Methodological Background 13

2 Evaluative Conditioning of Masked Nonwords Requires Perceptual Awareness 33

3 Why a Standard IAT Effect Cannot Provide Evidence for Association Formation: The Role of Similarity Construction 77

4 Does Mere Co-occurrence Affect Evaluation? MPT Modelling of Relational Evaluative Conditioning Cannot (Yet) Tell 97

5 General Discussion 125

References 133

Appendices 145

Abstract

In this dissertation, I address the question of whether association formation contributes to evaluative conditioning (EC). EC refers to a change in the evaluation of a stimulus (CS) that is due to its co-occurrence with another, valenced stimulus (US). EC can be explained from two distinct theoretical perspectives: according to the propositional approach, EC is mediated by the non-automatic formation of propositions about the relation between the co-occurring stimuli; whereas, according to the associative approach, EC is mediated by the automatic formation of associations between the mental representations of the co-occurring stimuli. While the contribution of proposition formation to EC is supported by unique evidence and therefore uncontroversial, a potential contribution of association formation is still hotly debated. In this dissertation, I try to contribute to this theoretical discussion in three critical lines of research that have the potential to provide unique empirical support for association formation in EC. In a first line of research, I tried to gather unique evidence for association formation in EC by investigating whether EC can emerge in the absence of awareness for the CS-US co-occurrence (a necessary condition for the formation of propositions but not of associations). In a second line of research, I investigated whether a recently reported dissociation between directly vs. indirectly measured CS evaluations (seemingly providing unique support for association formation) can be equally explained from a purely propositional perspective. Finally, in a third line of research, I demonstrated that a recently introduced methodology for studying association formation in EC is severely confounded with a propositional alternative explanation, and can therefore not provide unambiguous evidence for association formation (at least, not in its current state). Taken together, this dissertation therefore shows that unique empirical support for association formation in EC is still lacking, and that novel methodological approaches are needed for such support to ever be found.

Zusammenfassung

In dieser Dissertation geht es um die Frage, ob Assoziationsbildung zu evaluativer Konditionierung (EC) beiträgt. EC bezeichnet eine Veränderung in der Bewertung eines Stimulus' (CS), welche durch sein gemeinsames Auftreten mit einem anderen, valenten Stimulus (US) verursacht wird. EC kann aus zwei verschiedenen theoretischen Perspektiven erklärt werden. Laut dem propositionalen Ansatz wird EC durch die non-automatische Bildung von Propositionen über die Beziehung der gemeinsam auftretenden Stimuli mediiert. Dagegen besagt der assoziative Ansatz, dass EC durch die automatische Bildung von Assoziationen zwischen den mentalen Repräsentationen der gleichzeitig auftretenden Stimuli vermittelt wird. Während die Beteiligung propositionaler Prozesse an EC durch eindeutige Evidenz gestützt und daher unstrittig ist, wird eine zusätzliche Beteiligung von assoziativen Prozessen weiterhin heiß diskutiert. In dieser Dissertation versuche ich mithilfe von drei Forschungsbeiträgen zu dieser theoretischen Diskussion beizutragen. Im ersten Forschungsbeitrag versuchte ich, eindeutige Evidenz für Assoziationsbildung zu erbringen, indem ich untersucht habe, ob EC in der Abwesenheit von Bewusstsein für die CS-US Paarungen entstehen kann. In einem zweiten Forschungsbeitrag habe ich untersucht, ob eine kürzlich berichtete Dissoziation zwischen direkt vs. indirekt gemessenen CS Evaluationen (die scheinbar Assoziationsbildung nachweist) auch aus einer rein propositionalen Perspektive erklärt werden kann. Im dritten Forschungsbeitrag konnte ich nachweisen, dass eine kürzlich eingeführte Methode zur Untersuchung von Assoziationsbildung in EC durch ein propositionales Artefakt verfälscht ist und daher keine eindeutige Evidenz für Assoziationsbildung erbringen kann (jedenfalls nicht in seiner aktuellen Form). Insgesamt betrachtet zeigt diese Dissertation, dass es immer noch keine eindeutige Evidenz für Assoziationsbildung in EC gibt, und dass neue methodologische Ansätze nötig sein werden, um diese jemals zu erbringen.

Preface

Evaluative conditioning refers to a well established learning phenomenon in which the evaluation of a stimulus is changed through its mere co-occurrence with other, valenced stimuli (De Houwer, 2007). As such, this evaluative learning effect combines two cornerstones of everyday life — co-occurring stimuli and evaluations — and may therefore represent a key concept in explaining human thought and behavior. Due to its potentially immense importance, evaluative conditioning has been studied extensively and with an overwhelming focus on its underlying mental processes. One of these potential mediators is association formation and is assumed to register mere stimulus co-occurrences in order to translate them into simple memory links. These links are assumed to be completely unqualified by the exact relation between the co-occurring stimuli, and may therefore produce stimulus evaluations and downstream behaviors that are illogical by normative standards. Despite these potentially irrational effects, association formation is widely viewed as a generally beneficial mechanism allowing humans (and other species) to efficiently adapt to contingencies between environmental stimuli. As such, association formation provides an elegant explanation for various (associative) learning effects and has therefore shaped theorizing not only on evaluative conditioning, but also on Pavlovian conditioning as well as human contingency learning (for a review, see Mitchell, De Houwer, & Lovibond, 2009).

Having held sway for several decades, the association formation approach was gradually undermined by accumulating evidence showing Pavlovian conditioning to depend on the degree of statistical contingency between the co-occurring stimuli (Rescorla, 1968), on the one hand, and to require awareness for this statistical relationship (Dawson & Schell, 1985), on the other hand. Taken together, these and other findings gave rise to a novel theoretical perspective according to which Pavlovian conditioning and similar phenomena are driven entirely by higher-order reasoning (resulting in conscious knowledge about the predictive relationship between the co-occurring stimuli). After two decades of pre-eminence, this purely propositional approach was eventually reined in by novel research on the very learning effect that is the focus of this thesis. In the early 90s, evaluative conditioning was shown to be insensitive to statistical contingency and independent of awareness (e.g., Baeyens, Hermans, & Eelen, 1993), and thereby reinvigorated the debate over the role of association formation in human associative learning. Ever since, proponents of dual-process models — combining the propositional and association formation approach to learning — have relied heavily on this initial evidence for association formation in evaluative conditioning. Over time, however, this evidence turned out to be far less convincing than originally thought. For one thing, the propositional approach was easily reconciled with evaluative conditioning's insensitivity to statistical contingency by simply opening up to the possibility that associative learning may sometimes reflect non-predictive stimulus relations (that are therefore unrelated to the degree of contingency between the co-occurring stimuli). Similarly, early demonstrations of evaluative conditioning's supposed independence of awareness were shown to be based on flawed methodologies, and do therefore no longer count as convincing evidence for association formation (Gawronski & Walther, 2012).

These conceptual and methodological insights have swayed many scholars to adopt a purely propositional perspective on evaluative conditioning (and associative learning in general). However, just as many researchers still consider the possibility of association formation worthy of investigation, and are therefore engaged in an ongoing methodological discussion in order to develop more suitable approaches for addressing this far-reaching research question. By scrutinizing some of these methodologies, this thesis seeks to make a valuable contribution to this debate.

Chapter 1

Evaluative Conditioning: Theoretical and Methodological Background

What is evaluative conditioning?

As mentioned in the introductory passage, evaluative conditioning (EC) refers to a change in a stimulus' evaluation that is due to its co-occurrence with other, valenced stimuli (De Houwer, 2007). To further illustrate this definition of EC, imagine the following scenario. On a stroll through the neighborhood, you and your friend walk by a wall covered with graffiti. Your friend points to a meaningless string of letters such as "ESREN" and asks whether you like this tag. You respond by saying: "No, not really!". A few days later, you spot an identical "ESREN" tag on the lower edge of a beautifully painted rose graffito (which happens to be your favorite flower). Later that day, you meet up with your grandmother in front of a supermarket. She points to a freshly sprayed "ESREN" tag on the supermarket door and says: "I just hate graffiti!". This time, you respond by saying: "Oh really? I happen to quite like this one!". In this scenario, the change in the evaluation of "ESREN" has two possible causes. On the one hand, it may have been brought about by the joint sight of "ESREN" and the rose graffito. On the other hand, it may be due to the fact that the "ESREN" tag was repeatedly encountered. As evident from the prior definition, the change in the evaluation of "ESREN" qualifies as an instance of EC only in the first case. Given the second case, said change in evaluative behavior represents an instance of the mere exposure effect (see Bornstein, 1989).

In order to study EC as previously defined, one has to create a learning situation which ensures that an observed change in a stimulus' evaluation is in fact due to its co-occurrence with other, valenced stimuli (and not due to its mere repeated encounter). This is usually achieved by means of a simple pairing procedure in which neutral stimuli (e.g., meaningless letter strings or unknown cartoon characters) are repeatedly presented together with other, valenced stimuli. Crucially, one set of neutral stimuli is consistently paired with positive stimuli (e.g., images of flowers or recordings of pleasant music), while another set is always presented together with negative stimuli (e.g., images of spiders or recordings of human screams). Subsequently, the evaluation of the initially neutral stimuli is assessed using some measure of evaluation (usually a direct measure in the form of a rating scale). More often than not, these evaluations will (on average) indicate an assimilative effect of the stimulus co-occurrences on stimulus evaluations: that is, those (formerly) neutral stimuli that had been presented together with positive stimuli are evaluated more favorably than those that had been paired with negative stimuli. However, despite its empirical pre-dominance, such valence assimilation does not constitute a defining characteristic of EC. Under certain conditions (which will be described later on), stimulus co-occurrences can also have a contrastive effect on stimulus evaluations, and therefore produce more favorable evaluations of negatively paired stimuli than of positively paired stimuli. Though certainly not the default, such contrast effects can equally qualify as instances of EC — as long as they can be traced back to the joint presentation of two stimuli.

To reiterate, EC denotes any change in evaluative behavior towards a stimulus that is induced by its co-occurrence with other, valenced stimuli. As such, EC shares several features with Pavlovian conditioning (Pavlov, 1927), one of the most long-standing and widely studied phenomena in learning psychology. In its standard form, Pavlovian conditioning refers to changes in autonomous behavior towards a biologically neutral stimulus (called conditioned stimulus or CS) after its repeated presentation together with a biologically relevant stimulus (referred to as unconditioned stimulus or US). For example, in studies on the conditioning of an eyeblink reflex, a CS (e.g., a tone) is repeatedly followed by a US (e.g., a short pulse of air) inducing said reflex (i.e., an unconditional response or UR). Over time, the CS comes to induce a conditioned response (CR): that is, it will elicit (autonomous) behavior that is similar to the behavior induced by the US it was previously paired with (in this case, an eyeblink reflex). This conceptual framework is largely applicable to EC. In this reading, the valenced stimuli presented during the EC procedure represent USs intrinsically capable of evoking an evaluative response (i.e., a UR). In turn, the neutral stimuli in an EC procedure represent CSs that, due to their pairings with the USs, become capable of eliciting comparable evaluative responses (i.e., a CR). This conception of EC as a phenomenon similar to conditioned changes in autonomous behavior (such as salivation, sweating or the startle reflex) has shaped research on EC in at least two ways. For one thing, the vocabulary used to describe Pavlovian conditioning (in particular, the terms "CS" and "US") is also standard terminology in EC research (and will therefore be adopted in the remainder of this thesis). More importantly, however, theorizing on EC has also inherited the two most prominent approaches to explaining Pavlovian conditioning (for a review, see Mitchell et al., 2009). In short, the associative approach postulates that the behavioral effects of stimulus co-occurrence result from the automatic formation of associations between mental representations. By contrast, the propositional approach proposes that such effects are mediated by the non-automatic formation of propositions about stimulus relations. In the following section, the theoretical accounts of EC derived from these two approaches will be layed out in greater detail.

How can evaluative conditioning be explained?

Associative accounts

Associative accounts of EC incorporate one of the oldest ideas in learning psychology: namely, the notion that the mere observation of two co-occurring events produces an association between their representations in memory (e.g., Ebbinghaus, 1885). Such association formation is understood as a stimulus-driven process in that its operation requires no other input aside from the observation of the co-occurring events. This joint registration of two events will result in a co-activation of their respective mental representations which, in turn, produces and strengthens a simple memory link between them. Once established, such a link allows for a spread of activation from one representation to the other. Accordingly, the re-encounter of just one event, through the activation of its own mental representation, may bring to mind the mental representation of a previously co-occurring event.

According to associative accounts, a conditioned CS evaluation reflects a valenced mental representation that has become associated with the CS representation during the prior CS-US co-occurrence. Different associative accounts of EC differ, however, with regard to the nature of the representation that is assumed to become associated with the mental representation of the CS. According to stimulus-stimulus (S-S) models (e.g., Baeyens, Eelen, Crombez, & van den Bergh, 1992), the repeated CS-US pairings result in an association between the CS and US representations. In order to translate such a CS-US association into behavior (i.e., a CR), S-S models further assume that the US representation is also associated with its UR (which will become activated as a result of spreading activation from the CS to the US representation). To illustrate this mechanism, take the earlier example of the two graffiti. From the perspective of an S-S model, the sight of the "ESREN" tag next to the beautiful rose graffito produces a link between their mental images. The subsequent re-encounter of the "ESREN" tag will then activate the mental representation of the rose which (due to its status as your favorite flower) activates a positive evaluative response as expressed by your favorable statement about the tag (i.e., the CR). However, this CR can also be explained by another type of associative mechanism: that is, the direct linking of stimulus and response (S-R) representations. From the perspective of S-R models (e.g., Jones, Fazio, & Olson, 2009), the observation of a CS-US co-occurrence results in an association between the mental representations of the CS and the response elicited by the US. Accordingly, the positive evaluation of "ESREN" is assumed to reflect a direct spread of activation to the representation of the UR (without its prior passing through the US representation as postulated by S-S models).

Irrespective of whether they postulate S-S or S-R links, associative accounts view the formation of associations as a largely automatic mental process. Accordingly, association formation is assumed to require no awareness of its input (i.e., the registered stimulus co-occurrence) and very little (or no) attentional capacity. Moreover, the formation of associations is viewed as uncontrolled in the promoting as well as counteracting sense: that is, mental associations may arise without a goal to form them, and despite a goal to prevent them. Drawing on the earlier example once more, an association between "ESREN" and the rose graffito may therefore be formed even if the registration of their co-occurrence does not reach awareness, if one observes the two graffiti while thinking about something else and having no intention to make a connection between them, and finally, even if one tries not to store their joint sight in memory. Relatedly, the mechanism by which mental associations come to affect evaluative behavior (i.e. spreading of activation) is assumed to proceed in a similarly automatic fashion. Accordingly, mental activation may spread from "ESREN" to the rose graffito without awareness of the (co-)activated representations, in the absence of attentional resources or a goal for such co-activation, and finally, despite attempts to prevent it.

Taken together, associative accounts view EC as the product of largely automatic mental processes (association formation and spreading of activation) which thrive on passive stimulus observation, and operate without any input from other, higher-order mental processes. As such, the claims of associative accounts form a striking contrast with the assumptions of propositional accounts of EC the most important of which will be explained in the following section.

Propositional accounts

To recapitulate, associative accounts of EC assume that a CS-US co-occurrence must be merely registered in order to produce a change in CS evaluation. By contrast, propositional accounts propose that such an evaluative change requires not just a registration but also an interpretation of stimulus co-occurrence: that is, the observed co-occurrence of CS and US needs to be mentally construed in terms of a relation between them. According to propositional accounts (e.g., De Houwer, 2018), such a construction of stimulus co-occurrence results in the formation of a proposition; i.e., a mental representation holding information about the relation between (co-occurring) stimuli. As such, a proposition represents a statement about a state of affairs in the world and therefore has a truth value (i.e., it is either true or false depending on its correspondence with the de facto state of the environment). Accordingly, a proposition can be subjected to inferential reasoning: that is, it can be combined with other sources of (propositional) knowledge in order to derive and truth evaluate novel propositions. Crucially, such inferential processes allow for the generation of an evaluative proposition about the valence of a CS, which, according to current propositional accounts, drives the conditioned evaluative behavior.

From a propositional perspective, CS-US co-occurrence affects CS evaluation via evaluative inferences derived from the joint consideration of an *ad hoc* proposition about the CS-US relation and other (situationally formed or pre-existing) propositions. However, neither the exact nature of the constructed relation nor the precise content of the additionally considered propositions are (currently) specified by propositional accounts of EC. In a common illustration, the joint presentation of CS and US is assumed to produce a proposition simply stating that the two stimuli co-occur. In combination with a pre-existing belief that co-occurring stimuli tend to be similar, this situationally derived proposition allows for the (evaluative) inference that the CS valence is similar to the valence of its co-occurring US. Having said that, a propositional explanation of EC is not committed to any particular CS-US relation or additionally considered proposition. Instead it may (at least, in principle) adopt any combination of constructed relation and complementary propositions that allows for an evaluative inference about the CS valence. To illustrate this point, consider the earlier example of "ESREN" and the rose graffito. In this scenario, the spatial configuration of the two graffiti may give rise to a proposition stating that "ESREN" is the signature of the person who painted the beautiful image of your favorite flower. Furthermore, the combination of said proposition and additionally held beliefs (e.g., that signatures of cool artists are themselves cool) may give rise to a positive evaluation of the "ESREN" tag. Though not strictly logical, such an explanation of the change in evaluative behavior towards "ESREN" is in full compliance with the core assumptions of propositional accounts of EC: that is, it involves an evaluative inference about the valence of the CS that is derived from a proposition about its relation with the co-occurring US.

To recapitulate, propositional accounts explain EC as the joint product of constructive and inferential processes. The operation of these higher-order processes is widely assumed to require a conscious input (e.g., the co-occurring stimuli or the considered propositions), to depend on attentional capacity, and to be controlled (or controllable) by promoting and counteracting goals. However, the processes proposed by propositional accounts are also known to operate somewhat automatically under certain circumstances (De Houwer, 2018). For example, in line with other forms of problem solving, the construction of CS-US relations may sometimes proceed with only minimal mental resources, but should always require awareness of the CS-US co-occurrence and a promoting goal to construe such a relation. Similarly, (evaluative) inferences can sometimes be drawn efficiently as well as uncontrolled in the promoting and counteracting sense (but not without awareness for the involved propositions). Based on their varying degrees of (non-)automaticity, propositional processes are capable of producing EC effects under a variety of processing conditions. For example, the (consciously perceived) sight of "ESREN" next to the rose graffiti may spontaneously trigger its construction as the signature of the rose-painting artist and, subsequently, a quick-and-dirty inference about its valence. However, it is of course equally possible for such an inference to arise in a relatively non-automatic manner, e.g., if one spends the rest of the second walk deliberately contemplating the joint sight of the two graffiti.

In summary, propositional accounts of EC draw on mental processes that are decidedly different from those proposed by associative accounts. While (evaluative) propositions result from an often non-automatic integration of propositional input from various channels of information, associations between mental representations are produced by a non-cognitive and automatic process relying on registered CS-US co-occurrences as its sole informational source. Based on these differences in cognitive penetrability and automaticity, associative and propositional accounts make a number of predictions some of which can be used to demonstrate the (potential) contributions of association and proposition formation to EC. In the following section, the most prominent of these predictions as well as the results of (some of) their empirical examinations will be presented.

What are the unique implications of associative vs. propositional accounts and how are they supported by the literature?

Unique predictions based on cognitive (im-)penetrability

As explained in the previous section, propositional accounts view EC as the product of an evaluative inference that is derived from a proposition about the CS-US relation. Importantly, the nature of this relation between CS and US is not an inherent implication of their co-occurrence. Instead, it represents a mental construction achieved by flexible processes that are receptive to all kinds of propositional input — including information on the CS-US relation itself. As an illustration, consider once again the earlier example of the two graffiti. As previously described, the location of "ESREN" at the lower edge of the rose graffito is likely to induce an interpretation of "ESREN" as the signature of the artist who painted the beautiful rose. However, such a construction of the CS-US relation is by no means imperative, and may be altered even by minor changes in the positioning of the two graffiti. For example, were the "ESREN" tag to cover parts of the rose graffito, it would probably be viewed as an attempt at erasing the beautiful flower. Crucially, such an interpretation of the relation between the two graffiti should have a profound effect on the evaluative implications of their co-occurrence. While the earlier "signature" interpretation was assumed to result in a positive evaluation of "ESREN", the later "erasure" interpretation is likely to give rise to a negative view on the "ESREN" tag. Together with their flexible construction, such diverging implications of different CS-US relations allow for a first, unique prediction of propositional accounts. That is, the strength and direction of the conditioned CS evaluation should depend on information regarding the relation between the co-occurring stimuli. Associative accounts, by contrast, exclude such an influence of relational information on EC: due to its cognitively impenetrable nature, association formation cannot incorporate any input aside from the CS-US co-occurrence itself. Accordingly, neither direction nor strength of the conditioned CS evaluation should depend on additionally provided information concerning the CS-US relation.

This first set of unique predictions is usually tested by means of a modified EC procedure complementing the CS-US pairings with a manipulation of the perceived CS-US relation. In such relational EC procedures, some CS-US pairings are combined with an assimilative relation suggesting a convergence in the valences of CS and US (e.g., when the CS causes or starts the US), while others are accompanied by a contrastive relation implying an opposition between the two valences (e.g., when the CS prevents or stops the US). On direct measures of evaluation, such assimilative vs. contrastive CS-US relations have been shown to yield strong and highly consistent effects on the direction of the conditioned CS evaluation (e.g., Fiedler & Unkelbach, 2011). That is, while CS-US pairings combined with assimilative relations produce regular assimilative EC effects (i.e., more favorable evaluations of positively over negatively paired CSs), CS-US pairings accompanied by contrastive relations result in contrastive EC effects (i.e., a reversed preference for negatively over positively paired CSs). As previously explained, such an influence of relational information on EC confirms an exclusive prediction of propositional accounts, and cannot be explained by associative models. Accordingly, the aforementioned effects unambiguously demonstrate the contribution of proposition formation to the emergence of EC, and thus rule out purely associative explanations of said phenomenon. This demonstrable involvement of propositional processes cannot, however, preclude an additional contribution of association formation, and therefore leaves room for theoretical accounts that view EC as the joint product of associations and propositions (e.g., Gawronski & Bodenhausen, 2018). Though less parsimonious than purely propositional (or associative) explanations, such dual-process models may still be necessary to account for the potential results of testing other sets of competing predictions derived from propositional and associative accounts of EC.

Unique predictions based on (non-)automaticity

As explained earlier, associative and propositional processes do not only incorporate different types of information, they also require distinct mental conditions in order to operate properly. A first such distinction concerns the role of awareness for the CS-US co-occurrences. While such CS-US awareness is deemed indispensable for the construction of a CS-US relation, associations between co-activated representations are assumed to arise irrespective of whether their underlying CS-US co-occurrence is consciously perceived. This differential reliance on CS-US awareness allows for a second set of unique predictions implied by propositional vs. associative accounts. That is, while an EC effect based on (evaluative) propositions can only arise when the co-occurring stimuli have been consciously perceived, an EC effect based on mental associations may emerge even when the underlying CS-US co-occurrences have not reached awareness.

This prediction set has often been tested by means of a correlational approach examining the relationship between the conditioned CS evaluations and recollective memory for the US valences. In this line of research, EC without CS-US awareness (and, by implication, an additional contribution of association formation to EC) is assumed to be evidenced by significant EC effects in the absence of recollective memory for the US valences. The most convincing demonstration of this supposedly crucial pattern has so far been provided by Hütter, Sweldens, Stahl, Unkelbach, and Klauer (2012). Using a process dissociation procedure, these authors showed evaluative CS classifications to be biased towards the US valences even when these valences could not be recollected. Such demonstrations of EC without US valence memory indicate that the expression of a conditioned CS evaluation does not require concurrent awareness of its underlying CS-US co-occurrence, and are therefore well in line with associative accounts.

However, expressing a conditioned CS evaluation in the absence of CS-US awareness is not the same as acquiring this evaluation without having been aware of the co-occurring stimuli. After all, a consciously perceived CS-US co-occurrence may give rise to an evaluative inference about the CS, but be forgotten before this inference is expressed. As an illustration, consider once again the example of the two graffiti. The sight of "ESREN" next to the beautiful rose may be immediately translated into the previously described signature interpretation and, subsequently, into a positive evaluation of the "ESREN" tag. After its brief contemplation, this evaluative inference may be stored in long-term memory from where it is retrieved when your grandmother complains about a freshly sprayed "ESREN" tag on the supermarket door. Crucially, such a direct retrieval of a previously inferred CS evaluation may or may not be accompanied by a recollection of its underlying CS-US co-occurrence. To the extent that the joint sight of the two graffiti has been forgotten (or was never stored to begin with), a positive evaluation of the "ESREN" tag may therefore be expressed without recollecting the positive valence of the beautiful rose graffiti. As illustrated by this example, EC without US valence memory is equally well explained by propositional accounts, and can thus not demonstrate an additional contribution of association formation to EC. In order to provide unique support for associative accounts, the competing predictions concerning the role of CS-US awareness will have to be tested in other, more informative ways.

Though less absolutely so than in their differential dependence on CS-US awareness, associative and propositional processes also differ with regard to their demand for attentional capacity. While association formation is assumed to require little to no attentional resources, the constructive and inferential processes involved in proposition formation are known to depend on attentional capacity in many (though not all) situations. This difference in resource-dependence allows for another unique prediction implied by propositional accounts. That is, an EC effect based on (evaluative) propositions may be partially or completely eliminated by (sufficient) reductions in attentional capacity during the operation of its underlying processes. Associative accounts, by contrast, cannot account for such resource-dependent EC. Due to association formation's efficient nature, EC effects based on mental associations should be unaffected by the amount of attentional capacity available during the observation of the CS-US co-occurrences.

The possibility of resource-dependent EC (implied by propositional accounts) is usually tested by means of an experimental approach in which attentional capacity during the presentation of the CS-US pairings is taxed through an additional task (e.g., rehearsal of irrelevant information). In this line of research, resource-dependent EC is assumed to be demonstrated when "taxed" CS-US pairings (observed under attentional load) produce significantly smaller EC effects than CS-US pairings observed under regular "un-taxed" conditions. Providing yet more support for the contribution of propositional processes, this crucial pattern has been found repeatedly and across different manipulations of attentional capacity (e.g., Dedonder, Corneille, Yzerbyt, & Kuppens, 2010; Kattner, 2012). That being said, a number of studies have also reported significant EC effects based on "taxed" CS-US pairings (e.g., Walther, 2002, Experiment 5). At first sight, such instances of EC under attentional load may seem to provide unique support for associative processes that continue to operate despite substantial reductions in attentional capacity. Though well explained by an efficient formation of mental associations, EC under attentional load does, however, not confirm a unique prediction of associative accounts. As previously explained, the processes proposed by propositional accounts are known to be rather flexible in their demand for attentional capacity. Accordingly, the few reports of seemingly resource-independent EC effects may simply reflect a more economical operation of propositional processes, and can therefore not count as evidence for an additional contribution of association formation to EC. In order to provide unique support for associative accounts, future demonstration of EC under attentional load will have to be accompanied by additional evidence excluding the operation of propositional processes during the observation of the "taxed" CS-US co-occurrences.

A third set of predictions is based on the differential goal-dependence of association and proposition formation. As an uncontrolled process, association formation is assumed to operate independently of any promoting goal (e.g., to form an association or to generate a CS evaluation). By contrast, the formation of propositions involves goal-dependent constructive processes (that seem to always require the intention to construe a CS-US relation) as well as somewhat goal-independent inferential processes (that may operate with or without the intention to draw an evaluative inference about the CS). This differential demand for a processing goal of the promoting kind allows for another pair of competing predictions. While an EC effect based on mental associations may arise in the absence of such a processing goal, an EC effect based on (evaluative) propositions can only emerge after a goal to construe a CS-US relation has been activated.

The possibility of "unpromoted" EC effects (implied by associative accounts) has been tested by means of a modified learning procedure in which the intention to construe a relation between co-occurring stimuli is undermined in several ways (Olson & Fazio, 2001). Firstly, the CS-US pairings were never mentioned in the instructions leading up to the learning procedure. Secondly, this learning procedure featured only a relatively small number of CS-US pairings which were embedded in a stimulus stream presenting several hundreds of paired and un-paired images. Thirdly, the learning procedure included a detection task in which the stimulus stream had to be surveilled in order to detect occasional presentations of a target stimulus (that was unrelated to the CS-US pairings). Using this so-called surveillance procedure, several studies have reported significant EC effects in the (supposed) absence of a goal to construe a CS-US relation (e.g., Olson & Fazio, 2001). Though uniquely predicted by associative accounts, such demonstrations of unpromoted EC need to be interpreted cautiously. For one thing, neither a lack of reference to the CS-US pairings (during the instructions) nor attempts to distract from their presentation (during the learning procedure) can ultimately prevent a spontaneously formed goal to construe a CS-US relation. Moreover, even if dormant during the learning procedure, such a processing goal might still be activated by the subsequent assessment of the CS evaluations and (given recollective memory for the US valences) elicit *impromptu* propositions about CS-US relations as well as CS valence. Accordingly, in order to provide conclusive evidence for an additional contribution of association formation to EC, unpromoted EC (via the surveillance procedure or similar setups) will have to be demonstrated in a way that excludes the possibility of un-instructed yet otherwise triggered processing goals.

Last but not least, associative and propositional processes differ not only in their demand for a processing goal of the promoting kind, but also with regard to their receptivity to processing goals that somehow counteract their operation. While association formation is assumed to proceed irrespective of any such counteracting goal, the constructive and inferential operations involved in proposition formation may be intentionally refrained from in many (though not all) situations. This differential receptivity to counteracting goals allows for another unique prediction implied by propositional accounts. That is, an EC effect based on (evaluative) propositions may be eliminated by the goal to prevent an evaluative influence of the observed CS-US co-occurrences. Associative accounts, by contrast, cannot account for such "prevented" EC. Due to the uncontrolled nature of associative processes, EC effects based on mental associations should be unaffected by the goal to shield one's CS evaluations from the influence of the CS-US co-occurrences.

The possibility of "prevented" EC was tested by Gawronski, Balas, and Creighton (2014) who instructed participants to memorize the upcoming CS-US pairings, on the one hand, but to prevent their influence on CS evaluations, on the other hand. As predicted by propositional accounts, instructing such a counteracting goal resulted in non-significant EC effects when measured on an evaluative rating scale. This absence of EC on directly measured CS evaluations was, however, accompanied by significant evaluative priming effects reflecting the to-be-prevented influence of the previously presented CS-US co-occurrences. At first sight, such "unprevented" EC effects on indirectly measured CS evaluations may seem to demonstrate an additional contribution of associative processes (whose evaluative influence may be difficult to prevent under the processing conditions induced by indirect measurement procedures). Though well in line with associative accounts, unprevented EC on indirect measures of CS evaluation can, however, not provide unique support for association formation in EC. First of all, the response compatibility effects assessed by indirect evaluation measures are not necessarily driven by evaluative responses towards the CSs as such. Instead, they may reflect any (evaluative) information that is retrieved by a CS presentation and is capable of triggering one of the response options provided by the (indirect) measurement procedure. Hence, the evaluative priming effects reported by Gawronski et al. (2014) may simply reflect recollective memory for the to-be-memorized CS-US co-occurrences that were, however, not translated into proper CS evaluations (Stahl & Aust, 2018). Secondly, and more principally, unprevented EC effects on indirect evaluation measures may not confirm a unique prediction of associative accounts even when the response compatibility effects reflect evaluative responses towards the CSs themselves. As previously described, the inferential processes by which (evaluative) propositions are derived can be uncontrolled in the promoting as well as counteracting sense in certain situations. Accordingly, unprevented EC effects on indirectly measured CS evaluations may also be based on spontaneous inferences about the CS valences that are derived in an uncontrolled manner during the indirect measurement procedure (De Houwer, 2018). Taken together, future demonstrations of unprevented EC will therefore provide unique support for associative processes only to the extent that they can be shown to reflect proper CS evaluations, on the one hand, but not be driven by uncontrolled evaluative inferences, on the other hand.

To summarize the previous sections, associative and propositional contributions to EC can be distinguished by testing a number of competing predictions concerning the role of relational information, CS-US awareness, attentional capacity and processing goals of the promoting as well as counteracting kind. Furthermore, most unique predictions implied by propositional accounts have received strong empirical support demonstrating unequivocally the involvement of propositional processes in the emergence of EC. By contrast, an additional contribution of associative processes has not yet been convincingly demonstrated. So far, associative accounts have been empirically supported only through the confirmation of their non-unique predictions (that are shared by propositional accounts) or when their unique implications were tested in non-informative ways. Accordingly, demonstrating an associative contribution will therefore require other methodological approaches allowing for more informative tests of the unique predictions implied by associative accounts. As apparent by the previous and following sections, designing such critical tests poses a great methodological challenge, and requires careful consideration of any alternative explanation based on propositional processes. In the next section, the currently most promising methodological approaches as well as some of their hitherto unaddressed limitations will be presented.

How can an additional contribution of association formation to EC be demonstrated?

As illustrated by the previous section, demonstrating association formation in EC entails two equally important steps. Firstly, the distinct natures of association vs. proposition formation need to be translated into competing predictions of associative vs. propositional accounts. Secondly, these predictions need to be uniquely mapped onto the possible outcomes of an empirical examination of some kind. Thus far, all of the discussed approaches fail to meet at least one of the two criteria and, by implication, need to be modified in order to provide unique evidence for an additional contribution of associative processes to EC. However, some of the previously introduced approaches require methodological modifications that are currently infeasible, and do therefore not represent promising avenues for demonstrating association formation in EC. Before introducing the more promising candidates, these presently inadequate approaches as well as the reasons for their unsuitability will be presented.

Why can demonstrations of unpromoted or resource-independent EC currently not provide unique evidence for association formation?

As explained earlier, EC under attentional load does not constitute a unique prediction by associative accounts and can also be explained by an economical operation of propositional processes. Accordingly, achieving a unique mapping on an empirical outcome set requires not only a suitable manipulation of attentional load but also an additional measure demonstrating the non-operation of propositional processes under the implemented load manipulation. A similar modification is required for demonstrating association formation via EC effects in the absence of a goal to construe a CS-US relation. Though uniquely predicted by associative accounts, unpromoted EC effects cannot be reliably induced (due to the impossibility of precluding spontaneously activated processing goals) and need to therefore be accompanied by additional evidence excluding the involvement of propositional processes.

In both of these cases, the methodological challenge is two-fold. Firstly, an appropriate indicator of (inoperable) propositional processes has to be established. Secondly, this indicator needs to be assessed in a way that does not undermine the core purpose of the respective study design (i.e., to remove the attentional capacity or processing goal for constructive and inferential operations on the observed CS-US co-occurrences). Based on the current methodological repertoire, the (in-)operability of propositional processes cannot be directly assessed. Accordingly, a propositional contribution to a given instance of EC is typically not excluded by showing that propositional processes are themselves inoperable in a given learning situation. Instead, a propositional influence is usually excluded by demonstrating that its minimally required informational input (for producing an EC effect) has not been met in the particular learning context.

To this aim, reports of unpromoted EC effects (obtained in the surveillance paradigm) are typically complemented by additional evidence demonstrating a lack of recollective memory for the US valences (e.g., Olson & Fazio, 2001). In a similar vein, Mierop, Hütter, and Corneille (2017) used Hütter et al.'s process dissociation approach in order to investigate whether significant EC effects under attentional load were accompanied by US valence memory (or instead by a memory-independent evaluative bias towards the unrecollected US valences). In contrast to reports of unpromoted EC in the absence of US valence memory, Mierop et al. (2017) failed to find significant EC effects under attentional load whenever recollective memory for the US valences was absent. However, even had they done so, such a pattern could not have excluded a propositional mediation of a significant EC effect based on taxed CS-US co-occurrences (or, likewise, obtained in the surveillance paradigm). As explained earlier, expressing a conditioned CS evaluation without concurrent awareness for its underlying US valence is not the same as acquiring such an evaluation without having been aware of this crucial piece of information. Accordingly, excluding a propositional contribution to a given instance of (resource-independent or unpromoted) EC requires a demonstrable absence of awareness for the CS-US co-occurrences already during the pairing procedure (and not just at a later stage of the experiment).

Assessing awareness of the co-occurring USs (and/or their valences) during their actual presentation is, however, incompatible with the second methodological challenge for demonstrating a non-involvement of propositional processes (i.e., the non-interference with the core purpose of the study design). In the case of unpromoted EC obtained through the surveillance paradigm, an online assessment of CS-US awareness is certain to render use-less any precaution to distract from the critical CS-US pairings and, consequently, the construction of a CS-US relation. Similarly, assessing CS-US awareness during the taxed observation of the CS-US co-occurrences is likely to divert attention away from the taxing task and towards the co-occurring stimuli. Due to these inadvertent effects of excluding a propositional contribution, providing unique support for association formation via unpromoted or resource-independent EC appears rather inauspicious at the moment. Accordingly, current endeavors to demonstrate an additional contribution of associative processes to EC are focused on other, more promising approaches which will be presented in the following sections.

How can demonstrations of EC without CS-US awareness provide unique evidence for association formation?

As explained above, assessing CS-US awareness already during the pairing procedure is indispensable for excluding a propositional contribution to a given instance of EC. However, in a regular EC procedure with clearly visible stimulus presentations, such an assessment is likely to induce the very same awareness whose absence it is supposed to indicate. Accordingly, a lack of CS-US awareness during the pairing procedure can only be demonstrated when the awareness measure is applied to CS-US co-occurrences that cannot be consciously perceived.

These considerations form the basis of a second methodological approach to investigating the possibility of EC without CS-US awareness which is implied by associative accounts. In this approach, awareness for the CS-US co-occurrences (and, by implication, the operation of propositional processes) is precluded by rendering one of the two co-occurring stimuli (typically, the CS) perceptually invisible. Such a suppression of CS visibility can be achieved through several psycho-physical techniques. In the widely used subliminal presentation approach, the to-be-suppressed stimulus is shown only briefly and in close temporal succession with other stimuli masking its presentation (e.g., Stahl et al., 2016). Alternatively, conscious processing of the CSs has also been prevented by presenting them in the parafoveal area of the visual field (Dedonder, Corneille, Bertinchamps, & Yzerbyt, 2013). Finally, CS-US awareness can also be erased through continuous flash suppression creating a visual conflict between stationary CSs presented in grey colors, on the one hand, and a flashing stream of colorful USs, on the other hand (Högden, Hütter, & Unkelbach, 2018).

As demanded earlier, the effectiveness of these experimental manipulations of stimulus visibility is usually demonstrated through a separate assessment of stimulus awareness (administered either during the pairing procedure or, alternatively, under comparable conditions at a later stage of the experiment). Typically, unawareness for the to-be-suppressed CSs is demonstrated through chance performance in an identification task for which the supposedly invisible stimulus has to be selected from a list of response options (right after its presentation together with a clearly visible US). Across different materials and manipulations of stimulus visibility, a consistent result pattern has emerged: whenever participants were unaware of the CSs (and, by implication, the CS-US co-occurrences), no significant EC effects could be found.

Although such a lack of EC in the absence of CS-US awareness is well aligned with other evidence supporting propositional accounts, its evidential value is constrained by two methodological concerns. First of all, the previously mentioned techniques may suppress CS processing to an extent that undermines the co-activation of the CS and US representations and, in turn, the formation of an association between them. This concern is, however, alleviated by compelling evidence demonstrating neural processing of stimuli that are presented subliminally (Van den Bussche, Van den Noortgate, & Reynvoet, 2009) and parafoveally (e.g., Yan, Richter, Shu, & Kliegl, 2009), or rendered invisible through continuous flash suppression (Yang, Brascamp, Kang, & Blake, 2014). The second methodological concern relates to the implemented indicator of CS (un-)awareness. As an objective measure of stimulus awareness, the previously mentioned identification task is considered an exhaustive indicator of consciousness; that is, it is assumed to capture all consciously processed information that is relevant to its performance. However, CS identification performance may not necessarily represent an exclusive indicator of conscious processing. Instead, it could also be driven by unconsciously processed information evoking a sense of familiarity when the previously invisible CS is re-encountered in the identification task. While precluding conscious processing as intended, reducing CS awareness to chance-level performance in the identification task may thus undermine the very same unconscious processing that would otherwise result in CS-US associations (and, consequently, EC in the absence of awareness for the CS-US co-occurrences). In order to demonstrate EC without CS-US awareness, an experimental suppression of CS visibility may therefore be combined with other unawareness criteria that can indicate a lack of consciousness, on the one hand, but do not exclude unconscious processing, on the other hand.

Such an indication of CS unawareness can be achieved by combining the previously described identification task with a subjective measure of CS visibility in which participants are asked to report their perceptual experience of the to-be-suppressed stimuli. As exclusive indicators of conscious processing, such self-reports are assumed to be unaffected by unconsciously processed information. Accordingly, establishing CS unawareness through a subjective criterion (i.e., the perceptual experience of not having seen any of the CS features) does not require the suppression of (unconscious) neural processing which may drive not only above-chance performance in the identification task, but also the formation of associations between co-activated mental representations. Taken together, EC without CS-US awareness (and, by implication, an additional contribution of association formation to EC) may thus be demonstrated through significant EC effects based on CS presentations that were neuronally processed (as evidenced by above-chance identification performance) but not consciously perceived (according to subjective reports of stimulus visibility).

How can unqualified EC effects provide unique evidence for association formation?

As illustrated by the previous sections, current attempts at demonstrating association formation in EC are often focused on learning conditions that undermine the formation of propositions about CS-US relations and, consequently, an inference about the CS valence. Unique evidence for associative processes may, however, also come from learning procedures that do not preclude a propositional contribution, and instead allow for the simultaneous operation of associative and propositional processes. Given such a learning situation, subsequently assessed CS evaluations may reflect evaluative propositions, CS-US associations, or the joint influence of both evaluative sources. In order to separate these two (potential) influences on CS evaluation, two methodological requirements have to be met. Firstly, the implemented learning procedure needs to create conditions under which associations and propositions produce opposing effects on the conditioned CS evaluations. As previously explained, such opposing effects can be induced through relational EC procedures in which (certain) CS-US pairings are accompanied by relational information implying a contrast in the valences of the co-occurring stimuli. Similarly, propositional processes may also be channeled towards contrastive EC effects by inducing a certain type of counteracting goal. That is, instead of preventing their effect on CS evaluations, participants may be asked to reverse the evaluative influence of the upcoming CS-US pairings by deriving CS evaluations from the reversed (instead of veridical) US valences.

The second methodological requirement (for separating associative and propositional contributions) relates to the way in which the CS evaluations are subsequently assessed. As illustrated by the consistent empirical pattern produced by assimilative vs. contrastive CS-US relations, direct measures of CS evaluation are likely to be driven by evaluative inferences about the CS valence. On such measures, CS-US associations can therefore not be expected to produce a purely "associative" pattern (i.e., assimilative EC effects that are unqualified by the contrastive implications of CS-US relations or instructions to reverse the evaluative influence of the co-occurring US valences). Accordingly, in order to demonstrate an additional associative contribution to EC, a suitable assessment of CS evaluation will have to allow for some kind of distinct pattern that is not only uniquely implied by CS-US associations, but also likely to emerge despite the dominant influence of evaluative propositions on direct evaluation measures. In the following sections, the currently most prominent approaches for revealing such patterns will be presented.

Dissociations between directly vs. indirectly measured CS evaluations. As previously described, directly assessed CS evaluations are usually strongly affected by evaluative propositions. This susceptibility to propositional processes is hardly surprising given the processing conditions under which direct measures of evaluation are typically performed. For one thing, these measures come with a clear purpose: that is, to assess the perceived valences of the previously presented CSs. Moreover, direct evaluation measures are usually administered without time pressure and attentional load, and therefore allow for a deliberate weighing of information in order to comply with the stated purpose of their administration. Under such conditions, participants are likely to be both motivated and capable of engaging in propositional reasoning so as to arrive at normatively sound CS evaluations (reflecting the joint implications of the US valences and CS-US relations).

While explaining the strong influence of evaluative propositions on direct measures of CS evaluation, the previously mentioned processing conditions also hint at how such propositional influences on evaluative behavior could be undermined (in order to reveal an additional, associative source of CS evaluation). Specifically, CS evaluations may be assessed through indirect measurement procedures which are widely assumed to provide neither the goal nor sufficient time and attentional capacity for deliberate reasoning, and should thus preclude a propositional contribution to a given instance of EC. To the extent that they do not also preclude the influence of CS-US associations, indirect evaluation measures may therefore reveal the following pattern. Now driven by the co-activated US valences, CS evaluations in the contrastive condition should indicate a regular (assimilative) EC effect that is unqualified by the contrastive implications of CS-US relations or reversal instructions (hereafter, unqualified EC).

Such unqualified EC effects on indirectly measured CS evaluations have so far been obtained in three separate studies. In one such study, Gawronski, Mitchell, and Balas (2015) instructed participants to observe the CS-US pairings while forming reversed impressions of the positive vs. negative USs presented during the pairing procedure. In spite of these reversal instructions, the indirectly assessed CS evaluations were found to be driven by the veridical (i.e., un-reversed) US valences. In a second study, Hu, Gawronski, and Balas (2017) presented participants with fictitious pharmaceutical products (as CSs) preventing positive or negative health states (as USs). Despite contrastive implications of the CS-US relation, CS evaluations on an indirect evaluation measure again indicated assimilative EC effects (reflecting the un-reversed valences of the previously co-occurring health states). Though easily explained by an evaluative influence of CS-US associations, these demonstrations of unqualified EC need to be interpreted cautiously. Such caution is warranted by the fact that in both of these studies the unqualified EC effect on the indirect evaluation measure was not accompanied by a qualified (i.e. reversed) EC effect on the additionally administered direct measure of CS evaluation. This lack of contrastive EC on directly measured CS evaluations seems to point towards inadequate experimental manipulations (of the perceived CS-US relation or the goal to reverse the evaluative influence of the CS-US co-occurrences) and therefore precludes drawing strong conclusions from the unqualified EC effects reported in these two studies.

That being said, in a third and widely cited study, Moran and Bar-Anan (2013) did indeed obtain the crucial dissociation between directly vs. indirectly measured CS evaluations in the contrastive condition. In this study, participants learned about four families of unknown cartoon characters (as CSs) that either started or stopped a positive or negative sound (as USs). Demonstrating the effectiveness of the relational manipulation, evaluative CS ratings indicated regular (assimilative) EC effects for the "starting" families, and contrastive EC effects for the "stopping" families. Additionally, the CS evaluations were also assessed on two separate implicit association tests (IAT, Greenwald, McGhee, & Schwartz, 1998) one of which compared the two "starting" families, while the other one compared the two "stopping" families. Crucially, both of these IATs produced the same (assimilative) pattern: that is, the family connected to the positive sound was preferred over the family connected to the negative sound in both the "starting" and "stopping" IAT.

At first sight, this unqualified EC effect revealed by the "stopping" IAT seems to provide strong support for an additional, associative contribution to EC, and has thus been termed "the most compelling evidence for dual-process accounts" (Hu et al., 2017, p. 19). However, IAT effects have long been known to be driven (at least, partly) by additional mental processes that have little to do with the activation of associations in memory (e.g., Mierke & Klauer, 2001; Rothermund & Wentura, 2004). Most importantly, IAT effects may emerge through a constructive process by which the key-sharing (target and attribute) categories in a given IAT block are merged into a single category reflecting some kind of (mentally constructed) similarity between the two subsumed categories (De Houwer, Geldof, & De Bruycker, 2005). To the extent that this constructive process is differentially effective across the two IAT blocks, the aforementioned mechanism may therefore produce a performance difference between the two blocks (and, by implication, a significant IAT effect) even when the target and attribute categories are not connected by associations in memory.

Applying this non-associative mechanism to the findings reported by Moran and Bar-Anan (2013) allows for an alternative account of the previously described instance of unqualified EC. That is, instead of being driven by CS-US associations, the assimilative EC effect in the "stopping" IAT effect may reflect a similarity construction based on (propositional) knowledge about the valences of the previously co-occurring sounds. Given this plausible alternative explanation, an unqualified EC effect in the "stopping" IAT cannot, by itself, provide unique evidence for an associative contribution to EC. In order to do so, such an IAT effect needs to be accompanied by additional evidence excluding an underlying similarity construction based on the US valences.

Dissociations within directly measured CS evaluations. As illustrated by the previous section, the processing conditions induced by indirect evaluation measures cannot always preclude a propositional influence, and therefore do not (necessarily) provide exclusive access to mental associations. An equivalent point can be made about direct measures of evaluation: though conducive to propositional reasoning, the deliberate processing in these measures may not eliminate an additional influence of mental associations on the registered evaluative behavior.

The notion that directly assessed CS evaluations might also be affected by CS-US associations forms the basis of a second methodological approach for revealing an associative contribution to a given instance of EC. In this approach, the potential non-exclusivity of direct evaluation measures is harnessed in order to dissociate the respective influences of CS-US associations and evaluative propositions within a single (direct) measurement procedure. To this end, an evaluative classification task is administered in two conditions which are assumed to permit associative and propositional influences on evaluative (classification) behavior to similar degrees: that is, the likelihood of a given process driving the CS classification is assumed to be equal across the two conditions (hereafter, invariance assumption). Importantly, in one condition of the classification task (typically referred to as the inclusion condition), the two sources of CS evaluation should produce identical evaluative outcomes. That is, irrespective of whether the classification response is driven by evaluative propositions or CS-US associations, the selected CS classification should indicate the same (propositionally and associatively correct) CS valence. Conversely, in the exclusion condition of the classification task, associative and propositional sources of EC should result in opposing CS classifications. Accordingly, classification responses should indicate the propositionally correct CS valences only when they are driven by evaluative propositions, but not by CS-US associations (which will produce propositionally incorrect, yet associatively correct CS classifications). Taken together, an associative contribution to evaluative classification behavior may then be identified by comparing the response frequencies produced by the two conditions: that is, an influence of CS-US associations should result in overall lower shares of propositionally correct CS classifications in the exclusion than in the inclusion condition.

As mentioned earlier, Hütter et al. (2012) implemented a version of this process dissociation approach in order to demonstrate EC in the absence of recollective memory for the US valences. However, and as already described, EC without US valence memory can be explained by both associative and propositional accounts, and does therefore not provide unique support for an associative contribution to EC (whether demonstrated through a process dissociation procedure or otherwise). Accordingly, current attempts at demonstrating association formation in EC via process dissociation procedures are focused on other, more informative phenomena that are uniquely predicted by associative accounts. To this end, the possibility of unqualified EC has been investigated in three separate studies combining a process dissociation procedure with an MPT analysis of the resulting response frequencies. In one such study, Hütter and Sweldens (2018) instructed participants to observe the CS-US pairings while forming CS evaluations based on the veridical US valences (in the inclusion condition) or based on the reversed US valences (in the exclusion condition). Subsequently assessed CS classifications (as positive vs. negative) revealed the previously mentioned lower shares of propositionally correct classifications in the exclusion than in the inclusion condition. Equivalent classification asymmetries were obtained by Kukken, Hütter, and Holland (2020) as well as by Heycke and Gawronski (2020). Adapting the learning procedure introduced by Moran and Bar-Anan (2013), Kukken et al. (2020) reported substantially lower shares of propositionally correct CS classifications for cartoon characters that had stopped positive or negative sounds (i.e., the exclusion condition) than for cartoon characters that had started these sounds (i.e., the inclusion condition). Similarly, Heycke and Gawronski (2020) found that pharmaceutical products preventing (positive or negative) health states were less often classified according to their propositional valence than pharmaceutical products that had caused them.

As previously explained, such asymmetric response frequencies are indeed implied by an additional influence of CS-US associations on directly assessed CS evaluations (and are therefore well accounted for by dual-process accounts of EC). However, the classification asymmetries between the inclusion and exclusion conditions implemented by the three author teams can also be explained by other, non-associative mechanisms. Most importantly (in the present context), these classification asymmetries might also be driven by a certain violation of the previously mentioned invariance assumption. Specifically, the lower shares of propositionally correct CS classifications in the exclusion condition may result from a weaker propositional influence on classification behavior in the exclusion condition (or, in more technical terms, from a lower probability of the CS classifications being driven by a propositionally correct inference about the CS valence). Though not yet demonstrated, such a "propositionally disadvantaged" exclusion condition seems plausible due to the following reason. As explained earlier, a propositionally correct CS classification in the exclusion condition requires an evaluative inference that involves a mental reversal of the co-occurring US valence. Since such a mental reversal is more cognitively demanding than an un-reversed processing of a given US valence, participants may refrain more often from drawing an evaluative inference in the exclusion than in the inclusion condition of the learning procedure. Such an inference asymmetry between the two learning conditions should, in turn, result in weaker recollective memory for the propositionally correct CS valences in the exclusion condition during the subsequently administered CS classification task. In order to still deliver a classification response in the absence of such memory, participants may resort to pure guessing (or other sources of CS evaluation). Crucially, these alternative sources of classification behavior will produce relatively lower shares of propositionally correct CS classifications even in the complete absence of an additional, associative influence on the evaluative classification behavior.

Given this plausible alternative account in terms of a propositionally disadvantaged exclusion condition, a classification asymmetry obtained in a process dissociation procedure cannot, by itself, provide unique support for an additional, associative contribution to EC. In order to do so, it needs to be complemented with additional evidence demonstrating its associative mental underpinnings.

Summary and overview of empirical chapters

The preceding sections have identified the currently most promising approaches for demonstrating association formation in EC. In a first such approach, an associative contribution to EC may be revealed through a reliable demonstration of EC without CS-US awareness. To this end, unawareness of the CS presentations (and, consequently, the CS-US pairings) needs to be induced through one of several psycho-physical techniques on the one hand, and confirmed through an appropriate indicator of CS unawareness on the other hand. Though valid as such, this methodological approach has so far not demonstrated EC in the absence of CS-US awareness and, by implication, an additional contribution of association formation to EC. In a second approach, an evaluative influence of CS-US associations may be revealed through dissociations of directly vs. indirectly assessed CS evaluations. Specifically, CS-US pairings combined with contrastive information (e.g., reversal instructions) may be shown to produce reversed EC effects on a direct evaluation measure, yet un-reversed EC effects when the CS evaluations are indirectly assessed. As previously described, such unqualified EC effects have been demonstrated by Moran and Bar-Anan (2013). In a third and final approach, unique evidence for association formation in EC may come from process dissociation procedures revealing lower shares of propositionally correct responses in the exclusion than in the inclusion condition of an evaluative CS classification task. As mentioned earlier, such classification asymmetries have been obtained repeatedly and across different materials and experimental manipulations for inducing assimilative vs. contrastive CS evaluations during the pairing procedure.

Aside from introducing these three approaches, the preceding sections have also revealed (some of) their hitherto unheeded limitations. In the following chapters, these limitations will be addressed in three lines of research. Firstly, Chapter 2 is concerned with the previously described shortcomings of establishing CS unawareness through chance-level performance in an (objective) identification task. To this aim, a series of studies investigating the possibility of EC for subjectively invisible (yet objectively processed) CSs will be presented. Through this novel approach (combining subliminal CS presentations with a subjective criterion of CS unawareness), my co-author and I sought to allow for potentially higher levels of unconscious CS processing and, by implication, fairer conditions for the formation of CS-US associations. In Chapter 3, the previously described alternative account of Moran and Bar-Anan's unqualified EC effect in the "stopping" IAT will be tested. By providing evidence for or against similarity construction in the experimental setting created by Moran and Bar-Anan (2013), my co-authors and I sought to assess the evidential weight of their findings in favor of an additional, associative contribution to EC. Finally, Chapter 4 reports a series of simulation studies exploring the empirical implications of a propositionally disadvantaged exclusion condition in the process dissociation procedure. By comparing the thereby identified predictions with the results of validation studies conducted by Kukken et al. (2020) and Heycke and Gawronski (2020), my co-authors and I sought to re-assess the current evidential weight in favor of an "associative" interpretation of the classification asymmetry, on the one hand, and the alternative account in terms of a propositionally disadvantaged exclusion condition, on the other hand.

Chapter 2

Evaluative Conditioning of Masked Nonwords Requires Perceptual Awareness

The evaluative conditioning (EC) phenomenon is central to the study of preference acquisition and attitude formation. Some studies have reported EC in the absence of awareness, but more recent work has cast doubt on this conclusion. In previous work, using briefly presented and masked conditioned stimuli (CSs), we found that above-chance forced-choice identification of CSs is necessary for EC. Here we extend this work by addressing more directly the inherently subjective issue of consciousness. In two studies, we assessed the quality of perceptual awareness necessary for EC. Contrasting unconscious learning claims, EC was found to depend on the participants' subjective impression of having clearly perceived at least one (or more) of the CSs features. Additional findings suggest that the perceptual awareness measure can be collected during the learning phase without interfering with EC, and that it is more sensitive than US memory measures (which suffered from forgetting artefacts). We also found that the administration of a confounded variant of the forced-choice identification task during learning may induce artifactual EC effects. However, corroborating our previous findings, administering an unconfounded version of the task does not interfere with EC.

Evaluative conditioning (EC) — the effect of a valent unconditioned stimulus (US) on the evaluation of an initially neutral conditioned stimulus (CS) after both have been paired repeatedly — is perhaps the most widely studied phenomenon of evaluative learning and central for investigating questions of preference acquisition as well as attitude formation and change. Early research reported several dissociations of EC from expectancies (Baeyens, Crombez, Van den Bergh, & Eelen, 1988; De Houwer, Baeyens, & Eelen, 1994; De Houwer, Hendrickx, & Baeyens, 1997) that have led learning researchers to conceive of EC as resulting from a learning mechanism distinct from classical conditioning (J. De Houwer et al., 2001). Building on these conclusions, social-psychological theory has characterized EC as reflecting an automatic associative attitude learning process (e.g., Gawronski & Bodenhausen, 2014; for an overview see Corneille & Stahl, 2019). Both of these literatures have proposed dual learning processes, and a central claim of these accounts is that EC can operate without awareness of the CS-US co-occurrence (De Houwer et al., 2001; Gawronski & Bodenhausen, 2014). In contrast, single-process propositional learning models hold that learning involves controlled reasoning processes, and in particular that learning about a CS-US co-occurrence relation requires awareness of that relation (e.g., Lovibond & Shanks, 2002; Mitchell et al., 2009); hence, EC without awareness cannot be explained by propositional models. The issue of EC without awareness is therefore critical for both single- and dual-process learning models.

Several reports of EC without awareness have since been published (e.g., Rydell et al., 2006), and the results of a meta-analysis (Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010) suggest that, under certain conditions (i.e., when CSs are presented subliminally), the magnitude of EC effects is independent of awareness (see also Stahl et al., 2016). More recently, however, several of the studies have been criticized on methodological grounds (Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Stahl et al., 2016, 2009), and some findings have proven difficult to replicate (Heycke et al., 2018). The debate is adequately summarized by a recent review highlighting the fact that there is (still) disagreement about the role of awareness in EC (Sweldens, Corneille, & Yzerbyt, 2014). The present study uses improved methods and presents new data to test the hypothesis of EC without awareness.

We are concerned here with awareness of the CS stimulus. Whether a stimulus enters consciousness depends on its strength as well as the amount of attention it receives (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006): Weak stimuli that remain below a certain threshold are called *subliminal.* They do not enter consciousness; yet, if attention is directed toward them (e.g., if participants focus their gaze on the relevant screen location), subliminal stimuli can affect a variety of cognitive processes (e.g., Van den Bussche et al., 2009). Subliminal presentation of CSs prevents awareness of CS-US co-occurrences: Subliminally presented CSs do not enter awareness, and therefore participants necessarily remain unaware of the CS-US co-occurrence.[1] Because awareness of CS-US co-occurrences is considered necessary for propositional evaluative learning, an EC effect with subliminal CS stimuli cannot be accounted for by single-process models.

Assessing awareness is fraught with methodological difficulty. The meta-analytic evidence for EC with subliminal CSs (Hofmann et al., 2010) relied on a classification as “subliminal” of all studies using a presentation duration of 50 ms or less. Given that perception depends on many other factors beside presentation duration (e.g., stimulus size, contrast, presence and type of masks), this simplification (while justified in the context of a meta-analysis) is not satisfactory for the present purposes. In the set of studies reporting evidence for EC with subliminal CSs, awareness is typically assessed on the participant level, using interview questions presented to participants at the end of the study (asking, e.g., whether they noticed anything unusual during the study), or asking participants to report the stimuli they have seen during the learning phase (for more discussion of previous work see Stahl et al., 2016). These approaches have been criticized as insensitive assessments of awareness of CS stimuli during learning (Pleyers et al., 2007; Shanks & St. John, 1994): most importantly, these approaches are susceptible to forgetting (i.e., a CS-US co-occurrence that entered consciousness during the learning phase may fail to be recalled at the end of the study); and the aggregated level of analysis fails to consider the possibility that a given participant’s awareness of a single CS-US co-occurrence event during learning may suffice to enable propositional learning (if that participant is classified as unaware by some criterion, this will

[1]The case of subliminal US stimulus presentation is less straightforward because of the theoretical possibility of unconscious affective processing of US valence, which may lead to conscious affective information and would render resulting EC findings more difficult to interpret.

falsely support claims of unaware EC).

In our own recent work (Stahl et al., 2016), we have used more sensitive trial-by-trial awareness assessments to test the hypothesis that EC occurs in the absence of awareness. In a series of studies, participants were asked on each trial to identify the CS stimulus they have just been shown (i.e., select it from a set of options). In the critical condition, the CS stimulus was presented only very briefly (i.e., for 30 ms) and was masked (i.e., immediately preceded and/or followed) by other stimuli, such that conscious perception was interrupted, and identification often failed (i.e., identification performance was at or only slightly above chance). EC effects for such consciously unidentified CS stimuli would provide strong evidence for EC in the absence of awareness. This trial-by-trial identification assessment allowed for a more sensitive detection of awareness than the measures used in previous studies. Using this approach, we have so far consistently found no evidence for EC in the absence of awareness (Heycke, Aust, & Stahl, 2017; Heycke & Stahl, 2018; Stahl et al., 2016). Instead, we found that reliable above-chance identification performance was necessary but not sufficient for EC. Yet, the possibility remains that different conclusions are obtained with different awareness criteria.

Assessing the contents of consciousness

Awareness can be assessed using *objective* criteria as discussed above (i.e., identification or discrimination performance) or *subjective* criteria such as confidence or perceptual experience. While the former criteria are thought to exhaustively capture all conscious processing, they may fail to do so exclusively and instead reflect also unconscious processes; in contrast, the latter are thought to exclusively reflect conscious processes, yet may fail to exhaustively do so (Reingold & Merikle, 1990). In addition, defining an awareness threshold is straightforward for objective measures (i.e., chance-level performance); subjective awareness thresholds, on the other hand, are more difficult to define. We will address these issues in turn before introducing the approach adopted in the present study.

Objective awareness and the exclusiveness problem. Despite their sensitive assessment of awareness, our previous studies arguably do not provide a conclusive answer to the question of awareness in EC. This is because they defined consciousness as the presence of above-chance CS identification performance, an objective awareness criterion (Cheesman & Merikle, 1984) that has two advantages: It is considered an *exhaustive* measure of awareness (i.e., it captures all conscious processing; Reingold & Merikle, 1990), and it relies not on introspection but instead uses empirical evidence (i.e., above-chance performance) to draw conclusions about the presence or absence of awareness. However, the drawback of using such an objective criterion is that it may not be an *exclusive* measure of consciousness, but may also reflect unconscious influences. In other words, CS identification performance may not exclusively reflect conscious processing (e.g., Henley, 1984): Choosing the correct CS from a set of options may also be supported by unconsciously acquired traces, which may evoke fluency or familiarity that can then be used to inform choices. In the extreme, performance in some tasks may be driven entirely by unconscious processes (e.g., Greenwald & De Houwer, 2017). The use of an objective criterion can thus be said to effectively “define away” unconsciousness: If any evidence for an effect on performance is defined as reflecting conscious processing, it becomes impossible to find evidence for unconscious processing.

As an alternative to objective criteria, subjective measures of consciousness have been proposed that have the advantage of exclusively measuring conscious states. They benefit from an important theoretical argument: Because the notion of being conscious (or aware) of something is inherently subjective, it has long been argued that the scientific study of awareness requires subjective (i.e., first-person) reports of the contents of consciousness (Nagel, 1974). As an example, Cheesman and Merikle (1984) asked participants for their subjective estimate of how accurate they were in an identification task; the subjective threshold was defined as the point at which participants estimated they were performing at chance (i.e., 25% correct). When participants reported purely guessing, their performance was actually considerably and significantly above chance (i.e., 66% correct). This finding suggests that for awareness to rise above the subjective threshold (defined as the feeling of mere guessing) requires higher levels of stimulus intensity (i.e., CS duration) than exceeding the objective threshold (i.e., chance-level identification accuracy). It illustrates how high levels of performance may be accompanied by a lack of awareness. Applied to our previous work, in which the brief and masked presentation condition showed only just-detectable levels of above-chance performance (but no EC), this suggests that we may have realized unfair conditions for the EC-without-awareness hypothesis, which may yet turn out to be supported when higher levels of CS duration (and hence, objective performance) are realized. Here we ask whether EC can occur at CS duration levels above the objective chance-level threshold but below a subjective threshold.

Subjective awareness and the exhaustiveness problem. Subjective measures have in turn been criticized because of lack of exhaustiveness: Participants may be conscious of a stimulus but fail to report this conscious experience. As stated by Reingold and Merikle (1990), there are “serious doubts as to whether subjective reports constitute an adequate exhaustive indicator of conscious awareness”, and “most investigators [. . .] reject any approach for distinguishing conscious from unconscious perceptual processes that is based solely on subjective reports” (1990, pp. 17–18).

One possible reason for the lack of exhaustiveness is a conservative reporting bias: Participants may decide to (i.e., place their decision criterion such as to) underreport weak conscious experiences (e.g., Bengson & Hutchison, 2007). This very real possibility is illustrated by Kunimoto, Miller, and Pashler (2001), who replicated the finding by Cheesman and Merikle (1984) (i.e., participants estimated they were guessing despite levels of discrimination accuracy as high as 60%), and additionally showed that a less biased subjective method resulted in a much lower threshold (i.e., awareness was present already at 30% accuracy). Other studies have shown overconfidence biases (e.g., Odegaard, Chang, Lau, & Cheung, 2018). Taken together, this means we cannot interpret subjective reports (e.g., of a "guess" or "unaware" state) literally without additional independent evidence. Instead of directly relying on subjective reports (i.e., a literal interpretation), they are more fruitfully interpreted in relation to independent reference points.

Defining subjective thresholds. To address the interpretability problem, researchers have suggested using first-person subjective awareness reports as data and assess (objectively) whether subjective awareness reports can discriminate between independently defined conditions (Ericsson & Simon, 1980). For instance, Kunimoto et al. (2001) assessed whether participants were more confident in their accurate than their inaccurate responses. Such a correlation between subjective measure (confidence) and objective measure (accuracy) implies that subjective and objective measures make use of the same information, and hence, that participants are aware of the information that drives their discrimination performance. This approach is concisely summarized in a recent review: "if there is a systematic way in which awareness scores are related to performance, this is indicative of conscious knowledge" (Timmermans & Cleeremans, 2015, p. 36).

Metacognitive confidence versus perceptual awareness. Besides the confidence-rating approach introduced above, the most widely discussed subjective awareness measures are post-decisional wagering (Persaud, McLeod, & Cowey, 2007) and the *perceptual awareness scale* (PAS; Ramsøy & Overgaard, 2004).[2] The PAS asks participants to report their subjective perceptual experience. It was developed in a study on the perception of colored geometric figures (circle, square, triangle) which asked participants to describe their subjective experiences on a scale with the endpoints "no experience" and "clear experience" by defining their own intermediate scale points (Ramsøy & Overgaard, 2004). The result was an intuitive 4-point scale with the two intermediate levels "brief glimpse" and "almost clear experience". Conceptually, whereas confidence and wagering measures target meta-cognitive content (i.e., they are about a decision and its accuracy), perceptual awareness ratings aims to assess the sensory contents of consciousness (i.e., they are about the stimulus; Zehetleitner & Rausch, 2013). Recent work has shown that the conceptual distinction between meta-cognitive versus sensory awareness is empirically meaningful (Jachs, Blanco, Grantham-Hill, & Soto, 2015; Zehetleitner & Rausch, 2013): Participants can be confident about the accuracy of a discrimination response without a visual experience of the stimulus. The choice of subjective awareness measure should thus be based on theoretical considerations about the relevant type of subjective awareness. As the critical issue here is whether EC occurs without awareness *of the CS stimulus*, and we are therefore interested in the sensory content of consciousness, we decided to assess subjective awareness using the perceptual awareness scale. Its methodological properties have been found to be at least as good as (and sometimes better than) those of confidence ratings and post-decisional wagering: Direct comparisons have suggested that PAS is the most exhaustive of the subjective measures (e.g., Sandberg,

[2] Much like confidence ratings, post-decisional wagering asks participants to discriminate between correct and incorrect decisions by wagering either a high or low amount of money on their accuracy. This approach has been more recently been found to be biased by risk aversion (Dienes & Seth, 2010), to lack sensitivity, and to be susceptible to motivational influences (i.e., payoff matrices; Konstantinidis & Shanks, 2014); we therefore do not consider it here further.

Timmermans, Overgaard, & Cleeremans, 2010; Sandberg, Bibby, & Overgaard, 2013; Timmermans, Sandberg, Cleeremans, & Overgaard, 2010; Wierzchoń, Paulewicz, Asanowicz, Timmermans, & Cleeremans, 2014), and may more adequately capture intermediate or graded levels of consciousness (e.g., Fazekas & Overgaard, 2018).

The present study

Here we investigate whether awareness of the CS-US co-occurrence is necessary for EC. We manipulate CS duration to realize different levels of CS identification accuracy, and we assess whether EC arises in the absence of perceptual awareness of the CS.

Does EC obtain at intermediate levels of visibility? Because of the exclusiveness problem inherent in objective awareness criteria, to conduct a fair test of unconscious EC we had allowed for above-chance identification performance already in our previous work: When CSs were presented for 30 ms and masked, identification accuracy was (just) above chance level, yet we still failed to find EC. In contrast, when CSs were presented for 100 ms, or were not masked, identification accuracy was near ceiling, and EC was obtained. We concluded that this pattern speaks against the hypothesis of EC without awareness. Yet, this work realized only the extreme ends of the visibility range (identification accuracy at floor vs. ceiling), and did not address the possibility that unconscious EC may be found with intermediate levels of visibility: Specifically, intermediate presentation durations may be low enough for CSs to remain subjectively unaware, but may be high enough to allow for substantial identification performance as well as unconscious EC effects (e.g., Kunimoto et al., 2001; Zehetleitner & Rausch, 2013). The present work addresses these limitations: We investigated EC at low, intermediate, and high levels of visibility; and we assessed subjective awareness to characterize the degree to which the CSs were consciously experienced under these conditions.

Are participants perceptually aware of the CSs? We assessed subjective awareness using the PAS, which has been found to be the most exhaustive subjective awareness measure (Sandberg et al., 2010). We also assessed whether PAS is any less exhaustive than the objective measure (i.e., whether identification performance exceeds the threshold at shorter CS durations than PAS). We combine objective and subjective approaches by relying on independent empirical evidence to define subjective thresholds:

First, we define the subjective *discrimination* threshold (i.e., the point at which participants become aware of the information necessary to identify the CS from among similar stimuli) by reference to discrimination accuracy: Participants are considered aware of the information needed to identify the CSs when their perceptual awareness ratings reflect (i.e., are correlated with) identification task performance.

Second, we define the subjective *detection* threshold (i.e., the point at which participants become aware of the presence vs. absence of a stimulus) by comparing perceptual awareness between regular (i.e., CS-present) trials and CS-absent catch trials (i.e., trials in which no CS is presented in between masks). Participants are considered aware of stimulus presence when they report higher perceptual awareness on regular trials than catch trials (i.e., when their perceptual awareness ratings are correlated with stimulus presence vs. absence).

In sum, to assess whether EC occurs without awareness, CS duration is varied and subjective awareness of CSs during learning is related to (CS identification accuracy as well as) subsequently assessed EC effects.

Overview of experiments

We report two experiments whose main goal was to test whether EC operates without perceptual awareness. CSs were presented briefly and were masked so as to realize low (e.g., degraded or incomplete) levels of perceptual awareness. We investigated whether EC would occur under conditions of below-threshold perceptual awareness.

Additional issues. In addition (and connected to) the main theoretical issue, the present studies also addressed a number of methodological issues: First, we investigated whether administering the identification task or the perceptual awareness ratings during learning caused any interference with EC (or whether the additional attention directed toward the CS by these tasks would instead boost EC, as suggested by Avneon & Lamy, 2018). To foreshadow, across two experiments, performing PAS ratings on a trial-by-trial basis did not interfere with EC for these CSs (nor did performing the identification task interfere with EC).

Second, we investigated the relative sensitivity of online awareness measures implemented at learning (i.e., PAS ratings) and memory-based awareness proxies (i.e., measures of memory for CS-US pairings; Pleyers et al., 2007; Stahl et al., 2009). Results confirmed the higher sensitivity of on-line over memory-based awareness measures.

Third, we addressed a potential artefactual EC effect that may result from the presentation of CS stimuli as response options on the identification task display. In a given learning trial of the present studies, the (brief) CS is first presented together with the (longer-duration) US in a temporally overlapping manner. Subsequently, the forced-choice identification task is administered which comprises a display of several option stimuli taken from the CS set. One of these is the CS that had just been presented on the given trial; the other option stimuli may be other CSs paired with USs (of the same or the opposite valence) on other trials, or they may be distracters that appear only as response options on the identification task display. Because the forced-choice display appears directly after the offset of the US, a CS stimulus that is part of that display is presented in spatio-temporal contiguity with the US. Such *pseudo-pairings* may boost, and/or interfere with, evaluative learning:

If an option stimulus always appear on trials with USs of the same valence (*correlated* condition), this pseudo-pairing may artificially boost EC (i.e., increase EC effects for actual CSs, and/or create EC effects for distracter stimuli). On the other hand, pseudo-pairing may interfere with EC for stimuli appearing equally often on trials with USs of both valences (*uncorrelated* condition). This effect is relevant only for CSs (for which EC effects are expected due to their pairings during the CS-US sequence) but not for distracters (for which no EC is expected). To test this, we realized a correlated condition (in Exp.1) and compared the correlated and uncorrelated conditions (in Exp.2). To foreshadow, artifactual EC was found in the correlated (but not the uncorrelated) condition, yet no interference was found in the uncorrelated condition.

Pilot studies. Three pilot studies were conducted first to identify suitable presentation conditions, probe for effects of task and task order, and pretest the perceptual awareness scale adapted for the present studies. Here we briefly summarize the results (for details see Appendix A).

The first pilot manipulated stimulus strength via CS duration and assessed identification performance as well as subjective awareness ratings. An adaptive algorithm targeted three categories (low, intermediate, high) of subjective awareness: For the *low* category, we adapted CS durations so that subjective ratings were at floor levels. This category yielded presentation durations of 30-40 ms, which were too low to elicit above-chance identification performance. The *intermediate* category aimed at low levels of subjective awareness (i.e., on the lower half of the scale but without floor effects) and yielded presentation durations of 50-60 ms, which were accompanied by above-chance identification accuracy. The *high* stimulus category aimed at high levels of subjective awareness (i.e., on the upper half of the scale but without ceiling effects). It yielded presentation durations of 70-80 ms and high levels of identification accuracy.

The second pilot study compared confidence judgments and perceptual awareness ratings and tested whether the combined administration of objective and subjective tasks (and/or their order) affected performance. We investigated the effect of load (one vs. two tasks) and order (objective vs. subjective task first). Results showed that identification accuracy was unaffected by order and load. Similarly, perceptual awareness ratings were unaffected by order and load. Confidence ratings were, however, affected by task order and load: Confidence was higher when the objective task was performed first.[3] By contrast, collecting PAS ratings in addition to the identification task on each trial did not interfere with either task, suggesting that the perceptual awareness measure is suitable to assess participants' subjective experience independently of (and unaffected by) identification performance.

The third pilot developed and tested a custom 6-point perceptual awareness scale that increased the resolution of our subjective awareness measure and adapted it to the (more complex) materials used here (as recommended by Sandberg et al., 2013). To calibrate subjective ratings, CS stimuli from different materials (faces, nonwords, products) were presented at different presentation durations (30-100 ms in steps of 10 ms); in addition,

[3] These effects may reflect artefacts of the additional CS presentation during the identification task (i.e., such presentations could help consolidate the percept). Alternatively, they may also reflect an influence of performing the identification task on subsequent confidence ratings as discussed in metacognitive accounts. In any case, as confidence ratings were affected by the identification task, they were deemed unsuitable for the present purposes also for methodological reasons.

catch trials without CSs were presented. For a test under realistic conditions, this pilot study realized an EC procedure that repeatedly presented each CS together with USs of either positive or negative valence. Results established the adequacy of the PAS scale: Participants could readily implement instructions; PAS ratings increased monotonically with CS duration and covered a wide range of values and perceptual qualities (condition means on the 6-point perceptual awareness scale ranged from approximately 1.5 to above 5); and PAS ratings of CS-absent catch trials were reliably anchored at the low end of the scale. In addition, perceptual awareness was developed substantially faster for face and product stimuli than for nonword stimuli; we thus decided to use nonwords as they could be presented for longer durations without becoming subjectively aware, which may increase the chances that automatic associative learning processes can operate on them to produce EC in the absence of awareness.

Taken together, the pilot studies showed that the conditions in the focus of the present study—above-chance identification accuracy along with reduced perceptual awareness—were realized with CS durations of 50-60 ms for nonword CSs; that participants could readily report their perceptual experiences using the 6-point PAS scale we developed for the present materials; and that PAS ratings could be administered together with the identification task without mutual interference.

Experiment 1

The main goal of this study was to assess the role of subjective awareness in EC. To increase the chances of detecting a subliminal EC effect, we realized a valence focus task because this task has shown the largest effect sizes in our previous work (Stahl et al., 2016): During learning, participants were asked to focus on the CS-US pair and to report a valence judgment (pleasant or unpleasant) for the stimulus pair. CS duration was varied across four levels (30/40 ms, 50/60 ms, 70/80 ms, and 90/100 ms). Based on our previous work, we expected to find EC with 90/100 ms but not with 30/40 ms. The focus of interest is on the intermediate CS duration levels: Based on our pretests, we expected that CS presentation durations of around 50/60 ms would yield above-chance identification but only weak and fragmentary perceptual awareness (i.e., PAS ratings on the lower half of the scale). In contrast, presentation durations of around 70/80 ms were expected to yield high levels of both objective and subjective awareness. If EC depends on subjective awareness, it should be stricted to the latter condition.

Pilot studies confirmed that objective and subjective tasks did not interfere with each other, but these studies did not address their possible interference with EC: As a secondary goal, Experiment 1 tested whether the PAS ratings interfered with EC by comparing a baseline group (i.e., without online visibility checks) with a group in which subjective awareness was assessed using the PAS scale on each trial.

We also realized a third group in which both subjective and objective awareness were tested (for the same CSs, but on different trials). The (objective) CS identification task was designed such that the response options presented on a given trial depended on (i.e., were correlated with) the valence of the US on that trial: For CSs paired with positive (negative) USs, a set of options (CSs and distractors) was presented that appeared only on trials with positive (negative) USs. The distracters differed from regular CSs in that they only appeared as response options on the identification task display. An EC effect for these pseudo-paired distracters would support the above-mentioned artifact that an appearance during the identification task can also constitute a learning trial.

A final goal of the study was to investigate the relation between awareness assessments during learning and memory-based awareness proxies. At the end of the study, we therefore assessed participants' memory for US identity and valence.[4]

Method

Materials and experiment scripts are available at the accompanying OSF repository (doi: 10.17605/osf.io/f8y6w).

Participants and design. We realized a 3 (*Visibility Check*: none, subjective, both) $\times$ 3 (*Memory Assessment*: inclusion, exclusion, control) $\times$ 2 (*US Valence*: negative vs. positive) $\times$ 4 (*CS Duration*: 30/40ms, 50/60ms, 70/80ms, 90/100ms) mixed design, with the first two factors varying between participants. A sample of $N = 90$ University of Cologne students from different majors completed the experiment in exchange for either a monetary compensation or partial course credit, a third of which was randomly assigned to each of the three visibility-check groups. This resulted in a sample size of $N = 60$ in the conditions of main interest (*none* & *subjective*), which we deemed appropriate based on our experience that this sample size was sufficient to reliably detect EC effects for (unmasked) CSs presented for 30 ms [Stahl et al. (2016); Experiments 2 & 3]: With $N{=}60$, the design had a power of $1 - \beta$ = .95 (.8) to detect a main effect of the within-subject factor *US Valence* of small-to-medium size, f=.24 (.18).[5]

Materials. For each participant, 40 nonwords were randomly drawn from a set of 54 nonwords consisting of 5 to 7 vocals and consonants, 24 of which served as CSs. For participants in the *both* group (i.e., with both subjective and objective visibility checks), 8 nonwords served as additional response options in the identification task. The remaining 8 nonwords served as novel filler stimuli in the memory tasks. For participants in the other two *Visibility Check* conditions, the remaining 16 nonwords served as filler stimuli in the process-dissociation task. As USs, we used 50 positive and 50 negative IAPS pictures (512px $\times$ 384px; Lang, Bradley, & Cuthbert, 2008). For each participant, 48 positive and 48 negative pictures were randomly selected. These pictures were to be paired with one of the CSs, while the remaining 4 pictures served as USs in catch trials (see Procedure). Each CS

[4] As part of this final phase, we also administered a process-dissociation task (Hütter et al., 2012). This task was included for the purpose of assessing its validity; the results will be reported as part of another paper that focuses on the process-dissociation task.

[5] In a one-sided t test, power is sufficient to detect $d_z = .43$ (.32).

stimulus was paired with 4 different US images of the same valence; each of these CS-US pairs was presented 3 times. Each CS was assigned to one of four levels of *CS Duration* (see *Participants and design*), with the lower and higher duration of the respective level each being applied in 50% of trials for a given CS. Half of the CSs in each *CS Duration* condition were paired with positive USs and the other half with negative USs. Assignment of stimuli to experimental conditions was randomized for each participant anew. The set of response options presented in the identification task (see below) was constant for each combination of *US Valence* and *CS Duration* (e.g., the same options were used whenever the identification task followed a CS-US presentation that included a positive US and a CS presented for 50ms, irrespective of the specific CS that had been presented). Response options comprised the three CSs of a given combination of *US Valence* and *CS Duration* (one of which was the correct response option) as well as one additional distracter stimulus that was also a nonword but that only appeared as an option on the identification task display. Different distracter stimuli were used for different combinations of *US Valence* and *CS Duration*, to allow for testing the influence of the identification task on EC (i.e., whether artifactual EC could occur due to the pseudo-pairings resulting from a temporal proximity between the US presentation and the display of response options in the identification task). Forward and backward masks were drawn randomly from a set of 12 black-and-white patterns masks (256px $\times$ 64px) for each trial anew.

Procedure. The study was administered on personal computers equipped with CRT monitors at 100 Hz refresh rates and controlled by the OpenSesame software (Mathôt, Schreij, & Theeuwes, 2012). Participants were seated in a cubicle and were told that the study was about perception and appraisal of stimuli. They were told that on each trial of the learning phase, a picture and a nonword would be presented, and that nonwords would be presented at different durations, rendering some of them clearly visible, and others barely so. Participants were instructed to attend carefully to nonwords of both brief and long durations as well as to the pictures. Subsequently, participants in all three *Visibility Check* conditions were introduced to the valence focus task which required them to indicate their personal impression of (un-)pleasantness of the nonword-picture pair on each trial of the learning phase. Participants were told to give their responses in the valence focus task by clicking one of two buttons ("rather unpleasant" vs. "rather pleasant"). At this point, participants in the *none* group (without visibility check) received a brief summary of the instructions, which was followed by the learning phase. Participants in the *subjective* group were introduced to the 6-point PAS scale which they would use to report their perceptual awareness of the CSs during the learning phase. This scale differed from the piloted 6-point scale (see Appendix A, Pilot study 3) in that (a) it was tailored to nonwords, and (b) focused entirely on perception instead of addressing both perception and certainty (as recommended by Ramsøy & Overgaard, 2004). The six scale points were: (1) I could not perceive anything of the nonword, (2) I perceived a brief glimpse of something in between the two pattern masks, (3) I perceived the outline of a nonword (but no letters), (4) I perceived at least one (or several) letters quite clearly, (5) I perceived most letters cleary, and (6) I perceived the entire nonword clearly. Participants were given the opportunity to re-read the instructions on the use of the rating scale before reviewing the instruction summary and starting the learning phase. Participants in the *both* group (i.e., who worked on both visibility check tasks) were first introduced to the identification task: They were instructed to select the CS they believe to have seen on that trial by clicking on one of four response options (and to guess if they had not seen anything). Subsequently, participants received the same PAS instructions as the *subjective* group (and were also given the chance of a second reading). Finally, after being presented with a brief summary of task instructions, the learning phase started.

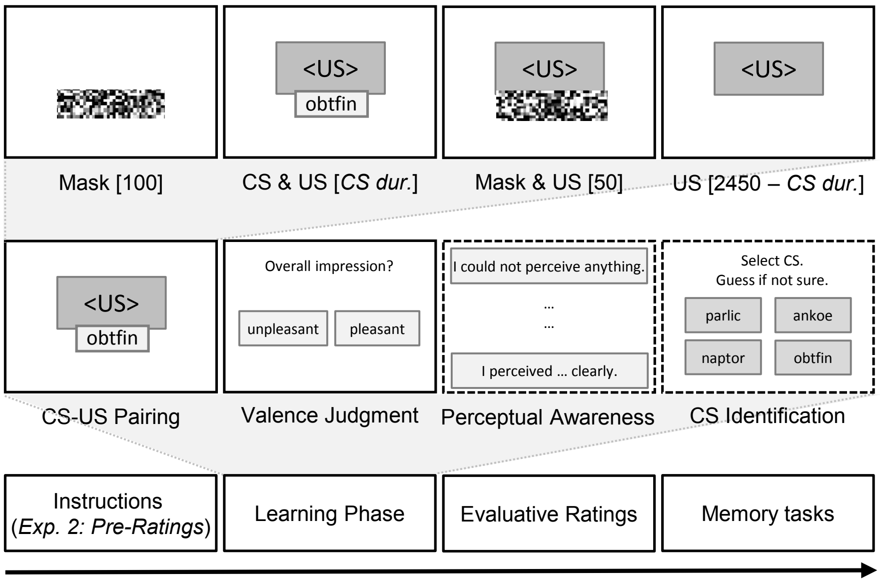

Figure 2.1. Upper row shows timing of CS-US pairing sequence (ms in brackets; initial 900 ms blank screen not shown). Middle row shows events during learning: After presentation of the CS-US pair, participants in all groups judged the valence of the stimulus pair. Visibility checks followed in the *subjective* condition (perceptual awareness ratings) and the *both* condition (perceptual awareness ratings on half of the trials; CS identification task on the other half). Bottom row shows experimental phases (in Experiment 2, CSs were evaluated both before and after learning).

The learning phase realized simultaneous (i.e. temporally overlapping) presentations of CS and US (see Figure 2.1). A trial proceeded as follows: First, a blank screen was presented for 900ms. Then a forward mask was presented centrally for 100ms (unless the CS was to be presented for 90 or 100ms, in which case no forward mask was shown and the blank screen remained in place for another 100ms). Next, the CS-US pair was presented, with the US shown in the upper half of the screen and the CS shown centrally, replacing the forward mask and slightly overlapping with the bottom margin of the US. Nonwords were presented in tiered shades of gray with lighter colors being used for shorter CS durations. CSs were always presented on a small light-gray background rectangle. The CS-US pair was presented for the assigned CS duration, after which the CS was replaced by the backward mask which disappeared after 50ms. The US remained in place for a total of 2500 ms, after which the screen was cleared (i.e., total presentation duration was 3500ms). US valence as well as CS duration varied on a trial-by-trial basis, and trial order was determined randomly for each participant anew.

Following each CS-US presentation sequence, participants of all three *Visibility Check* groups were required to perform the valence judgment task. For the *subjective* group, the valence task was followed by the PAS rating; for the *both* group, the PAS rating was administered on 50% of trials, and the identification task on the other 50% (each CS was assessed equally often on each task); for the *none* group, the next trial started immediately. In total, 6 blocks of 52 trials were administered. On 48 trials of each block, a CS-US pair was presented. Each CS appeared twice in the randomly ordered trials of one block, with the restriction that a CS could only be shown for the second time when all other CSs had been shown at least once in the given block. The remaining 4 trials were catch trials that presented an empty gray background rectangle together with a US. The rectangle was forward and backward masked and presented for 30ms or 40ms. These catch trials were administered to determine the subjective detection threshold.

After the learning phase, evaluative ratings and memory measures were collected for all 40 nonwords (regardless of whether they had previously been shown). Participants rated the CSs on a 20-point rating scale with endpoints labeled as *unpleasant* and *pleasant.* Memory for CS-US pairings was then assessed using three different tasks: the process-dissociation task, followed by an assessment of memory for US valence, and an assessment of memory for US identity. The process-dissociation task was administered first: Participants categorized nonwords according to either inclusion or exclusion instructions; a control group did not engange in this task. Because we do not report the results in this publication, details on the instructions and procedure will be omitted.[6] Next, to assess US valence memory, participants were asked to indicate, for one nonword at a time, whether it had been paired with pleasant or unpleasant pictures during the learning phase, or whether they did not remember US valence ("don't know"). Afterwards, participants' US identity memory was assessed: For one nonword at a time, participants were presented with 6 IAPS pictures as well as with a "don't know" option. Only one of these 6 pictures had previously been paired with the given CS during the learning phase (i.e., the correct response option). The remaining 5 options were randomly drawn from the total set of US pictures with the restrictions that (a) they were of the same valence as the correct option, (b) they had not been paired with the given CS during the learning phase, and (c) they had not been selected to serve as the correct response option in the present task (US identity memory test) for any of the other nonwords. For the filler nonwords, a random US valence was assigned and the six response options were randomly drawn except for restriction (c). For distractor stimuli from the

[6]Information about the implementation of the process-dissociation task (Hütter et al., 2012) can be obtained from the first author.

identification task, we provided a pseudo-correct option by randomly selecting one of the IAPS pictures that had served as a US in the trials in which the distracter had appeared (excluding IAPS pictures that served as correct options for other CSs, distracters, or filler stimuli); the remaining five options were randomly selected using restrictions (a) through (c). After the memory tasks, participants were thanked, debriefed and compensated.

Results

We first report the results of the objective and subjective awareness measures before analyzing EC effects. Tests of artifactual EC, as well as results of US memory measures, are reported in the section "Additional analyses" below. ANOVA results report Greenhouse-Geisser corrected *df*s. Data and scripts to reproduce these analyses are available at the accompanying OSF repository (doi:10.17605/osf.io/f8y6w).

Identification performance. A repeated-measures ANOVA of the proportion of correct CS identifications, with *CS Duration* as the only factor, revealed a significant main effect, $F(2.05, 59.45) = 303.12$, $MSE = 0.02$, $p < .001$, $\hat{\eta}^2_G = .859$, indicating better CS identification for higher CS durations.

As can be seen in Figure 2.2, identification performance was significantly above chance level for CSs presented for 50-60ms, $t(359) = 14.37$, $p < .001$, for 70-80ms, $t(359) = 56.38$, $p < .001$, and for 90-100ms, $t(359) = 117.17$, $p < .001$, but did not differ from chance when CSs were presented for 30-40ms, $p = .61$.

Perceptual awareness. A 2 (*CS Duration*) × 2 (*Visibility Check*: *subjective* vs. *both*) mixed-design ANOVA of the perceptual awareness ratings revealed significant main effects of *CS Duration*, $F(2.40, 139.03) = 919.50$, $MSE = 0.23$, $p < .001$, $\hat{\eta}^2_G = .862$, of *Visibility Check*, $F(1, 58) = 15.60$, $MSE = 0.82$, $p < .001$, $\hat{\eta}^2_G = .140$, as well as a significant interaction of the two factors, $F(2.40, 139.03) = 7.38$, $MSE = 0.23$, $p < .001$, $\hat{\eta}^2_G = .048$ The main effect of *CS Duration* reflected higher subjective awareness ratings for higher CS durations. The main effect of *Visibility Check* indicated higher subjective awareness ratings in the *subjective* than in the *both* group. The interaction reflected the fact that the simple main effect of *Visibility Check* was significant for CSs presented for 50-60ms, $F(1, 58) = 11.03$, $MSE = 0.47$, $p = .002$, $\hat{\eta}^2_G = .160$, and 70-80ms, $F(1, 58) = 25.51$, $MSE = 0.40$, $p < .001$, $\hat{\eta}^2_G = .305$, but not for CSs presented for 90-100ms, $F(1, 58) = 3.77$, $MSE = 0.24$, $p = .057$, $\hat{\eta}^2_G = .061$, or 30-40ms, $F(1, 58) = 2.23$, $MSE = 0.26$, $p = .141$, $\hat{\eta}^2_G = .037$.

In a second step, mean PAS ratings on trials in which CSs were presented for 30 or 40ms ($M = 1.89$) were compared to catch trials (i.e., when a blank rectangle was presented for 30 or 40ms). The mean PAS rating for catch trials was $M = 1.70$ (indicated by the horizontal line in Figure 2.2). A 2 (*Type*: "CS" vs. "catch") × 2 (*CS Duration*: 30 vs. 40ms) × 2 (*Visibility Check*) ANOVA with repeated measures on the first two factors yielded a significant main effect of *Type*, $F(1, 58) = 7.81$, $MSE = 0.07$, $p = .007$, $\hat{\eta}^2_G = .008$. All other main effects and interactions did not reach significance, all ps $\geq .147$. The main effect of *Type* reflected higher PAS ratings for CS trials than for catch trials, indicating that participants were able

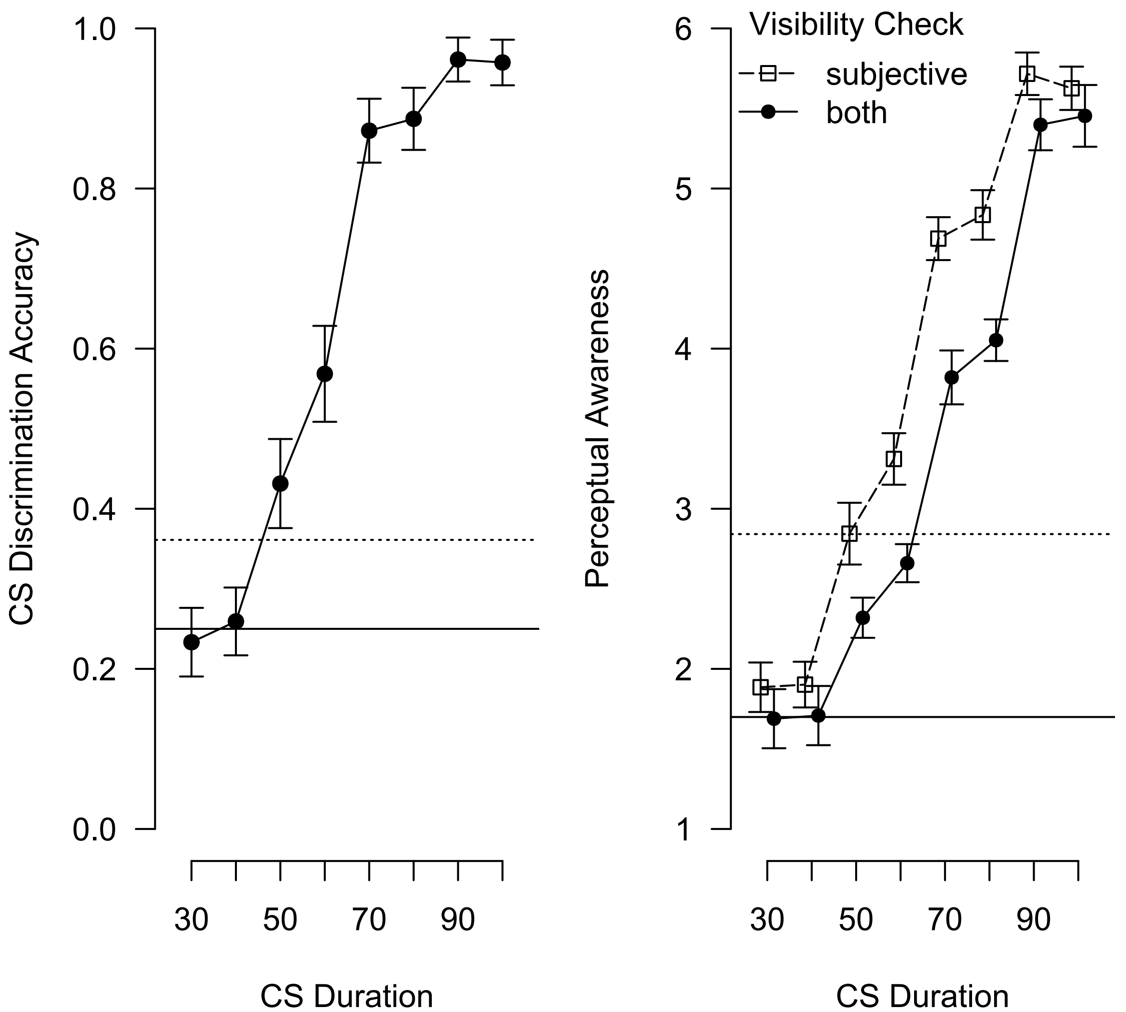

Figure 2.2. Experiment 1, result of awareness measures. Left: Mean proportion of correctly identified CSs as a function of CS Duration (solid horizontal line indicates chance level, dotted line indicates upper limit of 90% binomial confidence interval). Right: Mean perceptual awareness ratings as a function of CS Duration and Visibility Check. Horizontal lines illustrate PAS ratings for catch trials (solid line: mean; dotted line: 95% quantile). Error bars show 95% within-subject confidence intervals.

to discriminate between the presence and absence of a CS even at durations of 30 or 40 ms, and therefore, that CS presentation at this duration level was above the subjective detection threshold.

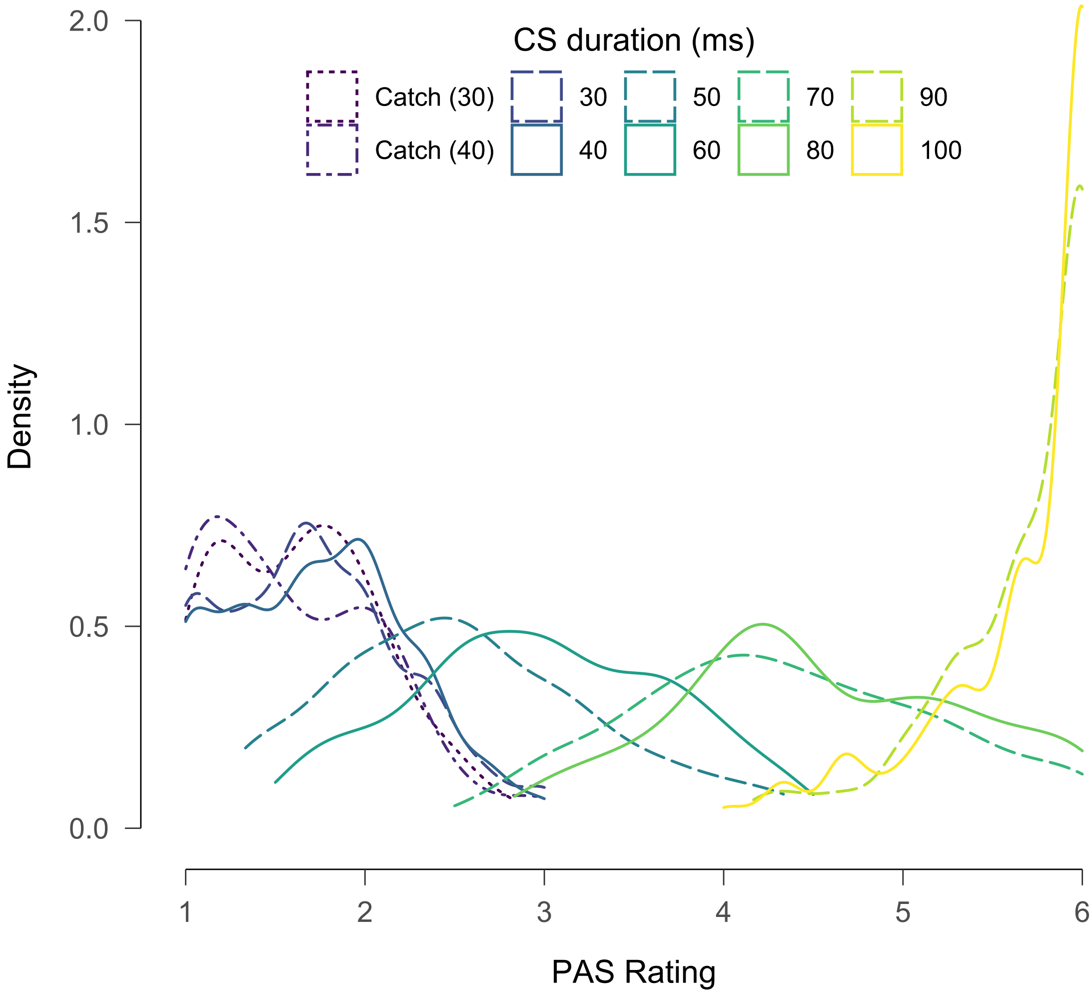

Figure 2.3. Experiment 1: Distribution of PAS ratings by trial type (only cases between the 5% and 95% quantiles are plotted). For CS durations of 30-40 ms, PAS ratings fell below 3. For CS durations of 50-60 ms, PAS ratings ranged between 1.5 and 4.5. For CS durations of 70-80 ms, PAS ratings were above 2.5. For CS durations of 90-100 ms, ratings were above 4.

Figure 2.3 shows that, even if mean PAS ratings were higher for CSs in the 30/40 ms condition than for catch trials, the range of values for both trial types was largely overlapping and the overall range of perceptual awareness quality was comparably low in both cases (i.e., regardless of the presence versus absence of a CS). Ratings of 1 (“nothing”) and 2 (“brief glimpse”) dominated (with the latter slightly more frequent for CS-present trials), and the majority of ratings fell below 3 (i.e., the 95% quantile of PAS ratings for 30/40 ms condition was 3.02, while that of catch trials was at 2.84, see also the dotted line in Fig. 2.2, right panel). In other words, participants indicated experiencing a diffuse impression of the CS stimulus on a small subset of trials in which a CS was actually presented, as well as in a comparably small subset of trials in which the CS was actually absent. This suggests that the 30/40 ms presentation conditions did not allow for a reliably subjective discrimination between CSs. In the same vein, participants were also unable to objectively discriminate between different CSs (i.e., showed chance-level identification accuracy) in the 30/40 ms condition.

In the 50/60 ms condition, awareness can be characterized as intermediate: Identification performance was significantly above chance, and condition means tended to exceed the range of ratings expected by chance (i.e., the 95% binomial confidence interval, assuming true chance performance), but they fell below the scale midpoint (i.e., .625, halfway between .25 and 1). Similarly, PAS ratings were robustly higher than in the 30/40 ms condition, but remained mostly on the lower half of the scale (with a median rating of 3.00 and condition means below the scale midpoint of 3.5).[7] Figure 2.3 shows that, while mean PAS ratings of CSs in this condition were clearly higher than those in the 30/40 ms condition, there was considerable overlap with the low-awareness (30/40 ms) condition, yet little overlap with the clearly conscious 90/100 ms condition (95% quantile: 4.67).

The 70/80 ms condition showed high levels of awareness: Identification performance was above .8 and only slightly lower than in the 90/100 condition, and PAS ratings of both conditions overlapped considerably (Fig. 2.3). Most PAS ratings were above the scale midpoint range (i.e., the 5% quantile of PAS ratings was 3.33), with a peak around 4 and approximately half of ratings at 5 or above (i.e., median = 4.83).

At the upper end, in the 90/100 ms condition, awareness was almost perfect Identification performance was at ceiling ($M = 0.96$), and so were PAS ratings ($M = 5.67$). Most PAS ratings were above 4 (i.e., the 5% quantile of PAS ratings was 4.33); the modal rating was 6 (“clear experience”), with a smaller proportion of 5s (“mostly clear”).

Reliability and correlations. We estimated split-half reliability of the objective and subjective visibility scores, based on the data from the condition in which both measures were obtained. This was done by computing two separate scores for each person and CS duration level (i.e., each based on half of the CSs in each condition and balancing US valence such that one of the two CSs of each US valence was included in each score). The correlations between the two resulting sub-scores were submitted to a Spearman-Brown correction and are reported in Table 1. Reliability was considerable for most scores except the objective visibility scores in the 30 and 40 ms conditions (this is likely because identification performance was at chance under these conditions and these scores therefore reflect only error variance). The reliability of the subjective score is high even with these brief CS

[7]The same conclusion is obtained when the range of observed PAS condition means was considered instead: It ranged from 1.7 (i.e., the experience of empty catch trials) to approximately 5.5 (i.e., the experience of clearly visible 90/100 ms CSs); the midpoint of this range lies at 3.6.

Table 1
Experiment 1: Reliablity and correlations

CS Duration	Ident.	PAS	r	r_{corr}
30	-0.83	0.95	-0.05	NA
40	0.08	0.89	0.02	0.06
50	0.52	0.89	0.80	1.00
60	0.65	0.84	0.80	1.00
70	0.73	0.76	0.68	0.90
80	0.55	0.67	0.52	0.85
90	0.85	0.65	0.64	0.86
100	0.44	0.50	0.40	0.85

Note. Split-half reliabilities, Spearman-Brown-corrected. Ident.: Identification task performance; PAS: Perceptual Awareness Scale; r_{corr}: Correlation corrected for attenuation. NA: Missing due to negative reliability estimate.

presentations, suggesting that perceptual awareness is a meaningful construct independent of identification performance. Furthermore, the reliability of the subjective score was found to be greater than that of the objective score in most cases.

Correlations between objective and subjective awareness are reported in Table 1. It shows that correlations did not differ from zero in the 30 and 40 ms conditions (reflecting the absence of systematic variance of the objective measure in that condition). In contrast, correlations were high in all other conditions (with a decreasing trend for the conditions with higher visibility that is likely due to restriction of range). In fact, correlations corrected for attenuation were (almost) perfect, indicating that objective and subjective indicators did not dissociate, and suggesting that in the present study, objective and subjective indicators of awareness were likely driven by a single latent dimension.

These analyses support the interpretation of the 30/40 ms condition as reflecting the absence of awareness. In contrast, the substantial correlations in the other conditions show that presentation was above the subjective discrimination threshold at CS durations of 50 ms or higher.

Evaluative ratings. The above analyses show that we have succeeded in creating conditions with low, intermediate, and high levels of objective and subjective awareness. We now turn to the issue of whether evaluative learning occurs under such conditions.

We focused here on the EC effect in the *none* and *subjective* groups. The evaluative ratings were submitted to a 2 (*US Valence*) × 4 (*CS Duration*) × 2 (*Visibility Check*) mixed-design ANOVA. The ANOVA revealed a significant main effect of *US Valence*, $F(1,58) = 12.77$, $MSE = 8.04$, $p = .001$, $\hat{\eta}^2_G = .021$, which was qualified by an interaction of *US Valence*

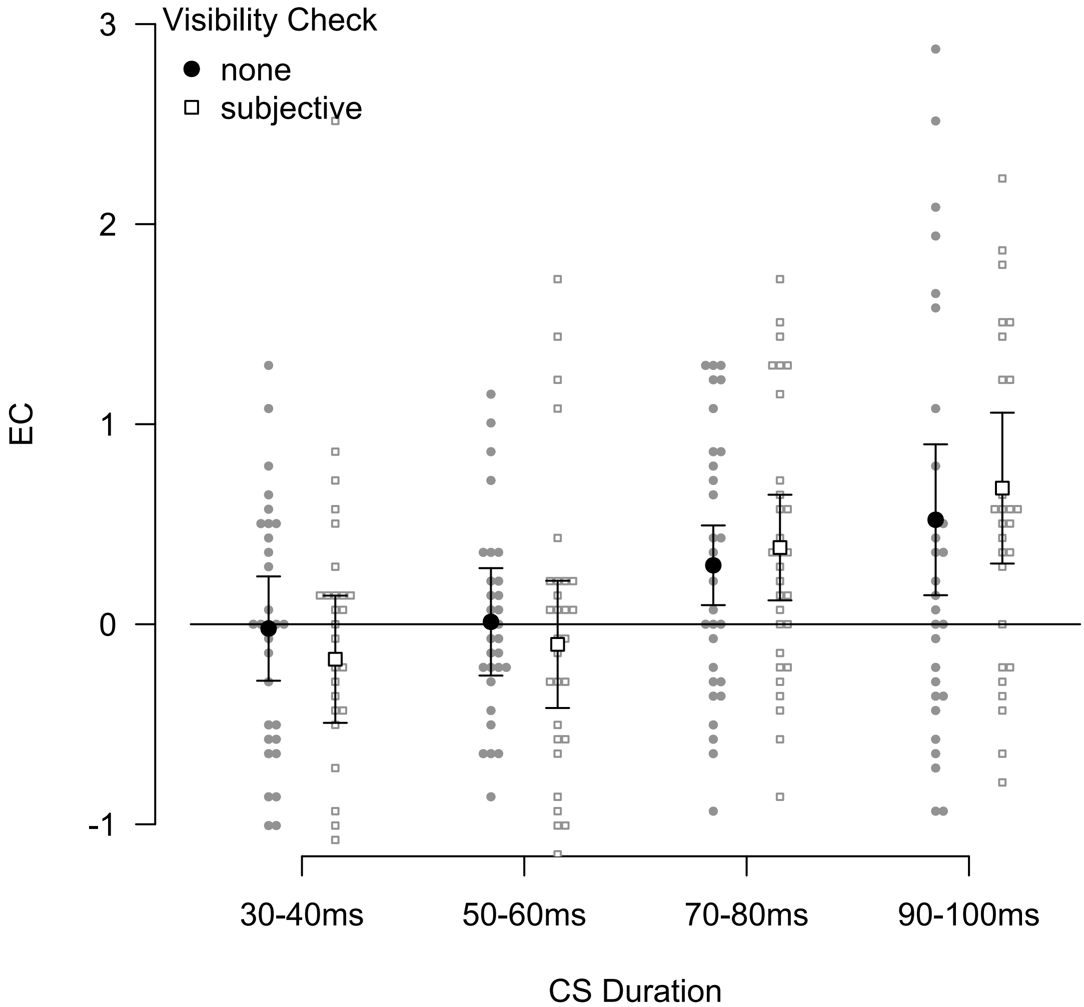

Figure 2.4. Experiment 1: Mean EC effects (and 95% within-subject CIs) as a function of CS Duration and Visibility Check. Positive EC effects reflect evaluative changes in the direction of the paired US.

with *CS Duration*, $F(2.55, 147.80) = 9.97$, $MSE = 8.36$, $p < .001$, $\hat{\eta}^2_G = .043$. All other main effects and interactions did not reach significance, all ps $\geq .09$. The results are illustrated in Figure 2.4. The interaction reflected the presence of an EC effect for 70/80 ms, $F(1, 58) = 15.51$, $MSE = 4.78$, $p < .001$, $\hat{\eta}^2_G = .060$, and 90/100 ms, $F(1, 58) = 16.15$, $MSE = 14.45$, $p < .001$, $\hat{\eta}^2_G = .119$; as well as the lack of an EC effect for both the 30/40 ms, $F(1, 58) = 1.12$, $MSE = 5.57$, $p = .295$, $\hat{\eta}^2_G = .007$, and the 50/60 ms conditions, $F(1, 58) = 0.28$, $MSE = 4.54$, $p = .599$, $\hat{\eta}^2_G = .001$.

The above results of the traditional ANOVA suggest that *US Valence* interacts with *CS Duration*, such that EC is absent with 30/40 and 50/60 ms but found with longer durations; and that trial-by-trial visibility checks did not influence evaluative learning (i.e., *Visibility Check* did not interact with *US Valence*). We computed a Bayesian ANOVA to evaluate whether the non-significant findings merely reflect the absence of evidence, or instead positive evidence for the absence of such effects. We report *Inclusion Bayes Factors* for matched models which indicate, for a given effect or predictor, how much more likely the data are under models that include the predictor versus under matched models that exclude the predictor (Hinne, Gronau, van den Bergh, & Wagenmakers, 2019). A *BF>1* is evidence for an effect of the predictor term, while a *BF<1* supports the absence of such an effect. First, we compared the set of models that include the *Visibility Check* by *US Valence* interaction with the set of matched models without the interaction, and found the posterior probability of the models without the interaction term to be higher by a factor of 7 (i.e., there was evidence against including the interaction, $BF_{Inclusion} = 0.14$).[8] Thus, we conclude that the administration of trialwise PAS ratings did not affect EC.

To probe the evidence for the presence versus absence of EC, we computed separate Inclusion Bayes Factors for the *US Valence* main effect at each level of CS Duration. Clear evidence for an influence of US valence (i.e., EC) was found in the 70/80 ms condition, $BF_{Inclusion} = 115.33$, as well as the 90/100 ms condition, $BF_{Inclusion} = 661.53$. By contrast, with shorter CS durations there was evidence in support of the null hypothesis of no EC for 30/40 ms, $BF_{Inclusion} = 0.33$; and 50/60 ms, $BF_{Inclusion} = 0.22$ (i.e., the more parsimonious models without the US valence main effect was favored under these conditions by factors of 3.02 and 4.55, respectively).

Discussion

Experiment 1 yielded EC for CSs presented for 70-80 and 90-100 ms, but not for 30-40 and 50-60 ms. First, replicating previous findings, EC was not found for 30-40 ms (presentation conditions associated with at-chance identification performance in the present study but slightly above-chance performance in our previous studies). Perceptual awareness ratings for these CSs were around 2 (“brief glimpse”), and were higher than those for catch (i.e., CS-absent) trials, suggesting that awareness was above a subjective detection threshold.

[8] The $BF_{Inclusion}$ was 108 for *US Valence*, and 16804 for the *US Valence* by *CS Duration* interaction.

More importantly, there was no EC at 50-60 ms despite clear above-chance discrimination (approximately 50% accuracy). While perceptual awareness in this condition was still low (i.e., it approximated PAS level 3 out of 6, suggesting that participants vaguely perceived the presence of the stimulus but none of its details or features), it was clearly above the subjective detection threshold (i.e., PAS ratings were higher than those for catch trials), as well as above the subjective discrimination threshold (i.e., PAS ratings were strongly correlated with identification performance). Thus, we conclude that participants were aware of the CSs under these conditions; still, the results suggested the absence of EC.

In contrast, reliable EC was obtained at CS durations of 70-80 ms. In this condition, identification accuracy was above 80%, and mean PAS ratings reached or exceeded level 4 out of 6 (i.e., participants indicated clearly perceiving at least one, or several, of the CSs features). Taken together, this pattern of findings suggests that EC requires (high levels of) subjective awareness.

In addition, we found that EC did not differ between the *none* and the *subjective* group, suggesting that the administration of perceptual awareness ratings during the learning phase did not interfere with EC.

An additional exploratory finding is worth mentioning: PAS ratings were reduced when participants additionally performed the identification task (i.e., in the *both* group when compared to the *subjective* group). This might come about because (a) participants who experienced the difficulty of performing the objective task were more conservative in their PAS ratings (i.e., a response-criteron effect); or (b) the additional task may have induced higher cognitive load, which in turn may have reduced the perceptual quality of their CS experiences (i.e., a strength effect). Before targeting these (possibly complementary) explanations, we assessed in Experiment 2 whether the effect is robust and replicable.

Experiment 2

Experiment 2 aimed at replicating and extending Experiment 1. We added a fourth condition in which both subjective and objective ratings were collected during learning, but with US valence unconfounded from the identity of the stimuli presented as response options on the identification task. If administering an identification task during learning indeed interferes with EC, then we should expect reduced EC effects in this condition. To reduce study duration, we also dropped the (objective and subjective) awareness tests for the 90/100 ms condition, both of which showed ceiling effects in Exp. 1.

Method

Participants and design. A sample of $N = 122$ University of Cologne students from different majors completed the experiment in exchange for either a monetary compensation or partial course credit. Excluding the condition with the correlated identification task, and assuming a power of $1 - \beta = .95$ (.8), this sample size (i.e., $N = 92$) allowed us to detect

small-to-medium effects of the *US Valence* within-subject factor, $f = .19$ (.15)[9]
We realized a 4 (*Visibility Check*: none, subjective, both/correlated, both/uncorrelated) × 2 (*Memory Assessment*: inclusion, exclusion, control) × 2 (*US Valence*: negative vs. positive) × 4 (*CS Duration*: 30/40ms, 50/60ms, 70/80ms, 90/100ms) mixed design with the first two factors varying between participants. For participants in the *both/correlated* and the *both/uncorrelated* groups, there was an additional within-subjects factor *Task* (subjective vs. objective) indicating for each CS which visibility check was administered.

Materials. For each participant, 72 nonwords were randomly drawn from a set of 90 nonwords consisting of 5 to 6 vocals and consonants, 32 of which served as CSs. For participants in the *both/correlated* and *both/uncorrelated* groups, 16 nonwords served as additional response options in the identification task, and the remaining 24 nonwords served as filler stimuli in the memory tasks. For participants in the other two *Visibility Check* conditions, the remaining 40 nonwords served as filler stimuli in the memory tasks. We used the same IAPS pictures and black-and-white pattern masks as in Experiment 1; again, for each participant, 48 positive and 48 negative pictures were randomly selected to be paired with one of the CSs, while the remaining 4 pictures served as USs on catch trials. Each CS stimulus was paired with 3 different US images of the same valence, with two CS-US pairs being presented 3 times, and one (randomly selected) CS-US pair being presented 4 times, totalling 10 presentations for each CS. Each CS was assigned to one of four levels of *CS Duration*, with the lower and higher duration of the respective level each being applied in 50% of trials for a given CS. Half of the CSs of each *CS Duration* condition were paired with positive USs and the other half with negative USs. For participants in the *both/correlated* and *both/uncorrelated* groups, perceptual awareness was to be rated for one half of the CSs, whereas identification was assessed for the other half (this is in contrast to Experiment 1, where each CS was rated on both tasks). This *Task* factor was orthogonal to the factors *US Valence* and *CS Duration*. In the *both/correlated* condition, the identification task presented participants with 4 response options: the correct CS; another CS with the same levels of *US Valence*, *CS Duration*, and *Task* (i.e. CSs for which PAS ratings were administered were not used as options in the identification task); and two distractor stimuli which were the same for all trials of a given combination of *US Valence* and *CS Duration*. Participants in the *both/uncorrelated* group were presented with 6 response options: the correct CS; three other CSs of the the same levels of *CS Duration* and *Task* (irrespective of *US Valence*); and two distractor stimuli which were the same for all trials of a given level of *CS Duration*. Assignment of stimuli to experimental conditions was randomized for each participant anew. Forward and backward mask were randomly assigned for each trial anew.

Procedure. Procedure, instructions, and stimulus presentation parameters of Experiment 2 were by and large identical to those of Experiment 1. Again, participants were told that the study was about perception and appraisal of stimuli. Departing from Experiment 1, before receiving instructions on the learning phase, participants rated all 72 nonwords on a 20-point rating scale with endpoints labeled as *unpleasant* and *pleasant* (i.e., the *pre-rating*

[9] In a one-sided t test, this corresponds to $d_z = .35$ (.26).

phase). Learning tasks and instructions (valence focus task, perceptual awareness rating, identification task) were the same as in the previous study. After each CS-US presentation, all participants performed the valence focus task. Participants in the *subjective* reported their perceptual awareness using the PAS scale on every trial; participants in the *both/correlated* and *both/uncorrelated* groups performed either the PAS rating or the identification task (depending on the level of the *Task* factor associated with the given CS). In total, 5 blocks of 68 trials (i.e., 2 presentations of each of the 32 CS-US pairs, plus 4 catch trials) were administered.

The learning phase was again followed by the evaluative rating of the CSs (i.e., the *post-rating* phase) and the memory tasks (process-dissociation task, and assessment of US valence as well as US identity memory). Each task was performed on all 72 nonwords, regardless of their role in the learning phase. In the US valence memory task (US identity memory task), participants had the additional option to respond "neither with pleasant nor with unpleasant pictures" ("with none of the pictures"). After the US identity memory assessment, participants were thanked, debriefed and compensated.

Results

We begin with reporting the results of the objective and subjective awareness measures before analyzing EC effects. Data and scripts to reproduce these analyses are available at the accompanying OSF repository (doi:10.17605/osf.io/f8y6w).

Identification performance. The left panel of Figure 2.5 illustrates identification task performance for both the uncorrelated and correlated groups as a function of CS duration. A 2 (*Visibility Check*: *both/correlated* vs. *both/uncorrelated*) $\times$ 3 (*CS Duration*: 30/40, 50/60, 70/80) mixed-design ANOVA of the proportions of correctly identified CSs revealed significant main effects of *Visibility Check*, $F(1, 60) = 4.45$, $MSE = 0.06$, $p = .039$, $\hat{\eta}^2_G = .042$, and of *CS Duration*, $F(1.96, 117.58) = 274.20$, $MSE = 0.02$, $p < .001$, $\hat{\eta}^2_G = .654$. The interaction of the two factors did not reach significance, $p = .349$. The main effect of *CS Duration* reflected better identification performance for longer CS durations. As suggested by Figure 2.5, the main effect of *Visibility Check* is likely the result of the different number of response options for the groups (4 options for *both/correlated*, 6 options for *both/uncorrelated*):[10] Performance levels were comparable at longer CS durations and, for the 30 and 40 ms conditions, identification accuracy was at their respective chance levels in both groups: For both groups, identification performance did not differ from chance level for CSs presented for 30-40ms, all ps $\geq .264$, but was above chance for all other CS durations, all ps $< .001$.

Perceptual awareness. A 3 (*Visibility Check*) $\times$ 3 (*CS Duration*) ANOVA of perceptual awareness ratings with repeated measures on the second factor revealed only a significant main effect of *CS Duration*, $F(1.27, 113.38) = 434.68$, $MSE = 0.57$, $p < .001$, $\hat{\eta}^2_G = .616$,

[10]Because the number of response options is confounded with the correlated/uncorrelated manipulation, the present data do not allow us to exclude the possibility that the correlation artifact may have contributed to the main effect.

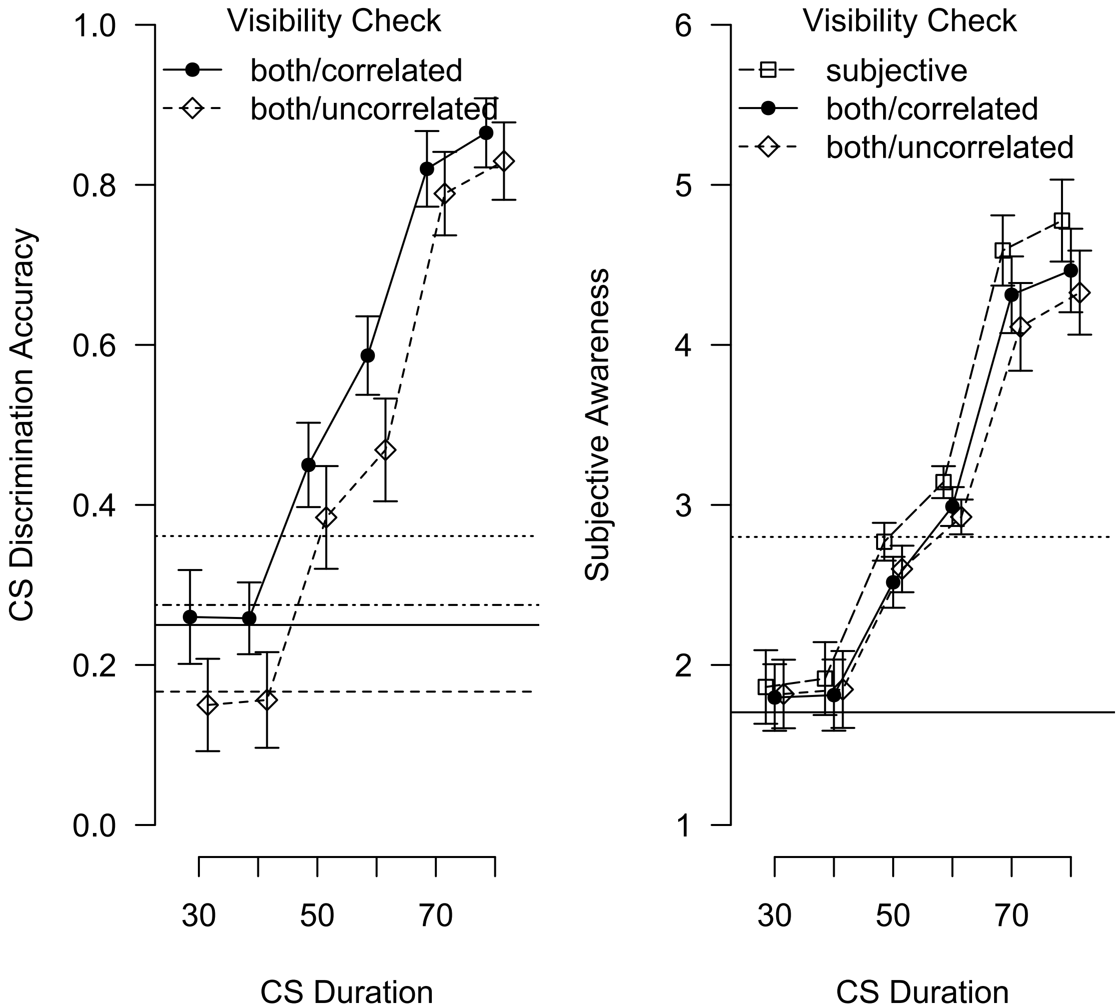

Figure 2.5. Experiment 2, results of awareness measures. Left: Mean percentage of correctly identified CSs (and 95% within-subject CIs) as a function of CS Duration and Visibility Check. Pairs of horizontal lines (both/correlated: solid, dotted; both/uncorrelated: dashed, dot-dashed) illustrate chance level and upper limit of 90% binomial CI, respectively. Right: Mean perceptual awareness ratings (and 95% within-subject CIs) as a function of CS Duration and Visibility Check. Horizontal lines illustrate PAS ratings for catch trials (solid line: mean; dotted line: 95% quantile).

indicating higher ratings for higher CS durations (right panel of Figure 2.5). All other main effects and interactions did not reach significance, all ps $\geq .374$.

The lack of a main effect of *Visibility Check* contrasts with the results of Exp.1, which found that participants who performed only the perceptual awareness ratings gave higher PAS ratings than participants who performed both the perceptual awareness ratings and the identification task. In an additional analysis, we collapsed the *both/correlated* and the *both/uncorrelated* groups and compared them to the "subjective" group in a 2 (*Visibility Check*: *subjective* vs. *both*) × 3 (*CS Duration*) mixed-design ANOVA on the perceptual awareness ratings. Once again, no significant main effect of *Visibility Check* was found, neither overall, $p = .166$, nor on any of the 3 levels of *CS Duration*, all ps $\geq .105$. This suggests that the effect of the *Visibility Check* factor in Exp.1 may have been spurious.

Next, perceptual awareness ratings on trials in which a CSs was presented for 30 or 40 ms were compared to ratings on catch trials (i.e., trials with blank rectangles presented for 30 or 40ms) in a 2 (*Type*: "CS" vs. "catch") × 2 (*CS Duration*: 30 vs. 40ms) × 3 (*Visibility Check*) mixed-design ANOVA. A significant main effect of *Type*, $F(1, 89) = 24.32$, $MSE = 0.07$, $p < .001$, $\hat{\eta}^2_G = .012$, was found. All other main effects and interactions did not reach significance, all ps $\geq .249$. The main effect of *Type* reflected higher perceptual awareness ratings for CS trials ($M = 1.85$) than for catch trials ($M = 1.70$), replicating the finding that participants were able to discriminate between the presence and absence of a CS already at 30/40 ms presentation durations.

Figure 2.6 shows that the range of perceptual awareness ratings replicated those observed in Experiment 1: The distribution of PAS ratings was comparable for catch trials (median: 1.60; 95% quantile: 2.80) and regular 30/40 ms trials (median: 1.80; 95% quantile: 3.20). In the 50/60 ms condition, we again observed low-to-intermediate perceptual awareness: PAS ratings were again at the lower end of the scale, considerably overlapping with ratings from the 30/40 ms condition (median: 2.80; 95% quantile: 4.40). The 70/80 ms condition again showed higher awareness levels and little overlap with the 30/40 ms condition: PAS ratings were again mostly at or above 4 (5% quantile: 2.20; median: 4.60).

Reliability and correlations. We again estimated split-half reliability of the CS identification and perceptual awareness scores (using the data from the two conditions in which both measures were obtained, and computed in the same manner as in Exp.1). The Spearman-Brown-corrected reliabilities, reported in Table 2, were again satisfactory for all scores except CS identification in the 30 and 40 ms conditions. As in Exp.1, the correlation between objective and subjective indicators was not significantly different from zero in the 30 and 40 ms conditions, but robustly so in the other conditions (again with a decreasing trend under higher visibility conditions).

As in Experiment 1, the above findings support the interpretation of the 30/40 ms condition as above the subjective detection threshold (i.e., PAS ratings discriminated between regular and catch trials) but below the subjective discrimination threshold (i.e., objective and subjective

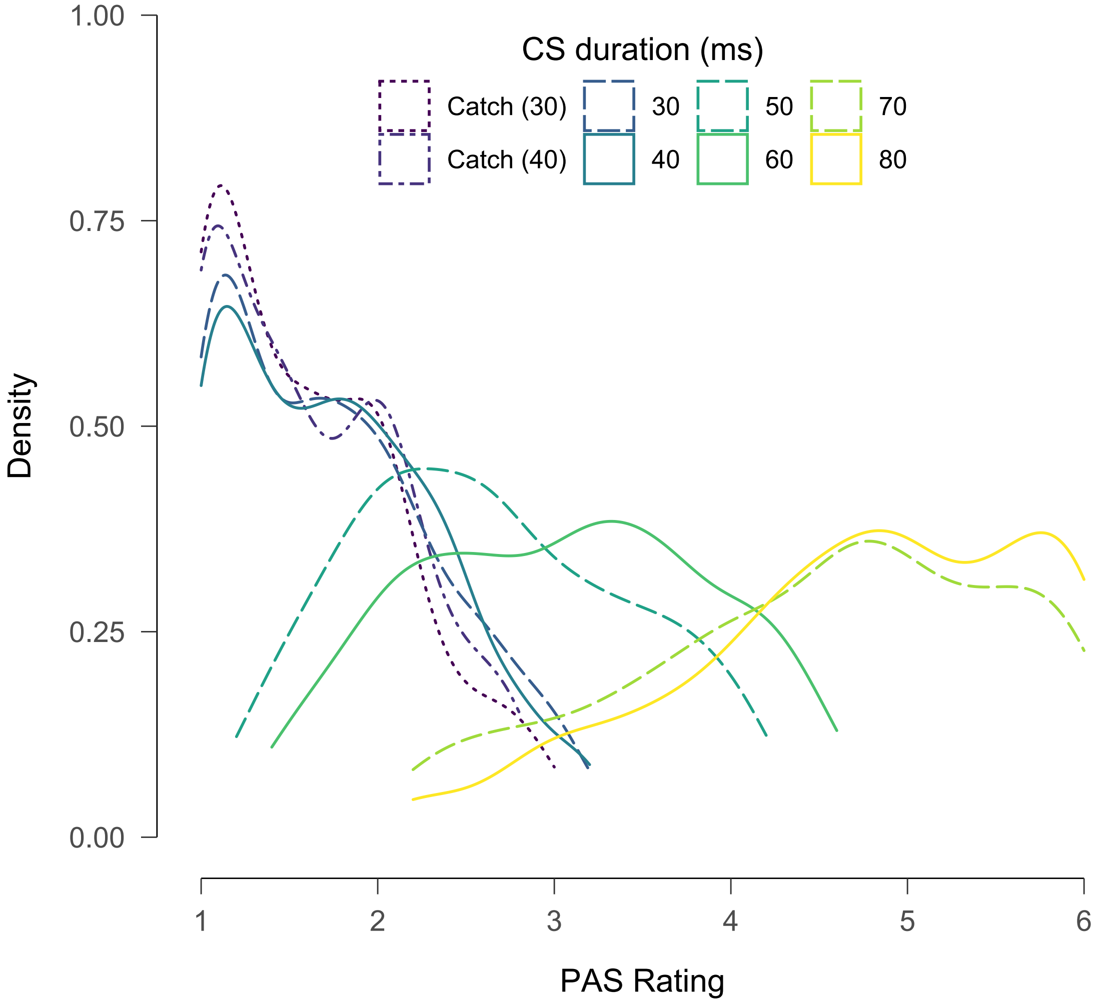

Figure 2.6. Experiment 2: Distribution of PAS ratings by trial type (only cases between the 5% and 95% quantiles are plotted). For durations of 30-40 ms, PAS ratings fell below 3. For 50-60 ms, PAS ratings fell between approximately 1.5 and 4.5. For 70/80 ms, PAS ratings were above approximately 2.5.

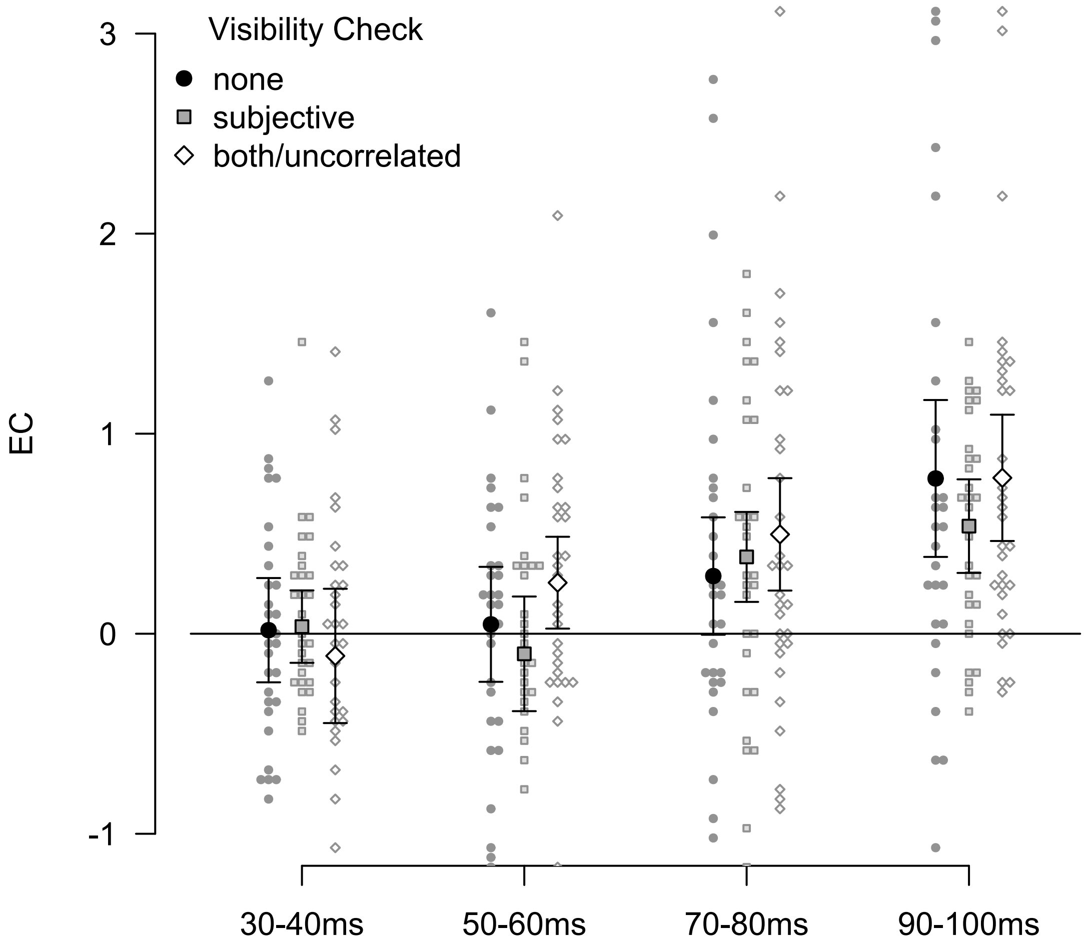

Figure 2.7. Experiment 2: Mean EC effects (and 95% within-subject CIs) as a function of CS Duration and Visibility Check. Positive EC effects reflect evaluative changes in the direction of the paired US.

Table 2

Experiment 2: Reliablity and correlations

CS Duration	Ident.	PAS	r	r_{corr}
30	-0.28	0.92	-0.05	NA
40	-0.19	0.89	0.02	NA
50	0.82	0.87	0.80	0.94
60	0.84	0.89	0.80	0.92
70	0.87	0.94	0.68	0.75
80	0.87	0.95	0.52	0.57

Note. Split-half reliabilities, Spearman-Brown-corrected. Ident.: Identification task performance; PAS: Perceptual Awareness Scale; r_{corr}: Correlation corrected for attenuation. NA: Missing due to negative reliability estimate.

indicators were uncorrelated). In contrast, the strong objective-subjective correlation in the 50/60 ms condition as well as the 70/80 ms condition supports their interpretation as above the (objective and) subjective discrimination thresholds.

Evaluative ratings. In Exp.2, each CS was evaluated both before and after the learning procedure, and the analyses focus on the *pre-post difference* in evaluations. The difference scores are shown in Figure 2.7. A 2 (*US Valence*) × 4 (*CS Duration*) × 3 (*Visibility Check*: *none*, *subjective*, *both/uncorrelated*) mixed-design ANOVA revealed a significant main effect of *US Valence*, $F(1, 89) = 33.82$, $MSE = 11.59$, $p < .001$, $\hat{\eta}^2_G = .049$, that was qualified by an interaction with *CS Duration*, $F(2.75, 244.31) = 16.77$, $MSE = 8.48$, $p < .001$, $\hat{\eta}^2_G = .049$. All other main effects and interactions did not reach significance, all ps $\geq .11$. The main effect of *US Valence* indicated more favorable evaluations for CSs paired with positive USs than for CSs paired with negative USs. The interaction reflected the fact that the simple main effect of *US Valence* was significant for CSs presented for 70-80ms, $F(1, 89) = 16.75$, $MSE = 11.02$, $p < .001$, $\hat{\eta}^2_G = .081$, and 90-100ms, $F(1, 89) = 48.44$, $MSE = 12.22$, $p < .001$, $\hat{\eta}^2_G = .205$, but not for CSs presented for 50-60ms, $F(1, 89) = 0.84$, $MSE = 6.56$, $p = .362$, $\hat{\eta}^2_G = .003$, or 30-40ms, $F(1, 89) = 0.09$, $MSE = 5.08$, $p = .768$, $\hat{\eta}^2_G = .000$.

A Bayesian ANOVA showed that the non-significant interaction between *Visibility Check* and *US Valence* again reflected evidence for the absence of an interference effect of trial-by-trial awareness assessments on evaluative learning: The resulting Bayes Factor ($BF_{Inclusion} = 0.08$) suggested that the set of models without this interaction term was favored by a factor of 12.03 over those including the term. Bayes-Factor analyses of the EC effect at specific CS duration levels again showed very strong evidence for an effect of *US Valence* in the 70/80 ms condition, $BF_{Inclusion} = 803.85$, and the 90/100 ms condition, $BF_{Inclusion} \approx 1.2e+09$, as well as evidence in support of the null hypothesis of no EC for 30/40 ms, $BF_{Inclusion} = 0.17$, and 50/60 ms, $BF_{Inclusion} = 0.24$ (i.e., the models assuming absence of EC were favored by factors of 5.94 and 4.10, respectively).[11]

Discussion

Experiment 2 replicated and extended the pattern of findings from Experiment 1. First, Experiment 2 showed again that administration of the PAS task during learning does not affect EC, supporting the validity of on-line assessments of perceptual awareness. Extending the findings of Experiment 1, we additionally showed that administering an (uncorrelated) identification task during learning also does not interfere with EC. This finding replicates a result from previous work on the role of above-chance identification in EC (Stahl et al., 2016) and lends further support to the validity of these previous findings. Moreover, the present studies replicated the central finding of that previous work—that EC is absent under presentation conditions associated with identification performance at (or just above) chance levels—in the *none* condition that excludes the possibility of interference from online awareness tasks.

The results of Experiment 2 again showed robust EC for objectively and subjectively aware CSs presented for 70-80ms and 90-100ms, but no EC for briefer CS durations. Most importantly, EC was again absent for CSs presented for 50-60ms, despite robust above-chance identification accuracy (ca. 50% correct), and perceptual awareness levels above the subjective detection and discrimination thresholds (i.e., PAS ratings were correlated CS presence vs. absence, as well as with identification performance).

Taken together, the findings from both studies consistently contrast claims of unconscious evaluative learning: First, replicating previous findings, there was no EC in the 30-40 ms condition, despite the finding that participants were able to subjectively detect CS presence. Most importantly, in the 50-60 ms condition, despite above-chance identification accuracy and above-threshold perceptual awareness, evaluative learning effects were absent. EC was found only in the 70-80 ms condition, when identification accuracy was high and perceptual awareness ratings were on the upper half of the scale (i.e., the 5% quantile at 3.5 suggests participants could consciously perceive at least parts of the stimulus on almost every trial). These findings confirm that EC requires not only high levels of identification performance (which could be explained as the result of unconscious processes) but also high levels of perceptual awareness of the CS stimuli.

Additional analyses

In addition to the main results reported above, we investigated the possibility that the identification task interferes with EC; and we compared immediate and delayed awareness assessments.

[11] Figure 2.7 suggests that the EC effect in the *both/uncorrelated* group may be significant in the 50/60 ms condition. Indeed, this effect was marginally significant, $F(1, 31) = 3.96$, $MSE = 6.95$, $p = .055$, $\hat{\eta}^2_G = .038$, but the evidence for versus against an EC effect in the 50/60 ms condition was inconclusive, $\mathrm{BF}_{10} = 1.35$. We briefly discuss this exploratory finding in the General Discussion.

Possible side effects of the identification task on EC

We consider two ways in which the identification task may affect EC. First, the focus on identifying the CS may distract from (incidentally) learning the CS-US contingencies, and this may interfere with EC. From an associative view, forming associations between CS and US may be interrupted when additional stimuli are presented with which associations may also be formed. As a consequence, EC effects should be reduced under conditions in which the identification task is performed (as compared to conditions without this task). Contrasting this notion, we found no such reduction in Experiment 2. This finding of a lack of an interference effect replicates the absence of a reduction effect in our previous work (Stahl et al., 2016, Exp. 2).

We focus here on the second possible side effect, namely that the stimuli presented as selection options in the identification task display may become associated with the US presented in the given trial. If the presentation of option stimuli is confounded with US valence (i.e., if they are pseudo-paired, as in the correlated condition), then this association may yield an unintended EC effect for option stimuli that have never been presented as actual CSs during learning. This unintended side effect may also boost EC for actual CSs, resulting in seemingly "unaware" EC effects for these stimuli. We tested these hypotheses using the data from the *both/correlated* groups from Experiments 1 and 2.

Experiment 1. Figure 2.8 shows EC effects for CSs and (pseudo-paired) distracters from the *both* condition as a function of *CS Duration*. A 2 (*US Valence*) x 4 (*CS Duration*) repeated-measures ANOVA of the evaluative ratings of paired CSs showed only a main effect of *US Valence*, $F(1, 29) = 18.87$, $MSE = 23.47$, $p < .001$, $\hat{\eta}^2_G = .150$, which was not qualified by *CS Duration*, $F(2.59, 75.01) = 0.68$, $MSE = 8.32$, $p = .548$, $\hat{\eta}^2_G = .006$. Follow-up t-tests confirmed the presence of EC on each level of the CS duration factor (30/40 ms: $M_d = 1.99$, 95% CI $[0.83, \infty]$, $t(29) = 2.90$, $p = .003$; 50/60 ms: $M_d = 2.54$, 95% CI $[1.22, \infty]$, $t(29) = 3.25$, $p = .001$; 70/80 ms: $M_d = 3.22$, 95% CI $[1.47, \infty]$, $t(29) = 3.13$, $p = .002$; 90/100 ms: $M_d = 3.11$, 95% CI $[1.54, \infty]$, $t(29) = 3.36$, $p = .001$).

Next, we tested for possible EC effects for pseudo-paired distracters. EC obtained for distracters on trials with a CS duration of 30-40ms, $M_d = 2.83$, 95% CI $[0.63, \infty]$, $t(29) = 2.19$, $p = .018$, but not for distractors on trials with longer CS durations (50-60ms: $M_d = 1.13$, 95% CI $[-1.27, \infty]$, $t(29) = 0.80$, $p = .214$; 70-80ms: $M_d = 0.63$, 95% CI $[-1.13, \infty]$, $t(29) = 0.61$, $p = .273$; 90-100ms: $M_d = -0.43$, 95% CI $[-2.32, \infty]$, $t(29) = -0.39$, $p = .650$).

We speculated that EC via pseudo-pairing was restricted to brief CS durations because the identification task was particularly difficult on these trials: Given that participants had not been able to clearly perceive the CSs on those trial due to their brief presentation, they would find it more difficult to identify the correct response, and therefore spend more time looking at the response options trying to find the correct one. By contrast, when CSs had been presented for longer durations and participants were able to perceive at least parts of the stimulus, response selection should be more efficient, and participants should therefore

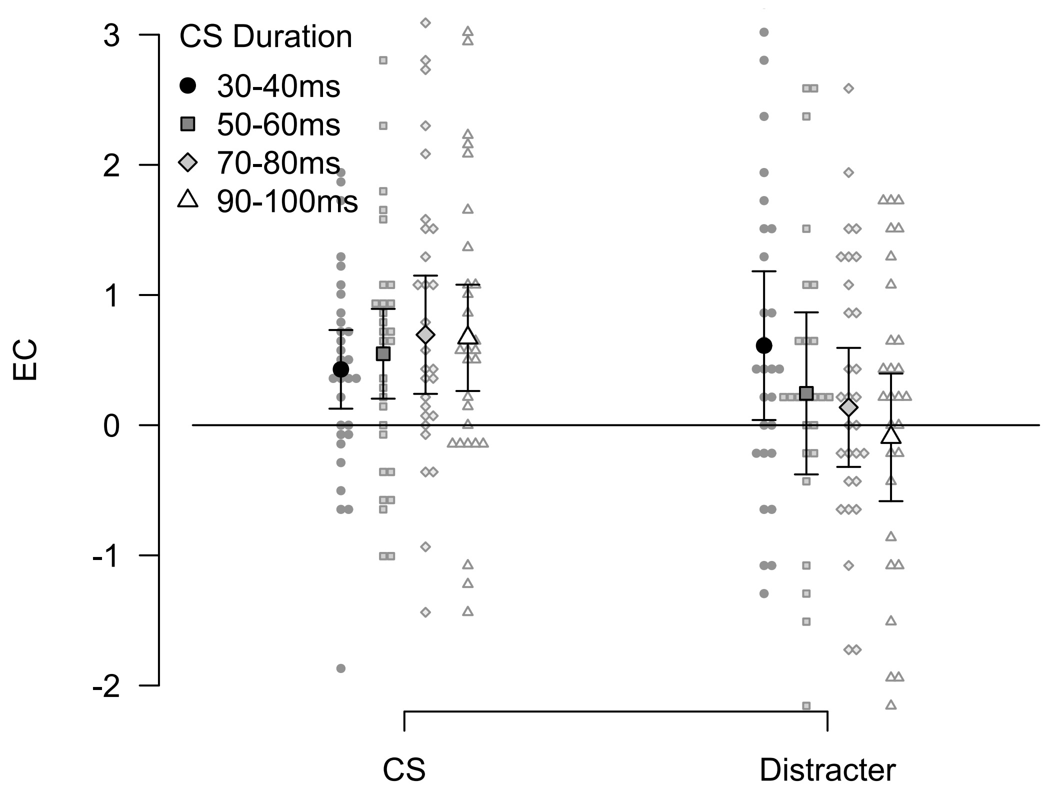

Figure 2.8. Experiment 1: Mean EC (i.e., difference of standardized mean ratings for positive vs. negative USs) for paired CSs and pseudo-paired distracters in the “both” group (error bars show 95% CIs).

pay less attention to the additional response options including the distracter. In support of this notion, Figure 2.10 illustrates that RT in the identification task varied as a function of CS duration, $F(1.73, 50.14) = 59.81$, $MSE = 884,962.65$, $p < .001$, $\hat{\eta}^2_G = .368$: RT was significantly greater for trials that presented the CS for 30-40ms than for trials with longer CS durations (e.g., 30/40 vs. 50/60 ms, $M_d = 193.98$, 95% CI $[69.57, \infty]$, $t(359) = 2.57$, $p = .005$).

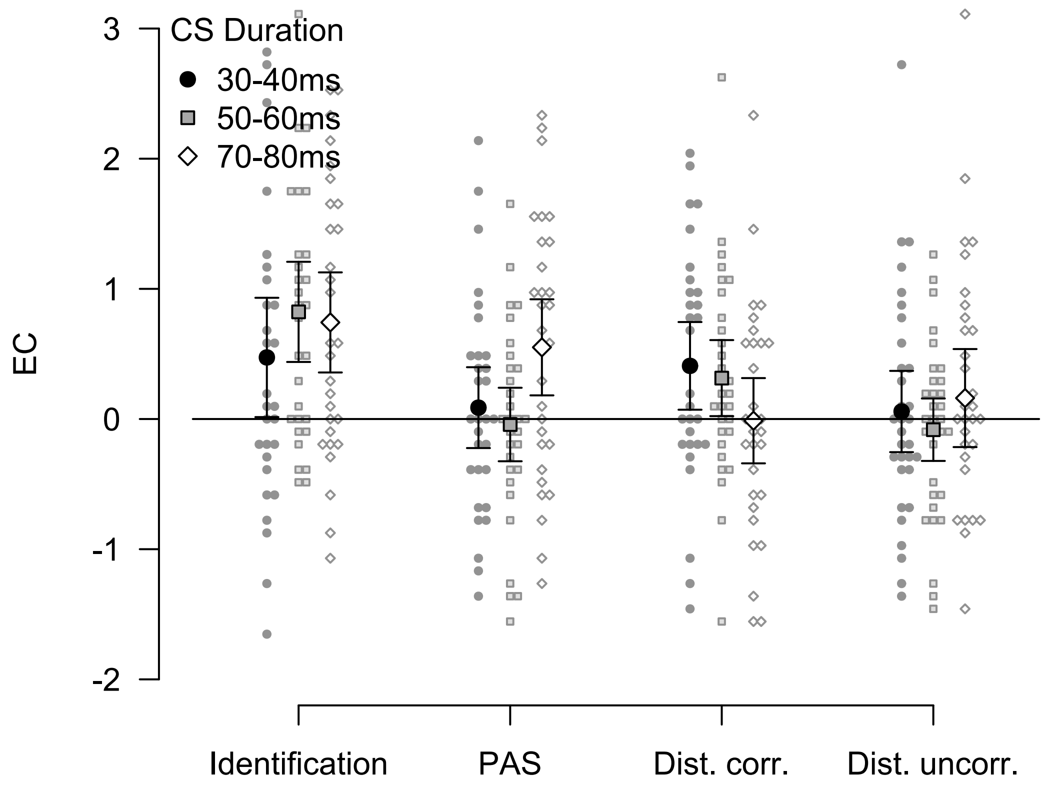

Figure 2.9. Experiment 2: Mean EC (difference of standardized mean ratings for positive vs. negative USs) for CSs in the correlated condition whose visibility was assessed with the identification versus PAS tasks, and for pseudo-paired distracters in the correlated versus uncorrelated conditions (error bars show 95% CIs).

Experiment 2. The data from Experiment 2 again yielded evidence for the pseudo-pairing artifact induced by the identification task that confounds response options with US valence (see Figure 2.9). In this study, the confound was present only in the *correlated* condition and absent from the *uncorrelated* condition. Furthermore, in each condition, only half of the CSs

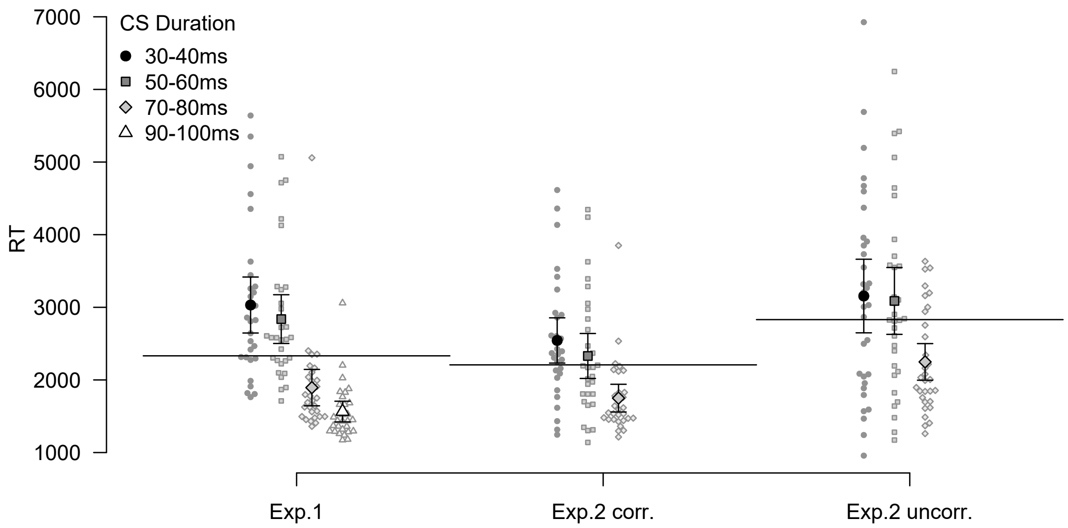

Figure 2.10. Mean identification-task RT in Experiments 1 and 2 as a function of CS Duration (error bars show 95% CIs). Horizontal lines indicate overall means.

were always assessed using the identification task; the other half of CSs was always assessed using the perceptual awareness task. An artifactual EC effect should therefore obtain only for the former subset of CSs, and only in the *correlated* condition.

As expected, only those CSs in the *correlated* group that were submitted to the identification task showed an artificial EC effect. A 2 (*US Valence*) by 3 (*CS Duration*) repeated-measures ANOVA was computed for paired CSs submitted to the identification task in the *correlated* condition. It yielded a significant effect of *US Valence*, $F(1,29) = 23.30$, $MSE = 23.59$, $p < .001$, $\hat{\eta}^2_G = .181$, which was unqualified by *CS Duration*, $F(1.99, 57.64) = 1.09$, $MSE = 12.27$, $p = .342$, $\hat{\eta}^2_G = .011$. In particular, pseudo-pairing produced an EC effect for paired CSs that was significant also at brief CS durations of 30-40ms, $F(1,29) = 4.46$, $MSE = 19.91$, $p = .043$, $\hat{\eta}^2_G = .086$; as well as 50-60ms, $F(1,29) = 19.15$, $MSE = 14.04$, $p < .001$, $\hat{\eta}^2_G = .250$. There was also an EC effect for paired CSs in the 70-80 ms condition, $F(1,29) = 15.58$, $MSE = 14.02$, $p < .001$, $\hat{\eta}^2_G = .226$; however, it is most likely not due to the pseudo-pairing artifact but instead reflects a regular EC effect that parallels the findings reported above for the data from the *none* and *subjective* conditions.

As predicted, a different pattern was obtained for the paired CSs submitted to the PAS task: Here, the effect of *US Valence*, $F(1,29) = 4.01$, $MSE = 11.71$, $p = .055$, $\hat{\eta}^2_G = .026$, was qualified by *CS Duration*, $F(1.95, 56.68) = 4.31$, $MSE = 9.17$, $p = .019$, $\hat{\eta}^2_G = .042$. Consistent with the above findings for all other CSs, EC was not obtained in the 30-

40ms, $F(1, 29) = 0.33$, $MSE = 9.15$, $p = .569$, $\hat{\eta}^2_G = .006$; nor in the 50-60ms condition, $F(1, 29) = 0.09$, $MSE = 7.56$, $p = .762$, $\hat{\eta}^2_G = .002$; it emerged only with CS durations of 70-80 ms, $F(1, 29) = 9.32$, $MSE = 12.93$, $p = .005$, $\hat{\eta}^2_G = .130$.

We also replicated the pseudo-pairing EC effects on identification-task distracters in the *correlated* condition; in this study, however, the effect was significant not only for the 30-40ms condition, $M_d = 2.10$, 95% CI $[0.66, \infty]$, $t(29) = 2.48$, $p = .010$, but also for the 50-60ms condition, $M_d = 1.62$, 95% CI $[0.37, \infty]$, $t(29) = 2.20$, $p = .018$; it was again absent from the 70-80ms condition, $M_d = -0.07$, 95% CI $[-1.47, \infty]$, $t(29) = -0.08$, $p = .532$. In contrast, effects of *US Valence* on distracters were entirely absent in the *uncorrelated* condition; for 30/40 ms, $M_d = 0.30$, 95% CI $[-1.04, \infty]$, $t(31) = 0.38$, $p = .354$; 50/60 ms, $M_d = -0.42$, 95% CI $[-1.45, \infty]$, $t(31) = -0.70$, $p = .755$; and 70/80 ms, $M_d = 0.83$, 95% CI $[-0.78, \infty]$, $t(31) = 0.87$, $p = .195$.

Further replicating Experiment 1, Figure 2.10 illustrates that identification-task RT in the *correlated* condition again varied as a function of CS duration, $F(1.81, 52.45) = 22.97$, $MSE = 487,656.73$, $p < .001$, $\hat{\eta}^2_G = .165$: RT was again significantly greater for trials with a CS duration of 30-40ms than for trials with longer CS durations (30/40 vs. 50/60 ms: $M_d = 213.84$, 95% CI $[94.27, \infty]$, $t(239) = 2.95$, $p = .002$; 50/60 vs. 70/80 ms: $M_d = 580.21$, 95% CI $[470.44, \infty]$, $t(239) = 8.73$, $p < .001$). In other words, pseudo-pairing EC effects were again accompanied by longer response latencies in the identification task. A similar pattern was obtained in the *uncorrelated* condition (with significant RT differences between 30/40 and 70/80 ms, $M_d = 907.19$, 95% CI $[766.68, \infty]$, $t(255) = 10.66$, $p < .001$, as well as 50/60 and 70/80 ms, $M_d = 839.56$, 95% CI $[665.93, \infty]$, $t(255) = 7.98$, $p < .001$, but not 30/40 and 50/60 ms, $M_d = 67.63$, 95% CI $[-92.13, \infty]$, $t(255) = 0.70$, $p = .243$), suggesting that the RT effect is not due to the confound but reflects task difficulty.

Taken together, the effect of *US Valence* on distracters, as well as on CSs for which the confounded identification task was administered, can be attributed to the pseudo-pairing artifact: When the stimuli presented as response options in the identification-task display were confounded with US valence (i.e., in the correlated condition), reliable and replicable EC effects were found for these (distracter) stimuli, even if they were never actually presented as CSs. This pseudo-pairing effect depended on the time participants spent studying the response options: It was found only on trials with briefly presented CSs because response selection was more difficult and therefore took longer; the pseudo-pairing artifact disappeared when response selection was less difficult and proceeded faster.

Notably, this artefact did not apply to CSs that were never submitted to the identification task; and it also did not apply to CSs (or distracters) that were submitted to an unconfounded version of the identification task (as in our previous work, Stahl et al., 2016). EC studies planning to use an identification task during learning should take care to use unconfounded response options.

Perceptual awareness and memory measures

The present studies combined two ways in which the role of awareness has been assessed in the EC literature so far: First, the most wide-spread approach has been to assess participants' memory for the CS-US pairing after the learning phase. This is typically done to avoid attentional effects during learning that may arise when, as a consequence of measuring awareness of CS-US co-occurrences during learning, attention is directed towards those co-occurrences; such attentional effects interfere with the goal of creating incidental learning conditions. The present studies focused on a different measure, perceptual awareness ratings, that provides an arguably more direct and—by virtue of being collected during the learning phase—possibly more sensitive awareness assessment. In support of this general approach, the present results show that assessing perceptual awareness during learning does not interfere with EC. We speculate that, because of their focus on the perception of the CS, perceptual awareness ratings should also be applicable unter incidental conditions as they do not direct participants' attention towards CS-US co-occurrences.

Here we explore the relative sensitivity of these two types of awareness measures, with a focus on two theoretically and methodologically relevant patterns:

(1) If we observe above-chance memory under conditions of low perceptual awareness, this would suggest that memory measures may be driven by unconscious processes (e.g., fluency or familiarity). If the finding of above-chance memory is then interpreted as evidence of awareness, this would imply that memory-based awareness proxies may falsely characterize unconscious learning as conscious.

(2) Conversely, if we observe chance-level memory under conditions associated with substantial perceptual awareness, this would suggest that awareness during learning may no longer be detectable at later stages; in other words, that forgetting occurred between learning and test. This would imply that memory-based awareness proxies may falsely characterize conscious learning as unsconscious.

We analyzed memory data for paired CSs from the conditions without an identification task from both experiments as a function of *CS Duration* and *Visibility Check* (the 90/100 ms condition was not included because of the lack of perceptual awareness data in Exp.2). We considered both US identification and US valence classification (the latter was found a better predictor of EC; Stahl & Unkelbach, 2009; Stahl et al., 2009). For both US identification and US valence classification, we analyzed accuracy (i.e., the proportion of correct responses among the correct and incorrect responses) and uncertainty (i.e., the proportion of cases in which the *don't know* option was selected).[12] Overall, memory was relatively poor and participants were uncertain in the majority of cases (US identification: 52% of responses; US

[12]We excluded *neither* responses from the reported analyses for comparability reasons (i.e., because this response option was presented to participants only in Exp.2). Because for paired CSs the *neither* option indicates absence of memory and can be considered a second *don't know* option, a second set of analyses that treated them as such was conducted; it yielded the same overall pattern depicted in Figure 2.11.

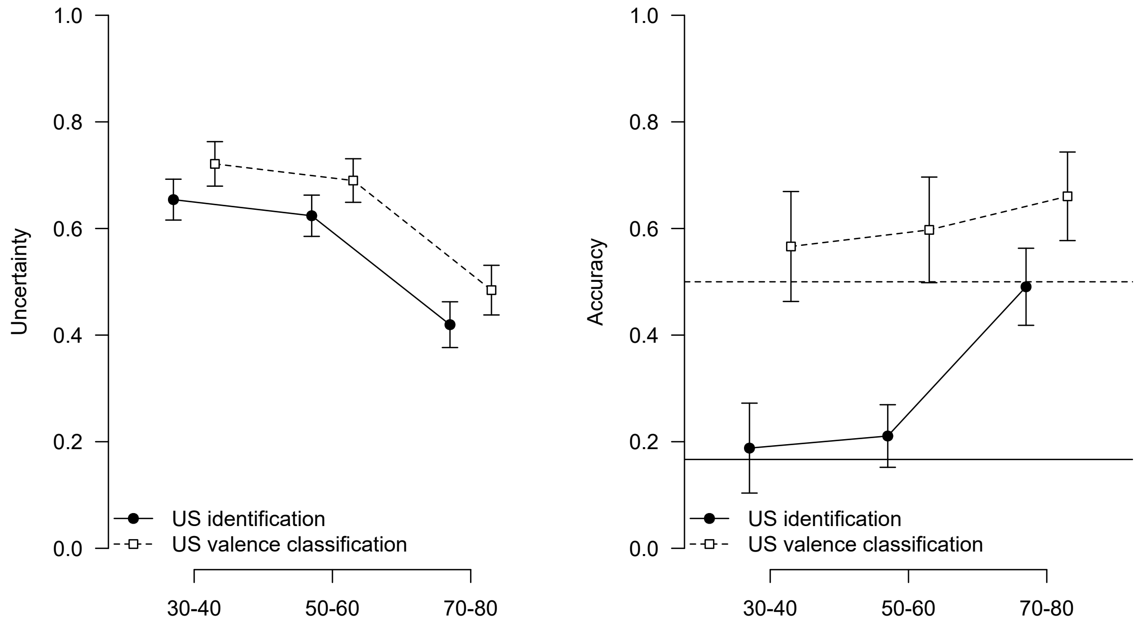

Figure 2.11. US identification and US valence classification as a function of CS duration (error bars show 95% CIs). Left: Uncertainty (i.e., proportion of *don't know* responses). Right: Accuracy (i.e., proportion correct out of correct and false responses). Solid line indicates chance level for US identification, dashed line indicates chance level for US valence classification.

valence classification: 51% of responses), with the accuracy data reflecting a smaller subset (i.e., 48% and 49%).

Figure 2.11 (left panel) shows that participants' uncertainty (i.e., the rate of *don't know* responses) mirrored the EC results: Uncertainty was at comparably high levels for the 30/40 and 50/60 ms conditions, but decreased markedly under the 70/80 ms condition. Two 4 (*Visibility Check* condition) by 3 (*CS Duration*) ANOVAs showed that uncertainty decreased as a function of CS duration for both US identification, $F(1.60, 111.66) = 13.98$, $MSE = 0.04$, $p < .001$, $\hat{\eta}^2_G = .043$, and US valence classification, $F(1.65, 103.99) = 9.56$, $MSE = 0.04$, $p < .001$, $\hat{\eta}^2_G = .037$ (no other effects were significant at $\alpha < .05$).[13] When the 70/80 ms condition was excluded, the *CS Duration* factor was no longer significant (US identification: $F(1, 70) = 0.01$, $MSE = 0.02$, $p = .929$, $\hat{\eta}^2_G = .000$; US valence classification: $F(1, 63) = 1.74$, $MSE = 0.02$, $p = .192$, $\hat{\eta}^2_G = .003$).

Figure 2.11 (right panel) shows that accuracy mirrored the uncertainty pattern, with above-chance memory only for the 70/80 ms condition. For US identification, in addition to a main

[13] After listwise exclusion of cases with missing values, $n = 74$ (75) and $n = 67$ (58) participants remained, respectively, in the ANOVAs of US identification and US valence classification uncertainty (accuracy).

effect of *CS Duration*, $F(1.90, 135.02) = 25.34$, $MSE = 0.07$, $p < .001$, $\hat{\eta}^2_G = .203$, an effect of condition was obtained, $F(3, 71) = 4.01$, $MSE = 0.05$, $p = .011$, $\hat{\eta}^2_G = .046$ (reflecting greater memory accuracy for the *subjective* group of Exp.2 than the other groups), as well as the interaction of *Visibility Check* condition with *CS Duration*, $F(5.71, 135.02) = 5.00$, $MSE = 0.07$, $p < .001$, $\hat{\eta}^2_G = .131$ (reflecting somewhat better memory for participants in the *none* than the *subjective* groups in the 30-40 ms condition). For US valence classification accuracy, there were no significant effects (in particular, the *CS Duration* main effect was not significant, $F(1.99, 107.42) = 0.03$, $MSE = 0.12$, $p = .969$, $\hat{\eta}^2_G = .000$). Mirroring the pattern for the uncertainty data, the main effect of *CS Duration* on US identification accuracy was also not significant after the 70/80 ms condition was excluded, $F(1, 74) = 0.28$, $MSE = 0.07$, $p = .601$, $\hat{\eta}^2_G = .002$.

Although the results were less clear for US valence classification, the main finding was again that accuracy followed the pattern observed for EC: Above-chance memory accuracy was restricted to the 70/80 ms condition (US identification: $M = 0.49$, 95% CI $[0.42, 0.56]$, $t(83) = 9.29$, $p < .001$, $\mathrm{BF}_{10} = 4.19 \times 10^{11}$; US valence classification: $M = 0.65$, 95% CI $[0.58, 0.72]$, $t(81) = 4.27$, $p < .001$, $\mathrm{BF}_{10} = 360.60$). Memory accuracy did not differ from chance in the 30/40 ms condition (US identification: $M = 0.19$, 95% CI $[0.13, 0.26]$, $t(68) = 0.84$, $p = .401$, $\mathrm{BF}_{01} = 5.37$; US valence classification: $M = 0.59$, 95% CI $[0.50, 0.69]$, $t(56) = 1.97$, $p = .054$, $\mathrm{BF}_{01} = 1.15$) and the 50/60 ms condition (US identification: $M = 0.21$, 95% CI $[0.14, 0.27]$, $t(71) = 1.29$, $p = .202$, $\mathrm{BF}_{01} = 3.50$; US valence classification: $M = 0.59$, 95% CI $[0.50, 0.68]$, $t(65) = 1.90$, $p = .061$, $\mathrm{BF}_{01} = 1.36$).

In sum, CS-US pairing memory depended on CS duration, but the pattern of this dependency deviated from that obtained for perceptual awareness (which was higher in the 50/60 ms than the 30/40 ms condition), while it conformed to the EC effect pattern (i.e., memory, as well as EC, was restricted to CS durations of 70/80 ms). In other words, perceptual awareness was more sensitive to increasing CS duration than pairing memory: While both perceptual awareness during learning and memory were at chance for 30/40 ms, an increase of CS duration to 50/60 ms was accompanied by the appearance of perceptual awareness (i.e., a correlation between PAS ratings and CS identification) but was not reflected in an increase in pairing memory (i.e., accuracy was still not different from chance). Regarding the two patterns instrudoced above, there was no evidence for the first, above-chance memory under low-PAS conditions, suggesting that memory-based proxies are not in danger of falsely characterizing an unconscious learning effect as conscious. However, results support the second pattern: The fact that memory measures did not reflect the increased perceptual awareness in the 50/60 ms condition (and the presence of awareness indicated by the correlation between subjective and objective measures) points to an effect of forgetting. Such forgetting may distort interpretations: If we had found EC in that condition, it would have been falsely classified as unconscious learning based on memory measures; in contrast, based on the perceptual awareness measure, the learning effect would have been more adequately characterized as being accompanied by (at least partial) consciousness.

General Discussion

In two studies, evidence for the absence of EC was obtained for masked CSs presented for 30-40 ms at the objective awareness threshold, confirming our previous findings that above-chance CS identification is necessary (but not sufficient) for EC (Heycke et al., 2017, 2018; Heycke & Stahl, 2018; Stahl et al., 2016).

The main contribution of the present work is the investigation of the role of subjective awareness for EC: If the subjective awareness threshold requries higher CS duration levels than the objective threshold, then we might find that EC can occur without subjective awareness. We attempted to realize conditions above the objective threshold but below the subjective threshold in two studies but found that the subjective awareness threshold was identical to the objective threshold in the present study: At CS durations of 50-60 ms, identification accuracy was robustly above chance, and perceptual awareness ratings were correlated with identification accuracy). Although awareness exceeded both the objective and subjective thresholds, EC was absent under these conditions of low or intermediate subjective awareness. This finding provides further evidence against the hypothesis that EC operates without awareness.

Additional findings

We also found that the perceptual awareness task did not interfere with EC; that the identification task can potentially induce unintended EC effects; and that perceptual awareness ratings are more sensitive than memory-based awareness measures.

Perceptual awareness ratings did not interfere with EC. Interference with EC may arise from the administration of perceptual awareness ratings, which, if implemented during the learning phase, may affect learning processes. To address this possibility, in two studies we compared EC effects obtained under conditions with on-line perceptual awareness ratings with those obtained under conditions in which participants did not perform any on-line awareness tasks. Across both studies, results did not show any evidence for interference of the perceptual awareness task. We conclude that administering this task during the learning phase of an EC study does not constitute a threat to its validity. However, we are careful to limit this conclusion to the present set of learning conditions (i.e., valence-focus orienting task, predictable CS-US sequence). Under other conditions (e.g., incidental conditions such as those realized in the surveillance paradigm; Olson & Fazio, 2001), administering PAS ratings may yet be found to affect EC (for instance via the additional attention on the CS that is afforded by the PAS task; see e.g., Avneon and Lamy (2018)].

A confounded on-line identification task can induce EC. Administering an on-line CS identification task allows for fine-grained tracking of objective CS visibility. However, it may not only serve as a measure, but it may also intervene with the cognitive processes during learning. In particular, the presentation of a set of CS stimuli as response options, immediately after the US, may itself constitute a learning opportunity because the CSs presented as response options are repeatedly paired with USs. This may induce (or interfere with) evaluative learning for those CSs: If the CS options are systematically presented on trials with USs of a specific valence (i.e., if CS options are correlated with US valence), they may acquire the USs' valence, producing an unintended, or artificial, EC effect. If, on the other hand, the CS options are presented equally often with USs of each of the two valences (i.e., if CS options are uncorrelated with US valence), an artificial EC effect is unlikely to obtain, but these additional pairings may disrupt or mask the effects of the critical pairings in the CSs' learning trials.

Here we found evidence for artificial EC in the correlated condition but no interference with true EC in the uncorrelated condition. Across two studies, when correlated with US valence, CSs presented as response options showed artificial EC: Distractor stimuli showed EC after being pseudo-paired (i.e., systematically presented as response options on trials) with USs of a specific valence. The artificial EC effect due to pseudo-pairings was modulated by the time participants took to process the response options. This demonstrates the necessity to avoid valence-correlated response options on the identification task.

Experiment 2 showed that the artifact is removed when uncorrelated response options are used. Importantly, the use of uncorrelated response options has not been found to interfere with EC effects: Regardless of CS presentation duration, EC effects in the condition with uncorrelated response options were not reduced when compared to the conditions without an on-line identification task. The lack of such an interference effect confirms our previous finding that the presence versus absence of an uncorrelated on-line identification task does not affect EC (Stahl et al., 2016, Exp. 2).

On-line awareness ratings versus memory-based proxies. As a secondary goal, we were interested in the relation between on-line awareness measures and memory-based awareness proxies. As expected, we found that on-line awareness measures are more sensitive to awareness at encoding than subsequent memory measures (e.g., Shanks & St. John, 1994). Given that on-line measures did not interfere with learning, one may prefer such measures because they tap more directly what is experienced during learning. On the other hand, it has been argued that EC is determined by memory retrieval (Gast, 2018; Stahl & Aust, 2018), and that memory for CS-US pairings at retrieval should therefore be a better predictor of EC than awareness at encoding (e.g., Gast, Gawronski, & De Houwer, 2012). We recommend relying on the on-line perceptual awareness task for studying the role of awareness at encoding. This would free the memory-based measures from serving as (often inappropriate) proxies for awareness at learning, and allow for their use in studying the role of memory for CS-US pairings at retrieval. Disentangling the two constructs, and using separate measures to assess them, may help move EC research forward.

Limitations and open questions

The present studies are limited because they rely on data that were aggregated across (trials and) items; and because the results may depend on the type of mask, the type of ongoing task, and the specific set of learning conditions and dependent measures implemented in the present studies.

A more fine-grained perceptual-awareness manipulation is needed. While awareness data were collected on the trial level, the present studies focused on EC under three presentation condition categories (i.e., 30-40 ms, 50-60 ms, and 70-80 ms). Within these categories, EC data were aggregated across trials with different CS presentation durations, as well as across items, and statistical tests were performed at the group level. Given this level of aggregation and the resulting wide range of awareness scores within each category, the present studies were clearly not designed to account for the role of interindividual differences in perception (nor inter-item differences in perceivability). Such interindividual differences may have resulted in some participants being able to consciously perceive some stimuli in the brief presentation conditions (and may have resulted in a marginally significant EC effect for 50-60 ms CSs in the *both/uncorrelated* condition of Experiment 2). Future research should study more fine-grained distinctions, and take into account individual differences, for example via individual adaption of presentation durations and computing of individual awareness thresholds. An adaptive approach would also be a useful basis for investigating whether a reduction of attention towards the CS (i.e., as expected under more incidental conditions) could be compensated via an increase in stimulus strength while keeping perceptual awareness constant. Studies aiming at a more fine-grained assessment should also consider using mixed-model analyses that are capable of taking into account differences between participants and items (and even trial-level effects).

Results may depend on the type of masking. In the present studies, as well as our previous work, we have used pattern masks to interfere with perceptual processing of the CSs. While pattern masks are known for interrupting at an earlier processing stage, other types of masks (e.g., meta-contrast) allow low-level perceptual processing to continue, and may therefore be more likely to spare unconscious perception (Breitmeyer, 2015). Indeed, a recent study investigating semantic priming found that, while priming depended on prime visibility with pattern masks, priming was independent of prime visibility with meta-contrast masks (Wernicke & Mattler, 2019). It is thus conceivable that EC may be found independently from (or in the absence of) consciousness when meta-contrast masking is used.

The type of ongoing task may influence encoding. The present results may be restricted to the set of tasks that were performed during learning. For instance, learning effects may depend on the quality of encoding required by the type of ongoing task (e.g., Nieuwenstein & Kromm, 2017): When participants processed pictorial information in anticipation of one of three different discrimination tasks (i.e., between different states of a specific exemplar; between different exemplars; or between different categories), they tend to encode visual stimuli at the specific level of detail necessary to perform the task; that is, they fail to encode additional detail if not required for the anticipated discrimination. Applied to the present studies, this notion may help explain the trend towards higher PAS ratings in the *subjective* as compared to the *both* conditions: If the PAS tasks requires a qualitatively different type of encoding than the identification task, then participants in the *subjective* group were able to optimize their encoding on each trial, whereas in the *both* group participants were unable to predict the upcoming task type, and so were unable to adapt CS encoding to the type of task. A second finding is relevant here: In Exp.2, a marginally significant EC effect was obtained for CSs presented for 50/60 ms when participants performed the (uncorrelated) identification task. This may be due to qualitative differences in processing of the CSs induced by the identification task in the *both/uncorrelated* condition (as compared to, e.g., the condition without awareness task): The level of detail that needed to be encoded for this task was presumably higher than that for the valence judgment task (which can be performed by reference to the US alone), and participants' anticipation of this requirement may have affected their attentional allocation during encoding. Yet, the present study was not designed to investigate such effects, and the discussed findings are at best suggestive; it remains to be seen whether such task type effects can be more robustly demonstrated in EC.

Results may depend on the set of learning conditions. The present findings should not be taken as conclusive evidence against consciousness-independent EC because they were obtained under specific set of learning conditions; they should be replicated and extended to different sets of learning conditions. Some authors have argued that the present conditions may have been detrimental to unconscious EC (Jones et al., 2009; Sweldens, Van Osselaer, & Janiszewski, 2010). For instance, the instructional focus on US valence (which was implemented here to increase the chances of obtaining EC; Gast & Rothermund, 2011) may have induced an analytic mindset and thereby suppressed a more heuristic evaluative processing mode (but see Heycke et al., 2017; Heycke & Stahl, 2018). Similarly, the pairwise presentation of CS and US may have informed participants's guesses about the goal of the study; a less predictable presentation mode may be more efficient in avoiding demand effects. Furthermore, the use of nonwords as CSs may be argued to favor propositional or analytic processes; while unconscious learning has been successfully demonstrated for nonwords (Greenwald & De Houwer, 2017), and although nonwords arguably have ecological validity in marketing contexts, it may still turn out that automatic associative processes may be more likely to operate on holistically processed material such as images of cartoon characters or human faces. Finally, it is not clear whether a simultaneous presentation of CS and US is indeed conducive to EC with brief and masked CSs, but this is an empirical question that remains to be investigated (Heycke et al., 2017). Taken together, the present studies arguably realized conditions at the intentional end of the spectrum: CS-US pairs were presented in a predictable event sequence, and participants' attention was focused on (the CS and) the valence of the CS-US pair. It remains possible that unconscious EC operates under more incidental learning conditions.

It is worth briefly discussing the findings by Greenwald and De Houwer (2017) in this context. In their studies, nonword CSs were presented for (a range of) brief durations and were subsequently masked, such that identification was at chance for a substantial proportion of participants. Immediately after the masked CS, a target (a valent word or a first name) was presented that was to be categorized (as positive/negative or male/female). Importantly, in contrast to the present studies, a response-window procedure was used to enforce fast categorization responses (i.e., between 333 and 467 ms). In the learning phase, a CSs was always associated with one of the categories (i.e., 100% contingency), and participants unconsciously learned this association. This was evident in a test phase with 50% contingency in which presentation of the CS facilitated the associated classification response. Importantly, this learning effect occurred even with chance-level identification performance; furthermore, the learning effect was uncorrelated with identification performance. While these findings clearly support the notion of unconscious learning, it is unclear as to whether they support unconscious *evaluative* learning in the sense of a learning effect that (generalizes beyond the learning task and) affects participants' direct or indirect evaluations of the CSs. Initial findings from our lab suggest that the unconscious learning process does not affect evaluative measures, and that awareness is required for evaluative learning also in that paradigm.

Conclusions and outlook

The present results are in line with recent work showing that EC is absent for brief and masked CSs, and that participants' ability to objectively identify and consciously perceive these CSs is a necessary (but not sufficient) precondition for EC (Heycke et al., 2017, 2018; Heycke & Stahl, 2018; Stahl et al., 2016). More broadly, they are also in line with a recent review on the automaticity of EC, which discusses the different features of automaticity and concludes that there is little evidence for automatic associative attitude learning (Corneille & Stahl, 2019). Taken together, there is currently little evidence for EC without (or independently of) awareness. Future studies—perhaps along the lines suggested above—may yet obtain EC for objectively above-chance but subjectively unaware CSs, but the evidence available so far suggests that perceptual awareness of CSs (and, presumably, of the CS-US co-occurrence) is a necessary precondition for evaluative learning.

Arguably, the present findings lend some support to the notion that EC may operate under partial or degraded forms of awareness. First, we show that EC can be reliably found with brief (i.e., 70 ms) presentation durations and given limited perceptual awareness of the CSs features (i.e., mean PAS ratings below 5). Under such conditions, only stimulus parts have entered awareness (Fazekas & Overgaard, 2018), while the rest of the stimulus perhaps remains in a state that has been described as "fleeting awareness" (Crick & Koch, 1990, 2003). If we hold that EC involves the entire CS stimulus, EC under such conditions can be said to involve at best fleeting or partial awareness of CSs (Kouider & Dupoux, 2004; Timmermans et al., 2010). Consistent with this notion of degraded awareness, we found that under these conditions participants may sometimes fail to report the source of their evaluative change (i.e., they are unable to identify the US) at the end of the study. It is an open question, however, whether the EC effect under such conditions is robust and replicable, and if so, whether it is any more stable over time than the memory for the US and/or its valence. A different interpretation may be proposed that focuses on the stimulus feature that has entered awareness: If the EC effect is linked only to the consciously perceived feature, it may generalize to other CSs sharing that feature, while the rest of the CS stimulus (that has not been consciously perceived) is not involved in the representation and should not affect EC. This is an empirical question that could be investigated by systematic manipulation of (more or less easily perceived) CS features and their association with US valence.

There is another sense in which EC may be said to operate in the absence of awareness: It has repeatedly been shown that EC can obtain incidentally and in the absence of memory for the source of the acquired valence, especially when attention to the source (i.e., the US and its pairing with the CS) is reduced and memory is held at low levels, as in the *surveillance* paradigm (Olson & Fazio, 2001; Stahl & Heycke, 2016). In this paradigm, participants are instructed to monitor a purportedly random stream of images and report the presence of a specific target stimulus. EC effects were repeatedly obtained under such conditions, but participants are assumed not to notice the presence of systematic CS-US pairings. Together with the subsequent lack of source memory for the pairings, this implies an inability to control for the influence of these pairings on CS evaluations. Even if EC for brief and masked CSs without perceptual awareness should turn out to be a reliable phenomenon, it is likely to be of relatively little consequence given its current ecological rarity (although future technical developments may soon require reassessment). Incidental encoding situations are probably much more frequent, and even if effects are small, over time and repetitions they may have considerable impact on cognition and behavior. In applied research on undesirable influences on behavior, researchers should focus at this point on effects of incidental exposure to clearly visible advertisements in distracting environments, as such conditions interfere with monitoring the sources of these influences and thereby limit our abilities to mitigate their effects (Biegler & Vargas, 2013, 2016).

Chapter 3

Why a Standard IAT Effect Cannot Provide Evidence for Association Formation: The Role of Similarity Construction

Moran and Bar-Anan (2013) demonstrated that evaluations on a direct measure reflected information on both US valence and CS-US relations, whereas evaluations on an indirect measure (IAT) reflected only information on US valence. This dissociation between measures supposedly tapping into propositional and associative processes apparently supports dual process models of EC. In the present study, we present an alternative explanation of this pattern, based on an interpretation of IAT effects in terms of flexible similarity construction processes. According to this account, processing draws on those features that discriminate between target categories, and help to align targets with attributes in the compatible block. Across two experiments, we consistently found that IAT effects did not reflect rigid associations, but instead depended on whichever information could be used for similarity constructions between targets and attributes in different variants of the IAT. The findings are discussed with regard to theoretical models of EC as well as in reference to prominent accounts of IAT performance.

Evaluative conditioning (EC) refers to a change in valence brought about by stimulus pairings (De Houwer, 2007). In a typical EC procedure, neutral “conditioned” stimuli (CS) are repeatedly paired with valent “unconditioned” stimuli (US). The most common finding is a spread of valence from the US to the CS; that is, CSs paired with positive USs are evaluated more favorably than CSs paired with negative USs.

Theoretical accounts of EC

Early accounts of EC were based on the notion of *association formation*, that is, the automatic formation of mental associations between stimuli caused by their mere co-occurrence in time and space (Baeyens et al., 1992). These associations are assumed to be relatively stable, strengthened through repetition, and stored in semantic memory. From this perspective, observing the co-occurrence of CS and US results in their simultaneous activation, which in turn creates a mental link between their mental representations. Upon later confrontation with the CS alone, the US is co-activated through this association, and affects behavior (evaluative responses) towards the CS.

Another account of EC champions the concept of *propositional learning* as the sole mediator of EC (De Houwer, 2009, 2018). Propositions are statements about events and their relationship to one another that are qualified by truth values, that is, whether they are deemed to be true or false. From this perspective, EC occurs when observing the CS-US pair triggers the formation of a conscious propositional representation about their relationship. While the process of proposition formation is assumed to be non-automatic, its outcome (i.e., propositional knowledge about the CS-US relation) may determine the future (dis-)liking of the CS via non-automatic as well as automatic processes.

Although the debate on whether all instances of EC can be explained by propositional learning is still ongoing, it seems safe to say that propositional processes play a crucial role in many empirical demonstrations of EC (for reviews, see Corneille & Stahl, 2019; Hofmann et al., 2010). Therefore, purely associative accounts of EC are currently off the table, and have given way to dual process models of evaluative learning (Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006) as the main competitor of purely propositional accounts. The shared tenet of dual process models is that both associative and propositional learning mediate the effect of CS-US pairings on CS evaluation. Their relative contribution to a given instance of EC is assumed to be determined by various context factors that relate to the conditions under which the CS-US pairings are administered and under which the CS evaluation is assessed. Put in a nutshell, dual process models claim that associative processes will dominate under conditions of automatic processing, whereas propositional processes will dominate under conditions of controlled processing.

Measures of EC effects

Initially, EC effects were usually assessed on direct measures of evaluation, by asking participants to evaluate the CSs on a bipolar rating scale. While this approach is still the most widely used, indirect measures of evaluation — such as the evaluative priming task (EP; Fazio, Sanbonmatsu, Powell, & Kardes, 1986), or the Implicit Association Test (IAT; Greenwald et al., 1998) — have become increasingly popular tools in EC research. The early argument for this methodological shift was based on the assumption that participants were less capable of controlling their evaluative responses on indirect than on direct measures. Hence, their use was seen as a way of establishing EC effects as a genuine change in liking, ruling out explanations in terms of mere demand effects (Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002).

The second reason for using indirect measures in EC research lies with their potential to foster theoretical advancement. According to dual process accounts, direct measures typically provide the opportunity for controlled processing and should thus reflect the result of propositional reasoning. This makes predictions for direct measures (oftentimes) indistinguishable from what is expected on the basis of purely propositional accounts.[14] By contrast, indirect measures are assumed to operate under automatic processing conditions and should therefore be immune against influences of propositional reasoning during measurement.

[14] Note that this assessment refers to direct measures that do not restrict controlled processing. By contrast, using a direct measure that imposes time pressure and/or requires participants to base their judgement on spontaneous feelings may allow for diverging predictions by dual process vs. propositional accounts.

However, demonstrating EC on indirect measures cannot — by itself — provide sufficient evidence for the role of association formation since the automatic evaluation reflected by the indirect measure could in principle be based on knowledge that was acquired in a propositional manner during the learning phase (De Houwer, 2006).

To address this problem, a test of association formation via indirect measures is now typically embedded in *US Valence* × *CS-US Relation* designs that allow for diverging predictions by propositional vs. dual process accounts. These designs draw on the concept of relational information that specifies the meaning of the co-occurrence of CS and US in an EC procedure. The classic EC paradigm does not provide such information (at least, not explicitly), but modified versions instruct participants to conceive of the pairings as instances of (dis-)similarity, or complement the CS-US pairings with relational qualifiers (e.g., cause vs. prevent) on a trial-by-trial basis.

The appeal of such modified EC procedures lies in the differential susceptibility of associations and propositions to this additional information. An EC effect mediated by propositional learning should reflect the interplay of information on US valence and information on CS-US relations. When the relational qualifier supports the evaluative implications of the CS-US pairing (e.g., CS causes US), propositional accounts predict a regular EC effect. Yet, when the evaluative implications of the CS-US pairing are inverted by the relational qualifier (e.g., CS prevents US), a reversed EC effect is predicted. By contrast, associations are conceptualized as the automatic product of co-occurrence, and therefore cannot incorporate additional information that specifies the meaning of this co-occurrence. EC effects mediated by association formation should thus be unaffected by relational qualifiers; specifically, CSs presented with inverting relational qualifiers should nevertheless be evaluated in line with their respective US valence.

Hence, although both accounts predict an influence of relational qualifiers on EC for direct measures, only a purely propositional account would predict a similar influence for indirect measures. According to dual process accounts, indirect measures provide an opportunity to unravel associative processes that are assumed to operate in parallel with propositional processes during the EC procedure. Therefore, these accounts predict a simple main effect of US valence on indirect measures that is not qualified by relational information.

Evidence for association formation

A recent study by Moran and Bar-Anan (2013) combining a *US Valence* × *CS-US Relation* design with an indirect measure provided the "most compelling evidence for dual process accounts" so far (Hu et al., 2017), and posed a serious challenge to purely propositional accounts of EC. In this study, participants were introduced to four families of alien creatures which differed by skin color. They were told that each family performed one of four actions (starting a pleasant melody, ending the pleasant melody, starting a horrifying scream, end ending the horrifying scream), and were instructed to identify and memorize which family performed which action during the upcoming learning procedure. After the learning procedure, each participant performed two IATs: a "starting" IAT contrasting the "starting melody" with the "starting scream" family, and an "ending" IAT contrasting the "ending melody" with the "ending scream" family. A direct measure of evaluation yielded the interaction of *US Valence* and *CS-US Relation* that is predicted by both accounts of EC: the "starting melody" was preferred over the "starting scream" family, and the "ending scream" was preferred over the "ending melody" family. By contrast, both IATs indicated a preference for the "melody" over the "scream" family. This pattern refutes the propositional account of EC (which predicts a reversed EC effect in the "ending" IAT), and supports dual process accounts. The latter explain (what we will call) the "ending" effect as the product of association formation: the temporal overlap of the "ending" families and their respective sounds automatically created evaluative associations in semantic memory, which then show up in the "ending" IAT despite participants' conscious knowledge about the inverting relationship between the families and sounds.

What does the IAT measure?

According to the standard view, IAT performance is driven by associations in semantic memory between the target and attribute categories included in a given IAT. Accordingly, an advantage in classification speed when one target category shares a response key with positive attributes, and the other target category shares a key with negative attributes (in comparison to the reversed key assignment) is attributed to relatively more positive associations for the first category than for the second category.

Even though many IAT effects can be explained by this *association account*, there is strong evidence that IAT effects can be driven by other factors, such as salience asymmetry or perceptual similarity (for a review, see Wentura & Rothermund, 2007). These findings have led to the development of the *similarity account* that assumes that IAT performance is driven by momentary constructions of similarity between targets and attributes (De Houwer et al., 2005). *Construction* means that a priori or ad hoc information about the target stimuli (or categories) is used in order to align them with the attribute categories. *Momentary* is supposed to convey that the construction of target-attribute similarity is flexible and depends on whatever information is made salient by a given IAT or the context in which the IAT is administered. This conception cannot only explain a wide range of IAT effects based on semantic similarity, salience, and perceptual features, but also accounts for their well-known malleability (Blair, 2002).

An alternative account of the "ending" effect

According to its original interpretation, the "ending" effect is driven by associations between the families and sounds. However, the "ending" effect can also be accounted for by an understanding of flexible similarity construction processes in the IAT. During the learning procedure, participants received two types of information about each family: information on

US valence and information on CS-US relations. While they certainly used both types of information when evaluating the families on the direct measure, they may have used only a subset of their knowledge when performing the IATs. Both “starting” and “ending” IATs compare creature families that share their CS-US relation, but differ with regard to their US valence. In this situation, focusing on US valence and ignoring CS-US relations is the easiest and most efficient way to discriminate between the targets and to derive target-attribute similarities in the compatible block of the “ending” IAT. First, this similarity construction requires processing of only one source of information and secondly, it avoids confusion between the target categories due to an overlap in CS-US relations. According to this explanation, the “ending” effect does not necessarily reflect evaluative associations but instead reflects the outcome of a specific similarity construction that is adaptive and efficient in the given circumstances. Importantly, however, this explanation is perfectly compatible with a possible influence of CS-US relations that might show up in a different IAT variant. Such an influence is predicted whenever target-attribute similarity based on the meaning of the action is emphasized by, or the best option in a given IAT.

Aims of the present study

The purpose of the present study was to demonstrate that information on US valence, information on CS-US relations, as well as the combination of both sources of information can impact IAT performance, and that their relative prominence depends on which type of information is useful in a given IAT. Thereby, we seek to address two different, yet related questions. First of all, we want to contribute to the theoretical debate over propositional vs. dual process accounts of EC by scrutinizing the evidence for association formation put forward by Moran and Bar-Anan (2013). Our second aim is to make a general point about the IAT, which is particularly relevant to researchers who want to use indirect measures of evaluation to investigate the mechanisms and boundary conditions of EC. Specifically, we seek to corroborate the validity of the flexible similarity construction account, and want to encourage other researchers to treat it on an equal footing with the widely adopted association account when designing studies and interpreting results.

Methodological remarks

In the present study, we calculated and compared family-specific D scores (D_f) for each family in a given IAT.[15] The calculation of D_f scores for a given family and IAT was based

[15] The reported results are based on the D_3 measure. For this measure, all trials from practice as well as main blocks with a response latency greater than 1000ms are excluded and response latencies on error trials are replaced by the sum of the block mean of correct responses and the block standard deviation of correct responses.

on the subset of trials in which a member of the respective family had to be classified.[16] This approach is akin to previous studies in which IAT data was analyzed separately for target and attribute trials (Brendl, Markman, & Messner, 2001), or for subsets of target trials (J. De Houwer, 2001; De Houwer et al., 2005; Gast & Rothermund, 2010). Since we are not interested in interpreting absolute IAT effects for single families but instead want to compare families with respect to their IAT effects, this is a perfectly legitimate way to analyze the data, and allows us to elegantly map specific patterns of results to theoretical hypotheses (see below). In order to compare our findings with those from Moran and Bar-Anan (2013), we also analyzed "overall" D scores (D_o, again based on the D_3 measure), which are reported in the online supplement.

Experiment 1

The primary aim of this experiment was to demonstrate that the "ending" effect does not provide unambiguous evidence for association formation, and can instead be explained by a flexible use of available information as predicted by the similarity account of the IAT. In order to assess whether the "ending" effect reflects stable associations between the "ending melody" ("ending scream") family and positive (negative) concepts or, alternatively, a momentary construction of "melody" ("scream") and positive (negative) concepts as being similar, we sought to create conditions for which the two accounts make diverging predictions. We therefore designed two new IATs each of which compared the families that shared a US valence but differed regarding their relationship with their common US. Specifically, the "melody" IAT compared the "starting melody" family to the "ending melody" family, and the "scream" IAT compared the "starting scream" family to the "ending scream" family (for an overview of possible IATs, see Figure 3.1). Given that a) the "starting" and "ending" families of a given US valence were equally associated with this valence (as posited by dual process accounts and implied by the parallel "starting" and "ending" effects reported by Moran and Bar-Anan [2013]), both the "melody" and "scream" IAT should yield null effects as the equally positive (or negative) associations of the two families in a given IAT would render the compatible and incompatible IAT blocks equally (in-)compatible. If, by contrast, the "ending" effect was based on a momentary similarity construction of "melody" ("scream") as positive (negative), this construction would no longer be helpful in the "melody" and "scream" IATs as US valence was constant for both target families within these IATs, and thus could not be used to construe target-attribute similarities that differentiate between the two target categories. Because similarity construction is assumed to be flexible and dependent on whatever information is salient in or most useful for a given IAT, we expected participants to instead consider the CS-US relations, and base target-attribute similiarity on

[16] Other than that, we followed the specifications provided by Greenwald, Nosek, and Banaji (2003). For each family and IAT alike, the D_f scores were calculated by subtracting the mean response time for trials in which the respective family shared a response key with positive attributes from the mean response time for trials in which it shared a key with negative attributes. This difference was then divided by the pooled standard deviation of both "positive block" trials and "negative block" trials for the respective family and IAT.

the meaning of the actions performed by the families. Based on the valence of their respective actions, we expected the "starting melody" ("ending scream") family to be construed as similar to positive concepts, and the "starting scream" ("ending melody") to be construed as similar to negative concepts. This would yield a preference for the "starting" ("ending") family over the "ending" ("starting") in the "melody" ("scream") IAT. For participants performing these two IATs (thereafter, the "valence of action" [VoA] group), we therefore predicted an interaction effect of *US Valence* and *CS-US Relation* on the D_f scores. In order to ensure that results in the "VoA" group were based on a learning experience that was comparable to Moran and Bar-Anan (2013) we implemented the same learning procedure and also ran a group of participants that were required to perform the original "starting" and "ending" IATs (thereafter, the "valence of sound" [VoS] group). We expected to replicate the original pattern, i.e., a preference for the "melody" family over the "scream" family in both IATs. For D_f scores, this translates to a main effect of *US Valence*, and the lack of an interaction of *US Valence* and *CS-US Relation*.

Method

Design. The study followed a 2 (*IAT Type*: "VoS" vs. "VoA") × 2 (*US Valence*: "positive" vs. "negative") × 2 (*CS-US Relation*: "starting" vs. "ending") mixed design with the first factor varying between participants. Furthermore, IAT order, block order of the first and second IAT, as well as the assignment of creature families to the four actions were counter-balanced using a Latin square.

Participants. For the Latin square, a total of sixty-four participants were needed (thirty-two per level of *IAT Type*). After the data of these participants were collected (and before analyzing evaluative ratings and IAT data), it was checked whether participants correctly remembered the actions of all four creature families (see section *Procedure*). Due to incomplete memory, the data of four participants were excluded and replaced. The final sample consisted of $N = 64$ University of Cologne students, who received partial course credit or monetary compensation for their participation (87.5% female, $M_{age} = 24.8$ years, $SD_{age} = 7.2$ years). The minimum sample size for such a balanced design ($N = 64$) provides sufficient power to detect medium effects ($d = .50$, one-tailed testing) within each *IAT Type* group ($1 - \beta = .87$), but has less power to test between-groups comparisons of IAT effects ($1 - \beta = .63$).

Materials. We used the same visual, auditory and verbal stimuli as the original study. The 12 positive and negative words (that served as attribute stimuli in the IAT) were translated into German.

Procedure. Participants were told that the experiment was about four creature families, which differed by color and head shape, and that each performed one of four actions: starting a beautiful melody, stopping the melody, starting a horrifying scream and stopping the scream. They were instructed to identify and memorize which family performed which action. The subsequent learning procedure used the same presentation parameters (number of trials, sound duration, temporal overlap of creatures and sounds) as the original study (Moran &

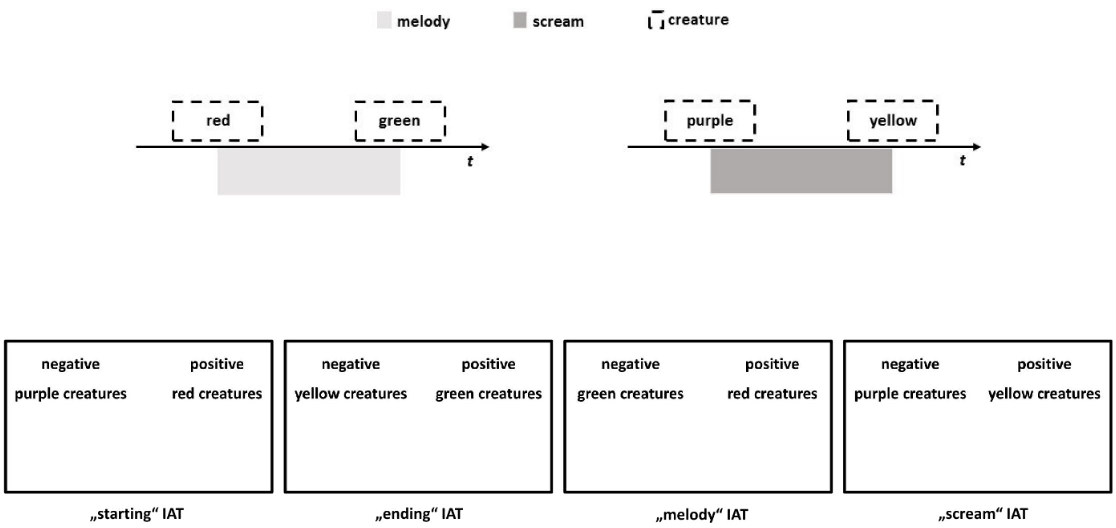

Figure 3.1. Upper half: Schematic depictions of a 'melody' trial (left) and a 'scream' trial (right) during the learning procedure. Lower half: Schematic depiction of the computer screen during the compatible blocks of the four IATs implemented in Exp.1. The white 'x' indicates the location of the to-be-classified target and attribute stimuli.

Bar-Anan, 2013). After the learning procedure, each participant performed two IATs the type of which depended on *IAT Type*. Each IAT started with a practice block comprising 21 trials whereof 10 trials showed members of one creature family present in the given IAT and the other 11 trials showed members of the other creature family. Which of the two families received 10 vs. 11 trials was determined randomly for each participant and IAT anew. Subsequently, participants worked through a second practice block of 21 trials whereof 10 trials showed words of one valence (negative or positive) and 11 trials showed words of the other valence respectively. Which valence received 10 vs. 11 trials was determined randomly for each participant and IAT anew. Next came a combined practice block of 17 trials whereof 5 trials showed words of one valence (negative or positive), 4 trials showed words of the other valence, 4 trials showed members of one creature family and 4 trials showed members of the other creature family. Which valence received 4 vs. 5 trials was determined randomly for each participant and IAT anew. The first trial of this block always showed a word of the valence that received 5 trials. The combined practice block was followed by a test block of 49 trials whereof 13 trials showed words of one valence (negative of positive), 12 trials showed words of the other valence, 12 trials showed members of one creature family and 12 trials showed members of the other creature family. Which valence received 12 vs. 13 trials was determined randomly for each participant and IAT anew. The first trial of this block always showed a word of the valence that received 13 trials. Subsequently, the practice block for the creature families was repeated with reversed response key assignments. Finally, participants worked through the combined practice block and the test block with reversed response key assignments for the two creature families. Prior to each block, participants were informed about the group of stimuli they would have to classify in the upcoming block and which response keys they were supposed to use for classification (see online supplement for exact wording of IAT instructions). These instructions referred to the creature families by their color and did not mention the actions the respective families had performed during the learning procedure (see online supplement for the exact wording of IAT instructions). During each block, the respective category labels (positive, negative, two of the following: red creatures, green creatures, yellow creatures, purple creatures) were presented in the corresponding top corner of the screen. For each participants, two of the four creature families appeared in first IAT, and the two other creature families appeared in the second IAT. The creature families' colors and performed actions differed between participants and depended on the Latin square that assigned the colors to the actions as well as on the level of *IAT Type* that a given participant belonged to. The two IATs were separated by a self-paced break. Following completion of the second IAT, participants were first asked to evaluate each creature family on an 8-point scale (1 = *very negative*, 8 = *very positive*). Afterwards, they had to indicate which creature family performed which action during the learning procedure. Finally, participants had to state whether they deemed ending the melody (scream) a negative (positive) action.

Results

Evaluative Ratings. A 2 (*US Valence*) × 2 (*CS-US Relation*) × 2 (*IAT Type*) mixed-design ANOVA on the evaluative ratings of the four families revealed a replication of the findings by Moran and Bar-Anan (2013) indicating the full integration of information on US valence and CS-US relations in line with propositional reasoning. Details can be found in the online supplement.

IAT effects. The D_f scores[17] were submitted to a 2 (*IAT Type*) × 2 (*US Valence*) × 2 (*CS-US Relation*) mixed-design ANOVA. Figure 3.2 depicts the mean D_f scores conditioned on these three factors. The ANOVA revealed a significant main effect of *US Valence*, $F(1,62) = 11.67$, $MSE = 0.23$, $p = .001$, $\hat{\eta}^2_G = .034$, indicating overall more positive D_f scores for "melody" families (M = 0.11, SD = 0.59) than for "scream" families (M = -0.10, SD = 0.55). The effect of *US Valence* was further qualified by significant two-way interactions of *US Valence* and *IAT Type*, $F(1,62) = 7.21$, $MSE = 0.23$, $p = .009$, $\hat{\eta}^2_G = .021$, and *US Valence* and *CS-US Relation*, $F(1,62) = 3.96$, $MSE = 0.50$, $p = .051$, $\hat{\eta}^2_G = .025$, as well as by a marginally significant[18] three-way interaction of *US Valence*, *CS-US Relation*, and *IAT Type*, $F(1,62) = 2.61$, $MSE = 0.50$, $p = .111$, $\hat{\eta}^2_G = .017$. The remaining main

[17] In this and the following experiment, positive (negative) family-specific D scores indicate that classification was faster when the family shared a response key with positive (negative) attributes.

[18] Note that because the shape of this interaction is predicted by our account of the IAT effects, and because of the equivalence of t tests and F tests with one degree of freedom in the numerator (Maxwell & Delaney, 1990), the three-way interaction of *US Valence*, *CS-US Relation*, and *IAT Type* can be considered marginally significant by a one-tailed test (p = .056).

effects and interactions did not reach significance, all ps $\geq .335$.

To follow up on the aforementioned three-way interaction, we calculated separate 2 (*US Valence*) $\times$ 2 (*CS-US Relation*) repeated measures ANOVAs for each level of *IAT Type*. The ANOVA for the VoS group revealed the predicted pattern: an unqualified main effect of *US Valence*, $F(1,31) = 12.43$, $MSE = 0.34$, $p = .001$, $\hat{\eta}^2_G = .120$, reflecting more positive D_f scores for the "melody" (M = 0.19, SD = 0.47) than for the "scream" families (M = -0.18, SD = 0.53). The other main effects and interactions did not reach significance, all ps $\geq .490$. The results in the VoS group thus replicated the findings by Moran and Bar-Anan (2013): EC effects as measured by the "starting" and "ending" IATs solely reflect information on US valence, and are unaffected by information on CS-US relations.

The ANOVA for the VoA group revealed a different pattern. The main effects of *US Valence* and *CS-US Relation* did not reach significance, all ps $\geq .472$. Instead, the predicted interaction of the two factors was found, $F(1,31) = 4.72$, $MSE = 0.69$, $p = .038$, $\hat{\eta}^2_G = .068$. This interaction reflected opposite simple main effects of *US Valence* for the two levels of *CS-US Relation*. Among "starting" families, D_f scores were significantly more positive for the "melody" (M = 0.22, SD = 0.65) than for the "scream" family (M = -0.14, SD = 0.58), $t(31) = 2.17$, $p = .019$ (one-tailed). Among "ending" families, the opposite was true: D_f scores were significantly more positive for the "scream" (M = 0.09, SD = 0.52) than for the "melody" family (M = -0.18, SD = 0.66), $t(31) = 1.85$, $p = .037$ (one-tailed). Hence, the findings in the VoA group demonstrate that EC effects as measured by the "melody" and "scream" IATs reflect the full integration of information on US valence and information on CS-US relations.

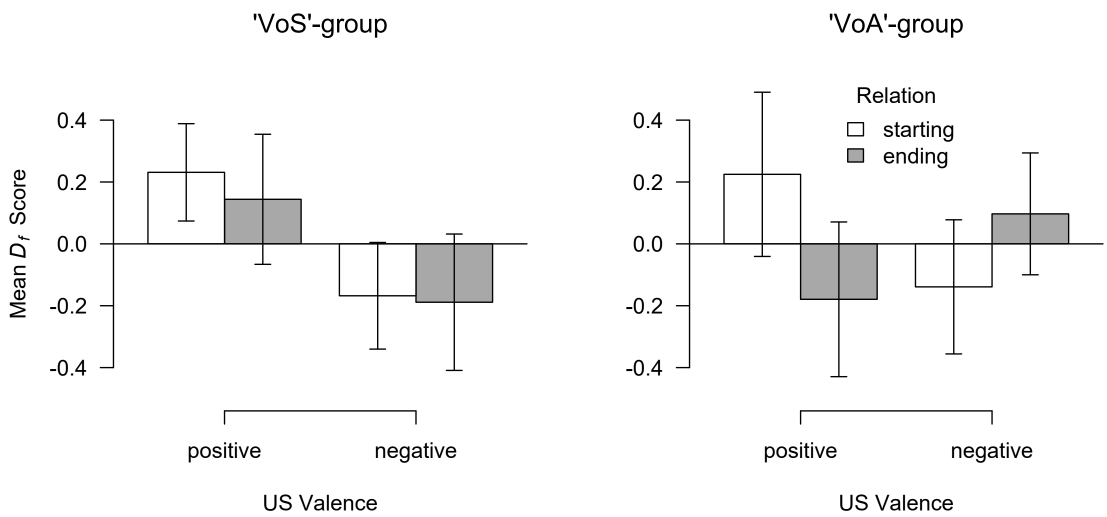

Figure 3.2. Family-specific D scores as a function of *US Valence*, *CS-US Relation*, and *IAT Type* (Exp. 1). Error bars represent 95% within-subjects confidence intervals.

Discussion

Experiment 1 demonstrated that the impact of CS-US relations on IAT performance depends on the type of IAT that is being used. In the "starting" and "ending" IATs, families were equal with regard to their CS-US relation, but differed in terms of their US valence. Here, stimulus classification was easier when the "melody" ("scream") family shared a key with positive (negative) attributes, irrespective of CS-US relations. In the "melody" and "scream" IATs, US valence was held constant across the two families that defined the target categories in the IAT, and families differed by their CS-US relation. Here, stimulus classification was easier when the family which performed the positive (negative) action shared a key with positive (negative) attributes. Since association formation is incapable of incorporating information on CS-US relations, the findings in the VoA group cannot be explained in terms of automatically acquired associations, and thus demonstrate that the IAT is sensitive to propositionally acquired knowledge.

In sum, the results of Experiment 1 indicate that both information on US valence as well as on CS-US relations influenced IAT effects. Which type of information was effective depended on the structure of the respective IAT that determined which type of information discriminated between the two target categories and could efficiently be used to align target and attribute categories in the compatible IAT block. Although the pattern of IAT effects within each of the groups exactly matched our predictions, the crucial three-way interaction just missed conventional criteria of significance. We therefore decided to conceptually replicate the study with a design that further extends the range of flexible similarity constructions.

Experiment 2

For the present study, we designed three IATs that again emphasized different types of target-attribute similarity. Departing from experiment 1, where only two creature families were contrasted in a given IAT, we now used all four families as target categories in each of the three IATs. Therefore, the IATs in the present study did not differ with regard to their target categories and stimuli (as in Exp. 1). Instead, the difference between IATs was now based on the way the four families were mapped onto the two response keys that were used for the classification task in the IAT (see Figure 3.3 for an illustration of the different IATs). For participants in the "valence of sound" (VoS) group, the IAT required them to press one response key whenever they saw a member of one of the two "melody" families, and to press the other response key whenever they saw a member of one of the two "scream" families. As in a traditional IAT, the two "melody" families ("scream" families) shared a key with positive (negative) attributes in one block, and with negative (positive) attributes in the other block. For participants in the "valence of action" (VoA) group, the IAT required them to press one response key whenever they saw a member of one of the two families that had performed a positive action (i.e., the "starting melody" and "ending scream" families), and to press the other response key whenever they saw a member of one of the two families that had performed a negative action (i.e., the "starting scream" and “ending melody” families). As in a traditional IAT, the two “positive action” families (“negative action” families) shared a key with positive (negative) attributes in one block, and with negative (positive) attributes in the other block. For participants in the “valence of relation” (VoR) group, the IAT required them to press one response key whenever they saw a member of one of the two “starting” families, and to press the other response key whenever they saw a member of one of the two “ending” families. As in a traditional IAT, the two “starting” families (“ending” families) shared a key with positive (negative) attributes in one block, and with negative (positive) attributes in the other block.

Designing these three IATs was motivated by two reasons. First, including all four families in each IATs raised comparability across IATs: each IAT consisted of the exact same categories and stimuli and only differed in how these stimuli were mapped to response keys. Second, the three IATs once again allowed us to test whether their effects are driven by rigid associations between the “melody” (“scream”) families and positive (negative) concepts, or, alternatively, by flexible similarity construction.

The association formation account makes the same prediction for all three IATs: namely, a preference for the “melody” families over the “scream” families that is unaffected by the relation between the family and the sound. For D_f scores, this translates to a main effect of *US Valence*, and the lack of an interaction of *US Valence* and *CS-US Relation* for all three IATs. This prediction is based on the fact that “melody” (“scream”) families are assumed to be equally associated with positive (negative) concepts and that in all three IATs the (hypothetical) net valence of the associations pertaining to the any of the families is more positive (for “melody” families) or more negative (for “scream” families) than the net valence of associations pertaining to the two families mapped onto the opposing response key.

By contrast, the similarity construction account makes diverging predictions for the three IATs. As the “VoS” IAT allows for a similarity construction based on the valence of the sound, the similarity construction account, here, predicts the same pattern as the association formation account: a main effect of *US Valence* on the D_f scores, and the lack of an interaction of *US Valence* and *CS-US Relation*. The “VoA” IAT does not allow for a similarity construction based on *US Valence*, but instead offers the valence of the performed action as a basis for aligning the two target categories with the attribute category. Therefore, an interaction of *US Valence* and *CS-US Relation* is predicted for this IAT. In the “VoR” IAT, neither the valence of the sound nor the valence of the performed action allows for a consistent similarity construction. We therefore expected participants to fall back on the information on *CS-US Relation* (which was constant for the two families assigned to the same response key) and to construe target-attribute similarity based on the (albeit weak) valence of the relational qualifiers themselves. Based on the fact that the concept of “starting” holds generally more positive connotations than the concept of “ending” (Eder, Rothermund, & Houwer, 2013), we expected IAT performance in the “VoR” group to indicate a preference for the “starting” families over the “ending” families. For D_f scores, this translates to a main effect of *CS-US Relation*, and the lack of an interaction of *US Valence* and *CS-US Relation*.

Method

Participants and design. A sample of $N = 120$ Friedrich Schiller University Jena students from different majors completed the experiment in exchange for a monetary compensation and a chocolate bar (62.5 % female, $M_{age} = 23.3$ years, $SD_{age} = 5.4$ years). The study followed a 3 (*IAT Type*: VoS vs. VoA vs. VoR) × 2 (*US Valence*: “positive” vs. “negative”) × 2 (*CS-US Relation*: “starting” vs. “ending”) mixed design with the first factor varying between participants. Furthermore, the order of IAT blocks as well as the assignment of creature families to the four actions were counter-balanced using a Latin square. The sample size for such a balanced design ($N = 120$) provides sufficient power to detect medium effects within each *IAT Type* group ($1 - \beta = .93$, $d = .50$, one-tailed testing), but slightly less power for the between-groups comparison of IAT effects ($1 - \beta = .68$, $f = .25$).
Materials. We used the original drawings of the alien creatures, but changed their colors to light pink, light yellow, light green, and light brown. This was done to prevent potential IAT task simplification based on a priori groupings of pairs of colors.

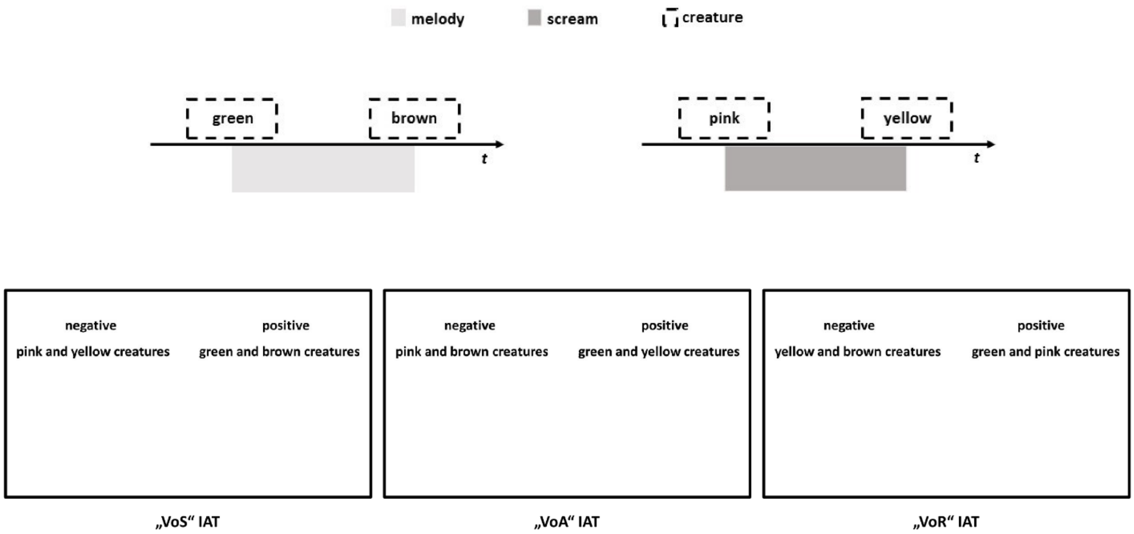

Figure 3.3. Upper half: Schematic depictions of a ’melody’ trial (left) and a ’scream’ trial during the learning procedure. Lower half: Schematic depiction of the computer screen during the compatible blocks of the three IATs implemented in Exp.2. The white ’x’ indicates the location of the to-be-classified target and attribute stimuli.

Procedure. We used the same instructions and learning procedure as in Experiment 1. After the learning procedure, each participant performed one of three IATs (see above). The number of trials was increased in order to compensate for the presence of four (instead of two) families in each IAT. The combined practice blocks consisted of 33 trials whereof 13 trials required a classification of an attribute, and 12 trials required a classification of a creature. The main blocks consisted of 130 trials whereof 66 trials required a classification of an attribute, and 64 trials required a classification of a creature. Prior to each block, participants were informed about the group of stimuli they would have to classify in the upcoming block and which response keys they were supposed to use for classification. Instructions *before* and category labels *during* the IATs referred to the families by their color (see the original publication's online supplement for exact wording). The category labels listed the two families that were assigned to the same key in one line (e.g., "pink and yellow creatures"), right below the respective attribute category (see Figure **??**). For each participant, the two creature families were randomly assigned to the first or second position of the category label. After completing the IAT, participants performed the same tasks as in Experiment 1 using the same measures as previously described. Finally, participants were asked whether the deemed ending the scream (melody) a positive, negative or neutral action.

Results

Evaluative Ratings. As in the first experiment, a 3 (*IAT Type*) × 2 (*US Valence*) × 2 (*CS-US Relation*) mixed-design ANOVA revealed a replication of the findings by Moran and Bar-Anan (2013). Details can be found in Appendix B.

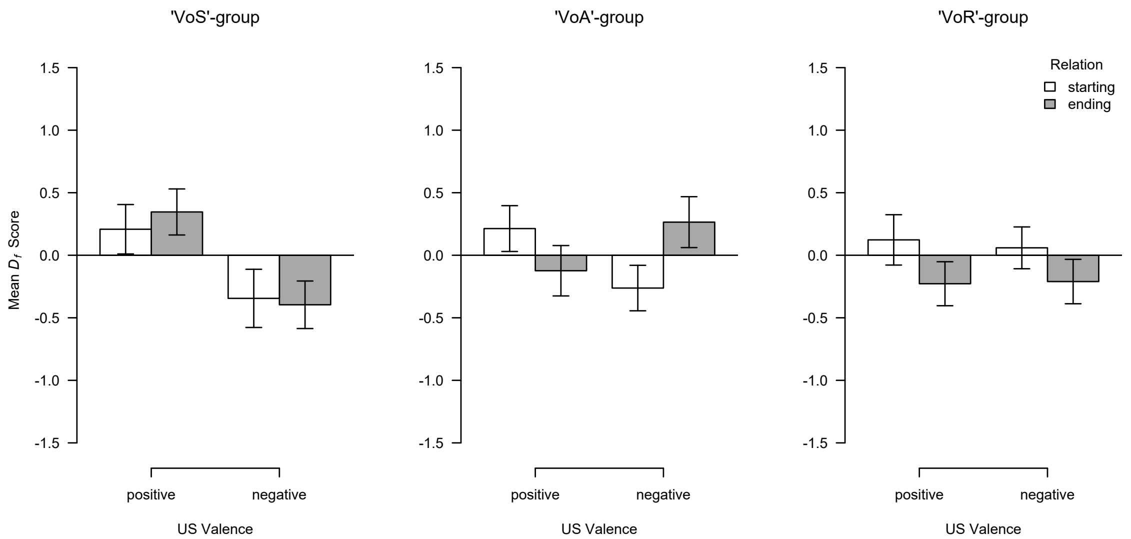

Figure 3.4. Family-specific D scores as a function of *US Valence*, *CS-US Relation*, and *IAT Type* (Exp. 2). Error bars represent 95% within-subjects confidence intervals.

IAT effects. D_f scores were submitted to a 3 (*IAT Type*) × 2 (*US Valence*) × 2 (*CS-US Relation*) mixed-design ANOVA. Figure 3.4 depicts the mean D_f scores conditioned on these three factors. The ANOVA revealed a significant main effect of *US Valence*, $F(1, 117) = 15.15$, $MSE = 0.45$, $p < .001$, $\hat{\eta}^2_G = .039$, significant two-way interactions of *US Valence* and *IAT Type*, $F(2, 117) = 11.21$, $MSE = 0.45$, $p < .001$, $\hat{\eta}^2_G = .056$, of *CS-US Relation* and *IAT Type*, $F(2, 117) = 5.47$, $MSE = 0.36$, $p = .005$, $\hat{\eta}^2_G = .023$, and of *US Valence* and *CS-US Relation*,

$F(1, 117) = 4.84$, $MSE = 0.39$, $p = .030$, $\hat{\eta}^2_G = .011$, as well as a significant three-way interaction of *US Valence*, *CS-US Relation*, and *IAT Type*, $F(2, 117) = 7.61$, $MSE = 0.39$, $p = .001$, $\hat{\eta}^2_G = .034$. All other main effects did not reach significance, all *p*s $\geq .25$.

To clarify the aforementioned three-way interaction, we calculated separate 2 (*US Valence*) × 2 (*CS-US Relation*) repeated measures ANOVAs for the three levels of *IAT Type*. The ANOVA for the VoS group revealed the predicted main effect of *US Valence*, $F(1, 39) = 15.87$, $MSE = 1.06$, $p < .001$, $\hat{\eta}^2_G = .212$, reflecting more positive D_f scores for the "melody" (M = 0.28, SD = 0.6) than for the "scream" families (M = -0.37, SD = 0.66). As expected, the main effect of *CS-US Relation* as well as the interaction of *US Valence* and *CS-US Relation* did not reach significance, all *p*s $\geq .167$. Hence, stimulus classification was easier when the "melody" families ("scream" families) shared key with positive (negative) attributes than vice versa. Furthermore, the lack of an interaction of *US Valence* and *CS-US Relation* indicates that this advantage in classification speed was present irrespective of the level of *CS-US Relation*.

The ANOVA for the VoA group revealed a different pattern. The predicted interaction of *US Valence* and *CS-US Relation*, $F(1, 39) = 9.03$, $MSE = 0.83$, $p = .005$, $\hat{\eta}^2_G = .116$, reflected the fact that the direction of the simple main effect of *US Valence* depended on the level of *CS-US Relation*: D_f scores for the "starting melody" family (M = 0.21, SD = 0.57) were more positive than those for the "starting scream" family (M = -0.26, SD = 0.57), $F(1, 39) = 12.59$, $MSE = 0.36$, $p = .001$, $\hat{\eta}^2_G = .151$, whereas D_f scores for the "ending scream" family (M = 0.26, SD = 0.64) were more positive than those for the "ending melody" family (M = -0.12, SD = 0.63), $F(1, 39) = 5.01$, $MSE = 0.60$, $p = .031$, $\hat{\eta}^2_G = .088$. The other effects did not reach significance, all *p*s $\geq .146$.

The ANOVA for the VoR group revealed yet another pattern. The predicted main effect of *CS-US Relation*, $F(1, 39) = 5.05$, $MSE = 0.76$, $p = .030$, $\hat{\eta}^2_G = .072$, was found. As expected, all other main effects and interactions did not reach significance, all *p*s $\geq .540$. The main effect of *CS-US Relation* reflected more positive D_f scores for the "starting" families (M = 0.09, SD = 0.58) than for the "ending" families (M = -0.22, SD = 0.55).

Discussion

As in the first experiment, the impact of CS-US relations on IAT performance depended on the type of similarity that was emphasized by a given IAT. Participants relied on US valence to align targets and valent attributes if this feature discriminated between families of the two target groups (VoS group), they relied on the combination of US valence and CS-US relations if this feature discriminated between the target families (VoA group), and they relied on CS-US relation alone, if this was the feature that allowed a discrimination between the two target groups (VoR group). Group differences in the pattern of family-based IAT effects were clearly reliable in this experiment, as were the patterns of findings within each group. Furthermore, the pattern of results cannot be explained by differences in stimulus composition as the three IATs were identical in this regard. In sum, the findings of the second experiment replicate and extend the findings of the previous study, indicating that information on US valence and CS-US relations is accessible during the measurement phase and can be used flexibly to construct target-attribute similarities in the IAT. Which information is used depends on which information is most helpful in simplifying the task by aligning target and valent attribute categories during the compatible block.

General Discussion

Across two experiments, we consistently found that the relative impact of information on US valence and CS-US relations on IAT performance depended on the type of IAT that was being used. Some IAT effects were driven solely by US valence, others solely by CS-US relations, and yet other IATs by the integration of US valence and CS-US relations. Moreover, the prominence of the different types of information in a given IAT was tightly connected to the way this IAT was constructed. First, IAT variants in which target families with positive US valence (on one response key) opposed target families with negative US valence (on the other key) reflected only the valence of the co-occurring sounds. Secondly, IAT variants in which target families performing an action with a positive meaning opposed target families performing an action with a negative meaning reflected the full integration of US valence and CS-US relations, that is, the valence of the actions. A third type of IAT, in which the "starting" families were contrasted with the "ending" families, reflected only the valence of the CS-US relations (Eder et al., 2013). We will first discuss these results with regard to theoretical accounts of EC, and later relate them to prominent accounts of IAT performance. Subsequently, we will address possible mediators of similarity construction, and propose a few, possibly attractive avenues for future research.

Implications for theoretical models of EC

The present findings reveal that the IAT is not inherently impervious to information on CS-US relations, and can very well draw on propositional knowledge that was acquired during the learning procedure. Therefore, our results seriously challenge the original interpretation of the "ending" effect as reflecting association formation. Because information on US valence and CS-US relations are equally accessible in the IAT, and the latter by definition reflects a propositional learning process, the "ending" effect — by itself — cannot count as unambiguous evidence for a second, association formation process. After all, the information behind the "ending" effect (i.e., US valence) could have obviously been acquired in the same, insightful and controlled manner as the information on CS-US relations that affects behavior in other IAT types. Hence, propositional accounts of EC pose a plausible and parsimonious explanation for the whole range of IAT effect patterns reported in this study: participants store propositional knowledge about the four families in memory, but the (strategic or automatic) application of this knowledge for evaluating the families depends on the particular IAT with which these evaluations are assessed; specifically, the relative susceptibility of similarity construction processes within a certain IAT to certain dimensions of this knowledge.[19]

However, our findings can also be accounted for by (certain) dual process models of EC. Whereas the notion that direct measures reflect the product of propositional learning, and indirect measures reflect the product of association formation (Rydell et al., 2006) is certainly untenable, one could still argue that — while being acquired through different processes — associations and propositions are nevertheless stored in the same memory unit (e.g., in the form of parallel, separate evaluative associations per family). Granting this notion, the variable impact of information on US valence and CS-US relations across IATs can be conceived of as the result of a conditionally automatic activation of valence that depends on which other objects are made accessible in the context of an IAT. Theoretically, such a context-dependent activation of evaluative associations can be explained in terms of pattern activation, that is, selective activation of evaluative associations driven by the set of external input stimuli present in a given IAT or context (Gawronski & Bodenhausen, 2006). From this perspective, the differential inclusion of families (Experiment 1) as well as their variable groupings (Experiment 2) count as different sets of input stimuli, and therefore activate different parts of the semantic network. In summary, both propositional as well as dual process models of EC can account for the different patterns of IAT effects reported in this study. On these grounds, the "ending" effect does not provide unambiguous evidence for association formation, and is more parsimoniously explained by propositional accounts.

Implications for IAT accounts

The similarity account posits that IAT effects are driven by flexible constructions of target-attribute similarity. This notion is further supported by our results in that some IATs reflected information on US valence (suggesting similarity construction based on the valence of the sounds), others reflected information on CS-US relations (suggesting similarity construction based on the valence of the relational qualifier), and yet others reflected the integration of both types of information (suggesting similarity construction based on the meaning of the action). Furthermore, the similarity account stipulates that the relative prominence of different types of information is determined by their salience in a given IAT. While salience

[19] Moran and Bar-Anan (2013) found the "ending" effect in the IAT as well as in another indirect measure of evaluation, the sorting paired feature task (SPF). As to whether the "ending" effect in the SPF can also be explained in terms of similarity construction, we can only speculate. Nevertheless, a number of procedural similarities between the IAT and the SPF make this possibility somewhat plausible. First of all, the "ending" effect in the SPF is also based on a version of the SPF that compares the two "ending" families. As in the "ending" IAT, the relational information is therefore constant which renders it less salient than the US Valence information that varies across the two families. Second, as is the case in the IAT, the SPF requires an active classification of the target stimuli making it more prone to attempts of task simplification than for example the evaluative priming procedure. Last but not least, the SPF requires the classification of pairs of target and attribute stimuli into four categories (the Cartesian product of the two attribute categories and the two target categories.) using four keys. It seems plausible to assume that participants would apply similarity construction based on US Valence, e.g., to memorize the key that represents the "ending melody" family and positive attributes as the "positive" key.

of information is undoubtedly crucial for the construction of target-attribute similarity and hence IAT performance, other factors, such as the amount of considered information, or the ease with which it lends itself to the alignment of targets and attributes, appear to be relevant as well. For example, both the valence of the sound (action) as well the valence of the action (relational qualifier) can be used to construe target-attribute similarities in the "ending" IAT ("melody" and "scream" IATs). Yet, when considering amount of information and ease of alignment, the respective predominance of one target-attribute similarity over the other makes perfect sense: Aligning target and attributes in terms of US valence requires only one type of information, and is therefore favored in the "ending" IAT. By contrast, similarity construction based on the positive versus negative meaning of the actions in the "melody" and "scream" IATs requires processing of two types of information, but the integration of this information is more unambiguously positive or negative than the relational qualifiers "starting" vs. "ending", and therefore retains the upper hand. Hence, our results confirm the predictions of the similarity account, but also draw attention to additional factors that might influence which features are used for the construction of target-attribute similarities in a given IAT.

We do not want to deny that the standard association account can in principle be reconciled with our findings if it is combined with the notion of pattern activation, which allows for a context-dependent activation of (different) associations. In order to compete with the similarity construction account, however, the concept of context-dependent pattern activation must be further elaborated so that specific predictions for different types of IATs can be derived from this account. As long as differential pattern activation is only a post hoc explanation for all kinds of complex results in different variants of the IAT, it remains impossible to test or refute the association account, rendering it empirically vacuous. In our view, it will be extremely difficult to account for the exact pattern of IAT results that obtained in our study in terms of a differential activation of associations. This is because the concept of pattern activation draws on a correspondence or overlap between specific features of the learning and measurement situations. This perspective, however, provides no clue to explain effects of different target groupings in the IAT that have no equivalent in the learning phase which consisted of episodes in which single target exemplars (CSs) were presented instead of pairs or groups of different targets.

Regardless of whether and how this debate will be decided, the crucial point may not be whether one interprets IAT effects in terms of associations or similarity constructions, but rather whether one acknowledges their tremendous flexibility. In our view, the striking ease with which different types of information flow in and out of IAT performance somewhat precludes the possibility of using a standard IAT to draw strong inferences about the relative strength of target-attribute associations in general, and about association formation in EC in particular. While certain patterns of IAT effects — such as a main effect of US valence that is unaffected by CS-US relations and consistent across different IAT types — may make a strong case for the latter, our results clearly show that such a pattern is certainly not the default, and should never be inferred based on only a single IAT.

Our findings also speak to a broader question that is related to the IAT as well as other indirect measures of evaluation, specifically, whether "unqualified associations" have a stronger impact on automatic evaluation (which indirect measures supposedly reflect) compared to deliberative evaluation (e.g., Moran & Bar-Anan, 2013). In our view, this question cannot be answered unambiguously, because the influence of "unqualified associations" (i.e., associations that solely reflect co-occurrence and do not incorporate information on CS-US relations) on automatic evaluation depends strongly on which kind of measure is used to assess this type of evaluation. Our finding that the impact of relational information on IAT performance depends on the structure of the IAT is crucial in this regard. Apparently, the IATs insensitivity to relational information is restricted to certain versions of it, namely those that emphasize target-attribute similarity based on US valence. When other types of similarity are highlighted or more helpful for similarity construction processes, the IATs sensitivity to relational information is comparable to that of direct measures. To the best of our knowledge, the sensitivity of other indirect measures of evaluation like the Evaluative Priming (EP; Fazio et al., 1986) and Affective Misattribution Procedures (AMP; Payne, Cheng, Govorun, & Stewart, 2005) to relational information seems to be comparable to that of direct measures of evaluation (see, e.g., Exp. 3 in Hu et al., 2017; Moran, Bar-Anan, & Nosek, 2017). Based on our findings, one may raise the question whether other indirect measures would also forfeit their sensitivity to relational information if they were to compare only those CSs that share a relational qualifier but differ in terms of their US valence (see, e.g., Exp. 1 in Hu et al. (2017), which also failed to find a significant interaction between US Valence and CS-US Relation in an EP paradigm when the CS-US-relation was manipulated on a between-subjects basis). Thus, in our view, whenever an indirect measure of evaluation is insensitive to the influence of relational qualifiers, this seems to reflect structural properties of the specific measure. Whenever co-occurrence information is highlighted during the measurement procedure or facilitates the task, this "unqualified" information will control responding. This kind of influence, however, may not indicate a general influence of unqualified associations on automatic evaluation.

Possible mediators of similarity construction

While our results clearly demonstrate that the IAT is highly flexible in its in- and exclusion of information, they can only hint at how this flexibility was brought about. In this regard, Experiment 2 indicates that the present IAT effects do not rely on the mere absence or presence of families; instead, the grouping of targets seems to be the crucial element in a given IAT. This fact suggests an interpretation in terms of a flexible construction of target-attribute similarity which draws on some (valent) feature of the target categories that (a) discriminates between the opposed target categories, and (b) provides a basis to differentially align them with the (valent) attribute categories. Such a notion is very close to the concept of *recoding* (Rothermund, Teige-Mocigemba, Gast, & Wentura, 2009), which we therefore deem to be the most likely mediator of similarity construction. Recoding is an IAT mechanism by which the instructed dual-classification is (strategically or implicitly) replaced or supplemented with a unidimensional categorization that allows for a simplification of the task in the compatible, but not in the incompatible block. Crucially, pairs of target and attribute categories in the compatible block are recoded based a feature that is shared by the two categories assigned to the same key, yet also distinguishes them from the categories assigned to the other key. Even though recoding provides a straightforward specification of similarity construction, the latter can in principle also be based on selective pattern activation of evaluative associations. Although it is barely possible to explain the entire, complex pattern of similarity-based effects of our study in terms of associations, they could nevertheless be responsible for some of the observed effects or could contribute to them. In order to quantify the relative contributions of recoding and associations in the generation of IAT effects, future research could use the ReAL model of IAT performance (Meissner & Rothermund, 2013), an MPT model that was designed to separate the effects of associations and recoding on the IAT. Specifically, the different IAT effects we presented in this paper could be compared with regard to their model parameters yielded by the ReAL model. While recoding and even pattern activation in semantic memory (Gawronski & Bodenhausen, 2006) can be conceived of as a strategic and therefore controlled process, their strategic nature should not be assumed offhandedly. Instead, future research should experimentally manipulate different features of automaticity, such as the availability of cognitive resources, time pressure, and processing goals, in order to gain insight into the (non-) automaticity of similarity construction in the IAT.

Conclusions

The IAT is an undoubtedly useful tool for measuring evaluations in an indirect and unobtrusive manner, and has quite a few advantages over other indirect measures, such as markedly higher reliabilities and effect sizes as well as its capability to channel the category-based processing of the presented stimuli. We therefore expect that this popular measurement procedure will continue to provide interesting and meaningful insights into the mechanisms and boundary conditions of evaluative learning. However, our findings strikingly illustrated the particular challenges that are connected with the IAT in gathering unambiguous evidence for association formation in EC. Due to their great flexibility concerning the in- and exclusion of information on US valence and CS-US relations, IAT effects in and of themselves do not allow for strong inferences about association formation as an additional mediator of EC, and are more parsimoniously explained by propositional accounts. Therefore, future theory-driven research combining the IAT with *US Valence* × *CS-US Relation* designs needs to incorporate creative, yet rigorous strategies of rendering the given IAT effects meaningful with respect to model predictions. We believe that the systematic comparison of effects on different IATs or the complementary use of MPT models in analyzing IAT data represent promising approaches in the pursuit of this aim.

Chapter 4

Does Mere Co-occurrence Affect Evaluation? MPT Modelling of Relational Evaluative Conditioning Cannot (Yet) Tell

In relational EC studies, the reversed EC effect induced by contrastive CS-US relations is typically smaller than the regular EC effect induced by assimilative CS-US relations. Dual-process accounts of EC explain this attenuated reversed EC effect with the opposing evaluative implications of CS-US propositions and CS-US associations. This interpretation is corroborated by two recent studies analyzing evaluative CS classifications by means of a multinomial processing tree model assuming task performance to be driven by propositional valence, US valence and guessing. While the larger-than-zero p and a parameters (indicating response generation based on propositional and US valence respectively) fit well with a dual-process view of EC, they are also compatible with a single-process propositional perspective. We argue that contrastive CS-US relations induce less extreme CS evaluations, and therefore have a lower probability of yielding an evaluative inference during the evaluative classification task. In three simulation studies, we demonstrate that this simple set of assumptions cannot only explain the US valence parameter itself, but also accounts for its correlations with evaluative CS ratings as well as for the somewhat contradictory results of previous parameter validation studies.

In an evaluative conditioning (EC) procedure, neutral "conditioned" stimuli (CSs) are repeatedly paired with valent "unconditioned" stimuli (USs). Subsequent CS evaluations usually indicate a spread of valence from the US to the CS; that is, CSs paired with positive USs are evaluated more favorably than CSs paired with negative USs.

Early accounts of this so-called EC effect focused on *association formation* as its sole mediator (e.g., Baeyens et al., 1992). Association formation is a mental mechanism that is assumed to be driven by the spatio-temporal co-occurrence of at least two environmental stimuli; specifically, the mere simultaneous activation of their mental representations is postulated to produce a mental association between them. These unqualified memory links are assumed to emerge in an automatic fashion (i.e., largely independently of people's cognitive resources, goals and awareness state) with their strength being solely determined by the frequency of co-activation between the involved stimuli. According to associative accounts of EC effects, changes in CS evaluation should therefore reflect the valence of the co-occurring USs, be amplified by repeated CS-US pairings, and remain largely unaffected by manipulations of cognitive processing. While the first two predictions are consistently matched by empirical demonstrations of EC (e.g., Baeyens et al., 1992), the third prediction has been repeatedly contradicted by a multitude of studies showing EC effects to be moderated by manipulations of cognitive resources (e.g., Dedonder et al., 2010), processing goals (e.g., Corneille, Yzerbyt, Pleyers, & Mussweiler, 2009) and awareness for the CS-US contingencies (e.g., Pleyers et al., 2007).

Mirroring these determinants of EC effects, an alternative account of the latter centers around *proposition formation*, a non-automatic mental mechanism generating propositional statements about the relationship between at least two environmental stimuli (e.g., De Houwer, 2007). Accordingly, propositional accounts of EC effects assume that the latter reflect evaluative inferences that are drawn from propositional representations of the relationship between the CSs and their co-occurring USs (e.g., De Houwer, 2018). In line with this notion, the use of a modified EC procedure has shown EC effects to be strongly moderated by manipulations of the perceived relation between the co-occurring stimuli (e.g., Förderer & Unkelbach, 2012). In such relational EC procedures, some CS-US pairings are combined with an *assimilative* CS-US relation implying a convergence between the valence of the CS and the valence of its co-occurring USs (e.g., when the CS causes the USs), while other CS-US pairings are combined with a *contrastive* relation implying a contrast between said valences (e.g., when the CS prevents the USs). Reflecting the evaluative implications of their respective CS-US relation, CS ratings in the assimilative condition usually indicate a preference for positively over negatively paired CSs (i.e., a regular EC effect), while CS ratings in the contrastive condition typically indicate a reversed preference for negatively over positively paired CSs (hereafter, a reversed EC effect).

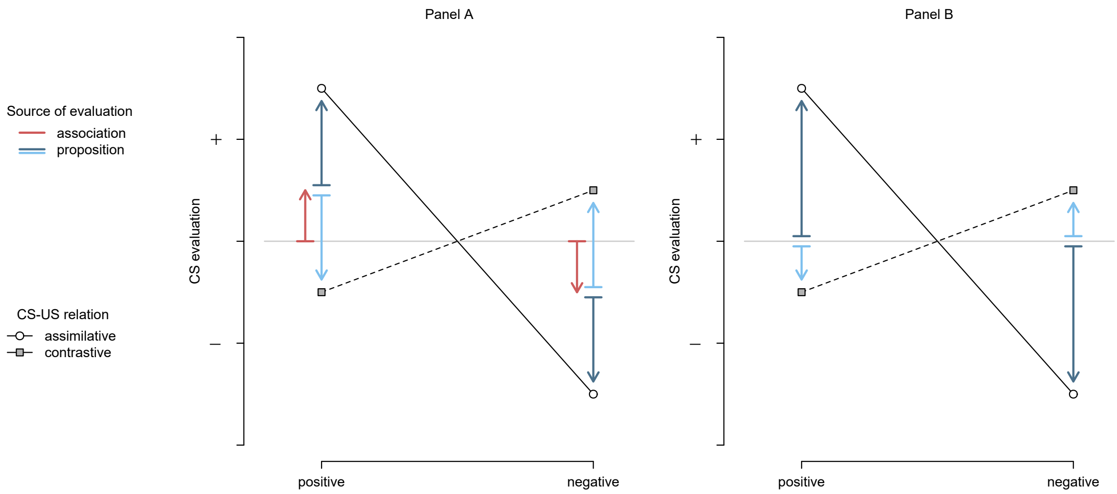

Figure 4.1. Schematic depiction of a dual-process (Panel A) vs. single-process (Panel B) propositional account of the asymmetrical EC effects in the assimilative vs. contrastive conditions of relational EC studies. Red arrows represent CS-US associations. Dark (light) blue arrows represent CS-US propositions involving an assimilative (contrastive) CS-US relation. The arrow head (length) indicates the direction (strength) of the evaluative implications of a CS-US association/proposition.

While the previously described interaction of *US valence* and *CS-US relation* clearly demonstrates the involvement of CS-US propositions in the emergence of (some) EC effects, it cannot preclude an additional influence of evaluative CS-US associations. Such a dual-process view on EC posits that associations and propositions jointly mediate the effect of the CS-US pairings on CS evaluations (Gawronski & Bodenhausen, 2018). In line with this notion, the reversed EC effect in the contrastive condition is usually smaller than the regular EC effect in the assimilative condition, resulting in a statistically significant main effect of *US valence* (see Figure 4.1, panel A). Dual-process accounts can explain this pattern as the joint product of two independent EC effects: one driven by CS-US associations and one driven by CS-US propositions (Moran, Bar-Anan, & Nosek, 2016). Specifically, observing the CS-US pairings produces equally positive (negative) CS-US associations for positively (negatively) paired CSs, irrespective of whether the CS-US pairing had an assimilative vs. contrastive quality (see red arrows in Figure 4.1, panel A). For CSs in the assimilative conditions, the evaluative implications of these associations point in the same direction as the CS evaluations derived from CS-US propositions (see dark blue arrows in Figure 4.1, panel A). However, for the CSs in the two contrastive conditions, the evaluative implications of their respective associations vs. propositions point in opposite directions (see light blue arrows in Figure 4.1, panel A). Assuming that both associations and propositions become activated when participants are asked to indicate their CS evaluations, the attenuated reversed EC in the contrastive condition can be explained by the fact that the evaluative impact of the activated propositions is summed up with and therefore mitigated by the opposing evaluative impact of the activated associations.

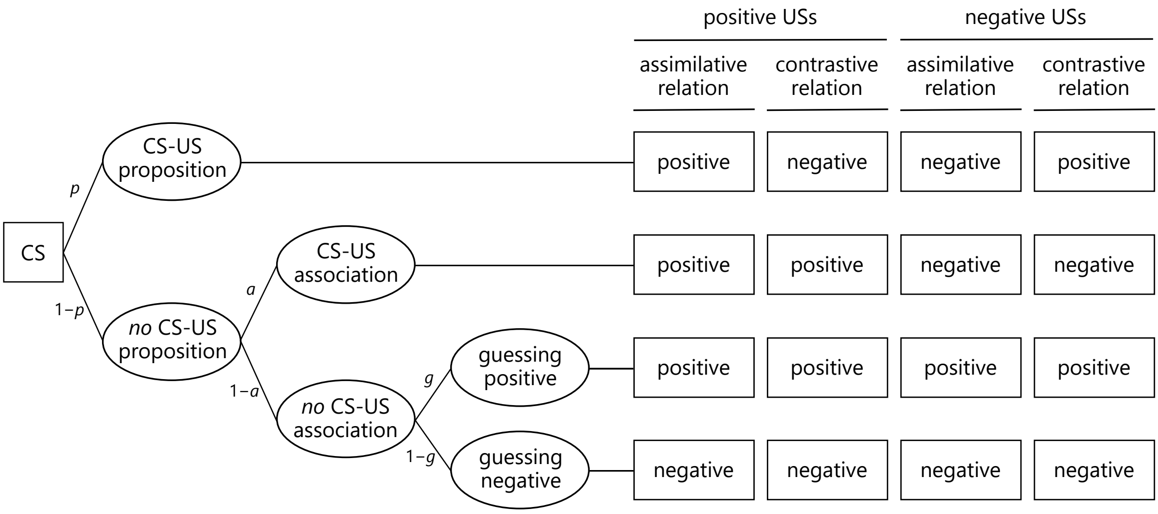

Figure 4.2. Schematic depiction of the MPT models used by Heycke and Gawronski (2020) and Kukken et al. (2020). Note that for Heycke and Gawronski (2020), a CS classification as "positive" ("negative") translate to a "yes" ("no") response.

A similar logic forms the basis of a recently introduced approach to studying the asymmetrical EC effects induced by assimilative vs. contrastive CS-US relations. In this new approach, evaluative CS classifications (as either "positive" or "negative") are analyzed by means of a multinomial processing tree (MPT) model mirroring the previously introduced dual-process account of EC effects: that is, an evaluative CS classification is assumed to be driven either by an evaluative inference derived from the available CS-US propositions, by the valence of the co-occurring USs, or by guessing.

A first application of such an MPT model was presented by Heycke and Gawronski (2020) whose participants observed CS-US-relation triplets consisting of a fictitious pharmaceutical product (as CS), positive or negative health states (as US), and the relational information of whether a given CS caused or prevented a given US. Subsequently, participants were subjected to a speeded choice task in which they had to indicate whether they would choose a given pharmaceutical product or not (with a "yes" ["no"] response indicating a positive [negative] evaluation of the respective CS). The resulting response frequencies mirrored the previously described asymmetrical EC effects on CS ratings: while the relative frequency of a propositionally correct classification surpassed chance level in all four conditions alike, CSs preventing a health state were substantially less often classified according to their propositional valence than CSs causing a health state. An equivalent classification asymmetry was obtained by Kukken et al. (2020) who presented their participants with pairings of unknown cartoon characters (CSs) and positive or negative sounds (USs) combined with information on whether a given CS started or stopped a given sound. In a subsequently administered evaluative classification task, CSs stopping a sound were classified according to their propositional valence in a substantial number of cases, but less often so than CSs starting a sound.

Both author teams analyzed these response frequencies with the MPT model depicted in Figure 4.2. This model specifies that the evaluative classification of a given CS is driven by the propositional valence of its *US valence* $\times$ *CS-US relation* condition with an unconditional probability of p. Given that the propositional valence does not drive response generation (with a probability of $1 - p$), the classification response is assumed to be driven by the US valence of a given CS with a conditional probability of a. Finally, given that neither propositional nor US valence drive response generation (with a probability of $[1-p]*[1-a]$), the classification response is assumed to be driven by guessing "positive" ("negative") with a conditional probability of g $(1 - g)$. In both empirical applications alike, this MPT model produced larger-than-zero probabilities for response generation based on propositional valence (hereafter, the p parameter). More crucially, both Heycke et al. and Kukken et al. also reported response generation based on US valence in a substantial number of cases (hereafter, the a parameter).

According to the two author teams, the larger-than-zero a parameter reflects a direct influence of US valence on CS evaluation that is independent of CS-US relations: specifically, the size of the a parameter is taken to indicate that in a substantial number of trials participants generate a CS classification by drawing on the valence of the USs that co-occurred with the to-be-classified CS. As discussed by Heycke et al., such a direct effect of US valence on response generation in the evaluative classification task can be reconciled with dual-process as well as purely propositional accounts of EC effects. From a dual-process perspective, said effect is accounted for by evaluative CS-US associations driving the classification responses whenever CS-US propositions are unavailable. From the propositional perspective, US valence may drive classification responses via *partial retrieval* of the CS-US propositions: that is, participants may not always be able to retrieve both types of information connected to a given CS (especially so when put under time pressure), and may rely on the retrieved US valence in order to generate a classification response whenever the CS-US relation is unavailable.

While any such direct effect of US valence on CS evaluation[20] poses a viable explanation of the a parameter, equating said MPT parameter with such an effect from the start is problematic because it conflates an empirical finding with a theoretical mechanism. Empirically, the a parameter represents nothing more than the previously mentioned classification asymmetry: i.e., CSs in the contrastive conditions are relatively less often classified according to their propositional valence than CSs in the assimilative conditions. As the MPT model employed by the two author teams assumes identical p parameters for the two *CS-US relation* conditions, this classification asymmetry is captured by the a parameter representing the theoretical idea of CS evaluation being driven by US valence irrespective of CS-US relations. However, said classification asymmetry can be equally well described by an alternative MPT model that does not contain an a parameter, but instead allows for a smaller p parameter in the contrastive than in the assimilative conditions: specifically, the classification asymmetry previously captured by the a parameter would now reveal itself in the difference between the two p parameters. Given that the evaluative classifications featuring the previously introduced asymmetry can also be described by an MPT model including two p parameters (instead of an a parameter), said asymmetry (and, by extension, the a parameter itself) can also be explained by any theoretical mechanism evoking a lower probability for a propositionally correct classification in the contrastive than in the assimilative conditions.

One such mechanism was already raised by Moran et al. (2016) in order to explain the previously mentioned asymmetrical EC effects on evaluative CS ratings: that is, the possibility of contrastive relations inducing weaker CS evaluations than assimilative relations (see Figure 4.1, panel B). While it has already been shown that certain assimilative relations induce more extreme CS evaluations than other assimilative relations (Hughes, Ye, Van Dessel, & De Houwer, 2019), a general asymmetry in valence induction between assimilative vs. contrastive relations has not been demonstrated yet. However, such an asymmetry may be plausible for a number of reasons. Firstly, the direction of the standard EC effect (obtained in a classical EC procedure without a manipulation of *CS-US relation*) seems to imply assimilation as the default mode of the mental process by which CS-US propositions

[20]Note that in the remainder of this article, we will use this term to refer to the possibility of evaluative behavior being driven by CS-US associations and/or partial retrieval of CS-US propositions.

are translated into evaluative inferences. The consequently greater familiarity with inferring valence from assimilative relations may produce a processing advantage in the assimilative conditions of relational EC procedures, and could therefore lead to the more extreme CS evaluations which they typically portray. Furthermore, idiosyncratic features of the materials used in relation EC procedures may be critical as well. For example, Kukken et al. used a pleasant melody as the positive US. While we typically enjoy listening to music — and may therefore generate a positive evaluation for the CSs starting the melody, we usually do not feel particularly harmed by not listening to it — leading to rather neutral (instead of negative) evaluations for CSs stopping the melody. In a similar vein, Heycke et al. presented their participants with images of drastic, yet uncommon health impairments as negative USs. The fact that most participants should dread such an ailment, but have no reason to expect it for themselves, may therefore explain why CSs causing these states of illness receive much stronger evaluations than CSs preventing them.

Irrespective of their exact origins, the possibly weaker evaluations induced by contrastive relations may translate into a relatively lower probability for classifying CSs presented with such relations according to the propositional valence of their respective *US valence* × *CS-US relation* condition (and may thereby produce the previously described classification asymmetry underlying the a parameter). The rationale behind this claim is that CS-US propositions inducing relatively more neutral (continuous) evaluations are more likely to result in the evaluative inference of a given CS being neutral — instead of positive or negative as required by the evaluative classification task. Furthermore, whenever the evaluative information provided by the CS-US proposition(s) connected to a certain CS does not produce an evaluative inference as positive vs. negative, participants will have to resort to other sources of CS evaluation (or outright guessing) in order to arrive at a suitable classification response. Crucially, response generation based on other stimulus characteristics (e.g., the color of the CS) as well as pure guessing should produce a higher share of propositionally incorrect CS classifications than would response generation based on evaluative inferences drawn from the *US valence* × *CS-US relation* condition of a given CS. Given that contrastive CS-US relations do indeed induce less extreme evaluations (and more often result in neutral inferences), the consequently higher share of classification trials driven by other sources of evaluation (or guessing) in the contrastive conditions should produce more propositionally incorrect responses in said condition and thereby give rise to the previously introduced classification asymmetry driving the a parameter.

Aims of the present research

The overarching aim of the present research is to demonstrate the viability of the previously introduced *inference asymmetry* account in explaining the a parameter reported by Heycke et al. and Kukken et al. To this end, we present three simulation studies each of which speaks to a distinct empirical finding related to the a parameters mental underpinnings. In the first simulation study, we address the question of whether the emergence of a larger-than-zero *a* parameter implies an independent effect of US valence on response generation in the evaluative classification task, or whether it can also be explained in terms of an inference asymmetry. In the second simulation study, we demonstrate that the correlations between the MPT parameters and evaluative CS ratings reported by Kukken et al. are equally compatible with an *a* parameter driven by an effect of US valence and an *a* parameter driven by an inference asymmetry between the assimilative vs. contrastive conditions. Finally, in the third simulation study we show that the puzzling increase of the *a* parameter under conditions of improved memory retrieval (which is inconsistent with dual-process as well as partial retrieval accounts) is perfectly compatible with the inference asymmetry account of the *a* parameter.

Simulation study 1

In Simulation study 1, we sought to show that a lower probability of generating an evaluative inference as positive or negative (hereafter abbreviated as "evaluative inference") in the contrastive than in the assimilative condition suffices to explain the emergence of a larger-than-zero *a* parameter which does therefore not imply an independent effect of US valence on CS evaluation. To this aim, we simulated individual evaluative inference (EI) probabilities for the assimilative vs. contrastive conditions in two samples. In the "symmetrical" sample, the mean EI probabilities were comparable across the two conditions. However, in the "asymmetrical" sample, the EI probabilities were systematically larger in the assimilative than in the contrastive conditions. We then used these EI probabilities to simulate CS classifications by means of a two-step algorithm that does not include any direct effect of US valence: whenever an EI was available for a given CS, the classification response was set to reflect the propositional valence of its US valence × CS-US relation condition; however, whenever no such EI was available, the classification response was generated by an unbiased guessing process. The resulting response frequencies were then analyzed with the previously introduced MPT model including the *p*, *a* and *g* parameters. For both samples alike, we expected the mean *p* parameter to be larger than zero reflecting the substantial EI probabilities in the assimilative as well as contrastive conditions. Based on the unbiasedness of the simulated guessing process, we furthermore predicted a mean *g* parameter close to .5 in both samples. Crucially, we expected the comparable EI probabilities in the assimilative vs. contrastive conditions of the symmetrical sample to produce a mean *a* parameter of (practically) zero. By contrast, the systematic EI asymmetry in the asymmetrical sample should bring about a mean *a* parameter that is substantially larger than zero.

Method

Sample sizes and design. For each level of "Sample", 500 participants were simulated. The data simulation followed a 2 (*Sample*: "symmetrical" vs. "asymmetrical") × 2 (*US Valence*: "positive" vs. "negative") × 2 (*CS-US relation*: "assimilative" vs. "contrastive") mixed design with the first factor varying between "participants".

Simulation of individual inference probabilities. In a first step, we generated individual EI probabilities for each participant: one for the assimilative condition and another for the contrastive condition. These probabilities were created by drawing a random value x from a normal distribution which was then probit transformed into its corresponding cumulative probability (with $x = 0$ corresponding to $p = .5$). For the "symmetrical" condition, we sampled from identical, uncorrelated normal distributions ($\mu_{\text{ass}} = \mu_{\text{con}} = 0$; $\sigma_{\text{ass}} = \sigma_{\text{con}} = 0.25$; $\rho_{\text{ass,con}} = 0$) for the assimilative and contrastive conditions. Accordingly, the sample mean of the individual EI probabilities was almost the same for the two conditions ($P_{\text{ass}} = .496$; $P_{\text{con}} = .503$) resulting in a mean EI asymmetry of $\Delta P = P_{\text{ass}} - P_{\text{con}} = -.007$ (see Figure 4.3, left panel). The individual EI probabilities in the assimilative and contrastive conditions were un-correlated with one another ($r = .01$, 95% CI $[-.08, .10]$, $t(498) = 0.20$, $p = .841$); but correlated significantly with the individual EI asymmetries (assimilative condition: $r = .70$, 95% CI $[.66, .75]$, $t(498) = 22.10$, $p < .001$; contrastive condition: $r = -.70$, 95% CI $[-.75, -.66]$, $t(498) = -22.13$, $p < .001$).

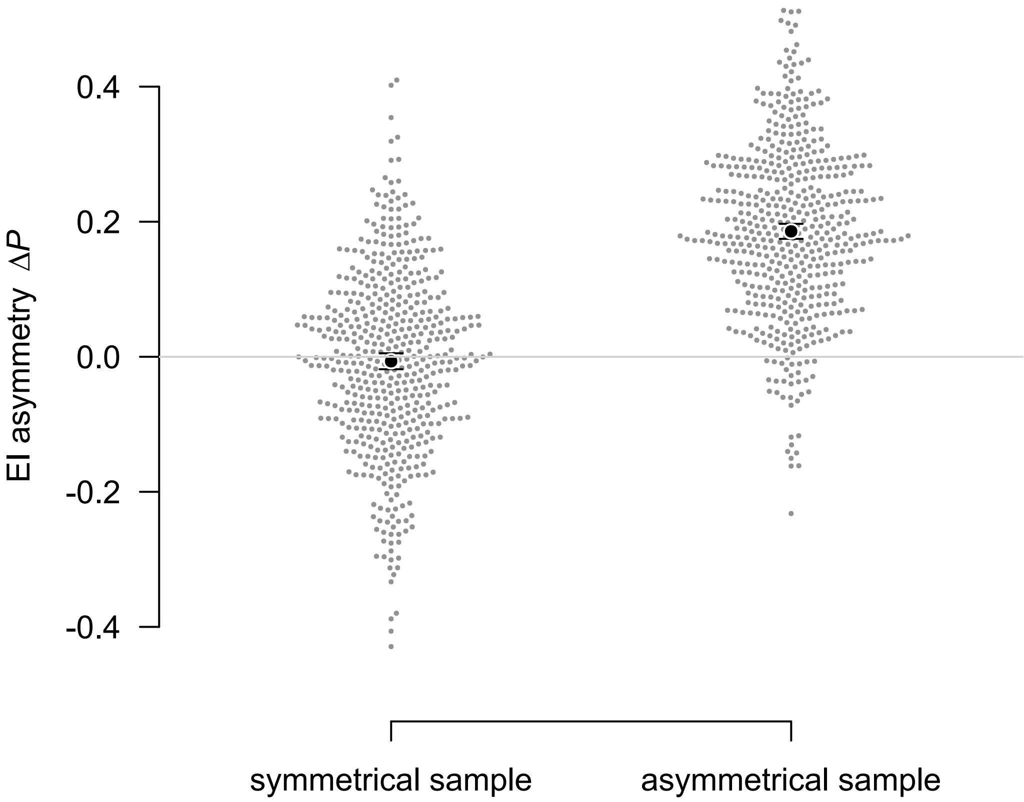

Figure 4.3. Simulation study 1: Individual EI asymmetries in the symmetrical and the asymmetrical condition.

For the "asymmetrical" condition, we again sampled from two uncorrelated normal distributions with identical standard deviations ($\sigma_{\text{ass}} = \sigma_{\text{con}} = 0.25$; $\rho_{\text{ass,con}} = 0$). This time, however, the mean of this distribution was higher in the assimilative than in the contrastive condition ($\mu_{\text{ass}} = 0$; $\mu_{\text{con}} = -0.5$). Accordingly, the sample mean of the individual EI probabilities was higher in the assimilative ($P_{\text{ass}} = .501$) than in contrastive condition ($P_{\text{con}} = .316$) resulting in a mean EI asymmetry of $\Delta P = P_{\text{ass}} - P_{\text{con}} = .186$ (see Figure 4.3, right panel). Again, individual EI probabilities in the assimilative and contrastive conditions were un-correlated with one another ($r = .06$, 95% CI $[-.03, .14]$, $t(498) = 1.29$, $p = .198$); but correlated significantly with the individual EI asymmetries (assimilative condition: $r = .75$, 95% CI $[.71, .79]$, $t(498) = 25.43$, $p < .001$; contrastive condition: $r = -.62$, 95% CI $[-.67, -.56]$, $t(498) = -17.41$, $p < .001$).

Table 3

Simulation study 1: Absolute (N) and relative (%) frequencies of propositionally correct CS classifications as a function of Sample, US valence and CS-US relation.

		Symmetrical sample		Asymmetrical sample	
US valence	CS-US relation	N	%	N	%
positive	assimilative	7,536	75.36	7,489	74.89
	contrastive	7,501	75.01	6,599	65.99
negative	assimilative	7,510	75.10	7,515	75.15
	contrastive	7,440	74.40	6,616	66.16

Simulation of individual classification responses. In the next step, the classification responses in the evaluative classification task were generated. For each participant, we simulated a total of 80 CS classifications (20 per US valence × CS-US relation condition) using the following algorithm. For each CS, we sampled an EI status from a Bernoulli distribution with a *1* (*0*) indicating that an EI was available (unavailable) for a given CS. The hit probability of the Bernoulli distribution depended on the level of *CS-US relation* of a given CS, but was independent of its level of *US valence*; specifically, for each CS, the hit probability was set equal to the individual EI probability of its respective *CS-US relation* condition. Subsequently, the classification responses were generated: all CSs with an available EI received a classification response according to their propositional valence; i.e., a *positive* classification for CSs presented with positive (negative) USs and an assimilative (contrastive) relation, and a *negative* classification for CSs presented with negative (positive) USs and an assimilative (contrastive) relation. For each CS without an available EI (i.e., a neutral inference), we simulated an unbiased guessing process by sampling a positive (*1*) vs. negative (*0*) response from a Bernoulli distribution with a hit probability of .5. As intended, the (overall) share of propositionally correct CS classifications was comparable for the assimilative vs. contrastive conditions in the "symmetrical" sample. By contrast, the "asymmetrical" sample featured the previously described classification asymmetry; that is, higher shares of propositionally correct CS classifications in the assimilative than in the

Table 4
Simulation study 1: Posterior summaries of MPT parameter estimates.

	Symmetrical sample		Asymmetrical sample	
Parameter	M	95% CI	M	95% CI
Means				
p	.500	[.489, .510]	.409	[.398, .421]
a	.006	[.000, .017]	.135	[.112, .156]
g	.501	[.491, .510]	.499	[.490, .508]
Correlations				
$\rho_{p,a}$	.395	[-.127, .823]	.499	[.238, .745]
$\rho_{p,g}$	.495	[.029, .873]	-.180	[-.814, .633]
$\rho_{a,g}$	.360	[-.387, .907]	-.194	[-.823, .671]

contrastive conditions[21] (see *Table 1*).

Model analysis. The model analysis of the simulated CS classifications was conducted in R (R Core Team, 2020) using the *TreeBUGS* package (Heck, Arnold, & Arnold, 2018). The response frequencies were modeled with the previously introduced MPT model (see Figure 4.2) using the latent-trait approach which allows for heterogeneity in parameter values between participants (Klauer, 2010). For each sample, parameters were estimated by running four Markov chains with 400,000 iterations each (200,000 were discarded as burnin iterations). We used 20,000 adaptation iterations and a thinning rate of 40. Convergence was monitored by means of the Gelman-Rubin statistic (Gelman & Rubin, 1992) using a criterion of $\hat{R} < 1.02$ for all parameters. Model fit was assessed by means of posterior predictive p-values (*PPP*; Heck et al., 2018) which were based on the resampling of 5,000 posterior samples.

Results

Symmetrical sample. In the symmetrical sample, the response frequencies were well accounted for by the MPT model, $T_1^{obs} = 0.01$, $T_1^{pred} = 0.01$, $p = 0.21$. The p parameter was larger than zero (see *Table 2*) indicating response generation based on propositional valence in 50% of classification trials. This number fits well with the mean of the individual EI probabilities which was close to .5 for both the assimilative and contrastive conditions. The a parameter was practically zero indicating response generation based on US valence in a mere 0.6% of classification trials (for which response generation was not driven by the propositional valence). This negligible number reflects the fact that the share of propositionally correct classifications was comparable for the assimilative and contrastive

[21]Note that the mean EI asymmetry of 18.6% translates to a mean classification asymmetry of 8.95%. This halving is due to the fact that the CS classifications are based on response generation based on propositional valence as well as unbiased guessing with the latter process partly compensating for the EI asymmetry between the assimilative and contrastive conditions.

conditions in the symmetrical sample. Finally, the g parameter was very close to .5 mirroring the unbiased guessing process which we simulated in order to generate classification responses whenever an evaluative inference was unavailable.

Asymmetrical sample. The response frequencies in the asymmetrical sample were also well explained by the MPT model, $T_1^{obs} = 0.01$, $T_1^{pred} = 0.01$, $p = 0.58$. As in the symmetrical sample, the p parameter was larger than zero and indicated response generation based on propositional valence in 40.9% of classification trials[22]. Similarly, the g parameter was very close to .5 again mirroring the unbiased guessing process by which classification responses in the absence of an evaluative inference were generated. Crucially, the a parameter was also larger than zero apparently indicating response generation based on US valence in 13.5% of classification trials for which response generation was not driven by propositional valence[23].

Discussion

Simulation study 1 demonstrated that a larger-than-zero a parameter can be brought about by a lower probability of generating an evaluative inference in the contrastive than in the assimilative conditions, and does therefore not require any systematic influence of US valence. Accordingly, the algorithm underlying the classification responses in the symmetrical as well as the asymmetrical sample relied either on propositional valence or unbiased guessing, but was never informed by the US valence of a given CS. As illustrated by the results in the symmetrical sample, this data-generating mechanism produced equally substantial shares of propositionally correct CS classifications in the two *CS-US relation* conditions when the EI probabilities were comparable across the two conditions. By contrast, in the asymmetrical sample (with its systematically lower EI probabilities in the contrastive condition), the same data-generating mechanism produced substantially lower shares of propositionally correct classifications for CSs in the contrastive conditions. As indicated by the MPT parameter estimates in the two samples, the size of the classification asymmetry between the *CS-US relation* conditions determines the size of the a parameter. Therefore, the mere emergence of a larger-than-zero a parameter can be simulated by an asymmetry in EI probabilities, and thus does not imply an independent effect of US valence on response generation in the evaluative classification task.

While the classification asymmetry underlying the a parameter can of course be explained by such an effect (whether based on CS-US associations or partial retrieval of CS-US propositions), it is equally well accounted for by asymmetrical EI probabilities in the assimilative vs. contrastive conditions. A similar point was already raised by Kukken et al. who discussed an asymmetry in the formation and/or retrieval of memory for CS-US propositions as a potential mechanism underlying the a parameter. These authors suggested

[22]The size of the p parameter corresponds to the overall mean of the individual EI probabilities in the asymmetrical sample, $P_{\text{overall}} = 0.5 * (P_{\text{ass}} + P_{\text{con}}) = .408$.

[23]The size of the a parameter reflects the magnitude of the classification asymmetry in the asymmetrical sample (see *Footnote 2*), but is slightly larger due to the fact that the a parameter is modeled as a conditional probability.

that the correlations between the a parameter and evaluative CS ratings may distinguish between an a parameter driven by an influence of US valence vs. by a memory asymmetry between the *CS-US relation* conditions. They argued that if the a parameter reflected a mental process deriving evaluative CS classifications from US valence, the same process should affect evaluative CS ratings as well. Consequently, the correlations between the a parameter and evaluative CS ratings in each of the four *US valence* × *CS-US relation* conditions should follow an "associative" pattern; that is, for both levels of *CS-US relation*, larger a parameters should come with comparatively more positive ratings for positively paired CSs and with comparatively more negative ratings for negatively paired CSs. By contrast, according to Kukken et al., an a parameter based on a memory asymmetry should only come with comparatively more positive (negative) ratings for positively (negatively) paired CSs in the assimilative conditions, but be un-correlated with evaluative ratings for the CSs in the contrastive conditions. Since the empirical correlations between the a parameter and evaluative CS ratings followed the "associative" pattern implied by an a parameter based on an independent effect of *US valence*, the authors concluded that said MPT parameter cannot be driven by a memory asymmetry between the assimilative vs. contrastive conditions. As the respective mechanisms by which a memory vs. inference asymmetry bring about a larger-than-zero a parameter are highly similar[24], the reasoning presented by Kukken et al. should also apply to the correlations between CS ratings and an a parameter based on an inference asymmetry. However, for mathematical reasons presented in the following section, the "associative" pattern of correlations reported by Kukken et al. is in fact equally predicted by an a parameter based on lower EI (or memory) probabilities for CSs presented with a contrastive CS-US relation. In order to corroborate this EI asymmetry account of the purportedly "associative" pattern of correlations between the a parameter and evaluative CS ratings, we conducted a second simulation study in which we simulated CS classifications as well as evaluative CS ratings, and investigated their correlations with one another.

Simulation study 2

As illustrated by the previous simulation study, the size of the a parameter reflects the size of the classification asymmetry between the *CS-US relation* conditions. Viewed through the lens of an EI asymmetry, the classification asymmetry is brought about by a mental process that derives classification responses from propositional valence, albeit with a lower probability in the contrastive than in the assimilative condition. Accordingly, the classification asymmetry should be particularly large (small) for participants with a high (low) EI probability in the assimilative condition as well as a low (high) EI probability in the contrastive condition. Given that the generation of evaluative CS ratings is driven by the same mental process,

[24]In the asymmetrical sample of Simulation study 1 we simulated CS classifications based on perfect availability of CS-US propositions and asymmetrical inference probabilities in the two *CS-US relation* conditions. However, the individual EI probabilities can also be construed as individual memory probabilities. When assuming an inference probability of 1 whenever a CS-US proposition is available, such asymmetrical memory probabilities would produce the same results as in Simulation study 1.

an equivalent logic applies to them as well. Specifically, participants with a high (low) EI probability in the assimilative condition should produce evaluative ratings for CSs in the assimilative condition that more (less) strongly reflect their respective propositional valence (whose sign is identical to that of their US valence). Furthermore, participants with a low (high) EI probability in the contrastive condition should produce evaluative ratings for CSs in the contrastive condition that less (more) strongly reflect their respective propositional valence (and therefore relatively more [less] strongly mirror their respective US valence). Taken together, larger a parameters should therefore come with more "associative" CS ratings (more strongly reflecting the US valence of a given CS) irrespective of whether the CS belongs to the assimilative or contrastive condition.

In Simulation study 2, we sought to corroborate this EI asymmetry account of the "associative" correlations between the a parameter and evaluative CS ratings. Using the same data-generating mechanism as in Simulation study 1, we again simulated (asymmetrical) individual EI probabilities and used them to generate evaluative CS classifications that were driven by propositional valence or guessing, but never by the mere US valence of a given CS. Furthermore, we used the same EI probabilities to independently simulate a second measure of CS evaluation representing a (somewhat simplified) rating measure. We then analyzed the CS classification responses with the previously introduced MPT model and investigated the correlations between the resulting parameter estimates and the CS ratings. Based on the previously introduced rationale, we expected positive correlations between the a parameter and evaluative ratings for positively paired CSs, and negative correlations between the a parameter and evaluative ratings for negatively paired CSs. Furthermore, we also expected to replicate the "propositional" correlations between the p parameter and evaluative CS ratings reported by Kukken et al.: participants with larger (smaller) p parameters produced more positive ratings for propositionally positive CSs (i.e., those paired with an assimilative [contrastive] relation and positive [negative] USs) and more negative ratings for propositionally negative CSs (i.e., those paired with an assimilative [contrastive] relation and negative [positive] USs). These predictions were based on the fact that the p parameter should be particularly large (small) for participants with high (low) EI probabilities in both *CS-US relation* conditions, and the very same participants should produce evaluative ratings that more (less) strongly reflect the propositional valence of a given CS (for all four *US valence* × *CS-US relation* conditions alike).

Method

Sample size and design. We simulated a single sample of 500 participants. The data simulation followed a 2 (*US Valence*: "positive" vs. "negative") × 2 (*CS-US relation*: "assimilative" vs. "contrastive") within-subjects design.

Simulation of individual inference probabilities. The individual EI probabilities for the assimilative and contrastive conditions were again generated by drawing a random value from a normal distribution and transforming it into its corresponding cumulative

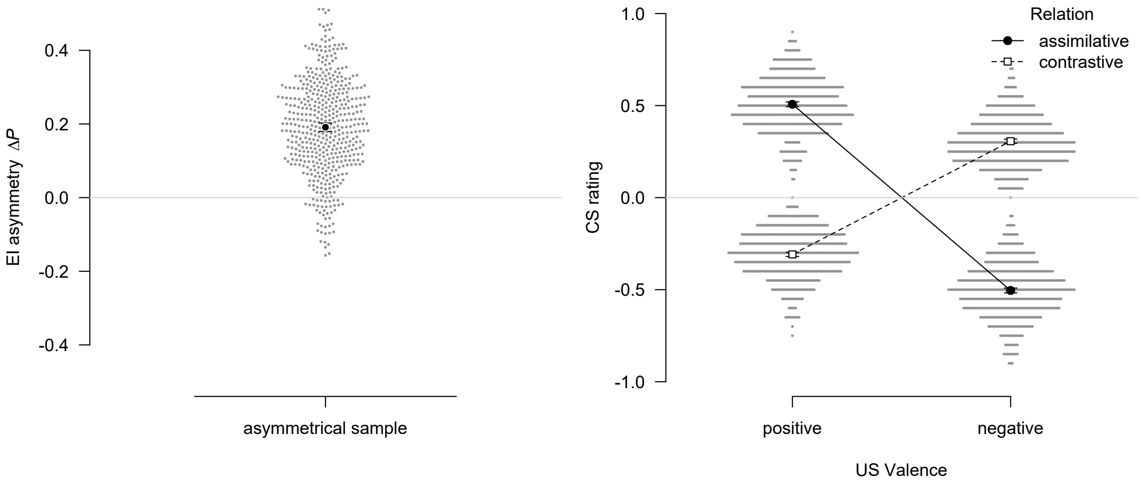

Figure 4.4. Simulation study 2: Individual EI asymmetries (left panel) and CS ratings as a function of US valence and CS-US relation (right panel).

probability. We simulated an "asymmetrical" sample using the same parameters as for the "asymmetrical" sample in Simulation study 1: $\mu_{\text{ass}} = 0$, $\mu_{\text{con}} = -0.5$ $\sigma_{\text{ass}} = \sigma_{\text{con}} = 0.25$; $\rho_{\text{ass,con}} = 0$). Accordingly, the sample mean of the individual EI probabilities was higher in the assimilative ($P_{\text{ass}} = .501$) than in contrastive condition ($P_{\text{con}} = .310$) resulting in a mean EI asymmetry of $\Delta P = P_{\text{ass}} - P_{\text{con}} = .191$ (see Figure 4.4, left panel). As in Simulation study 1, the individual EI probabilities in the assimilative and contrastive conditions were un-correlated with one another ($r = -.01$, 95% CI $[-.10, .08]$, $t(498) = -0.18$, $p = .857$); but correlated significantly with the individual EI asymmetries (assimilative condition: $r = .75$, 95% CI $[.71, .79]$, $t(498) = 25.26$, $p < .001$; contrastive condition: $r = -.67$, 95% CI $[-.71, -.62]$, $t(498) = -20.04$, $p < .001$).

Simulation of individual CS ratings. For each participant, we simulated a total of 80 CS ratings (20 per US valence × CS-US relation condition) using the following algorithm. For each CS, we sampled an EI status from a Bernoulli distribution with a *1* (*0*) indicating that an EI was available (unavailable) for a given CS. As for the simulation of CS classifications, the hit probability of the Bernoulli distribution depended on the level of *CS-US relation* of a given CS, but was independent of its level of *US valence*. In the next step, the CS ratings were generated: all CSs with an available EI received a rating according to their propositional valence; specifically, CSs presented with positive (negative) USs and an assimilative (contrastive) relation received a rating of $+1$, and CSs presented with negative (positive) USs and an assimilative (contrastive) relation received a rating of -1. All CS without an available EI received a neutral rating of 0. The resulting CS ratings featured the asymmetrical regular and reversed EC effects known from previous relational EC studies (see Figure 4.4, right panel).

Simulation of individual classification responses. The CS classifications were generated using the same algorithm as in Simulation study 1. As intended, the resulting response frequencies featured a higher share of propositionally correct CS classifications in the assimilative than in the contrastive conditions (see *Table 3*).

Table 5
Simulation study 2: Absolute (N) and relative (%) frequencies of propositionally correct CS classifications as a function of US valence and CS-US relation.

		Asymmetrical sample	
US valence	CS-US relation	*N*	%
positive	assimilative	7,527	75.27
	contrastive	6,591	65.91
negative	assimilative	7,524	75.24
	contrastive	6,530	65.30

Model analysis. The model analysis was conducted using the same methods and settings as in Simulation study 1. We again used posterior predictive *p*-values (based on the resampling of 5,000 posterior samples) to assess model fit.

Results

Model analysis. The results of the MPT analysis were highly similar to those for the asymmetrical sample in Simulation study 1. As before, the response frequencies were well accounted for by the MPT model, $T_1^{obs} = 0.01$, $T_1^{pred} = 0.01$, $p = 0.55$. The p parameter was larger than zero and indicated response generation based on propositional valence in 40.7% of classification trials (see *Table 4*). The a parameter was also larger than zero apparently indicating response generation based on US valence in 14.6% of classification trials for which response generation was not driven by propositional valence. The g parameter was again very close to .5 mirroring the unbiased guessing process by which classification responses in

Table 6
Simulation study 2: Posterior summaries of MPT parameter estimates.

	Asymmetrical sample	
Parameter	*M*	95% CI
Means		
p	.407	[.396, .418]
a	.146	[.125, .167]
g	.498	[.488, .507]
Correlations		
$\rho_{p,a}$	.265	[-.042, .552]
$\rho_{p,g}$	.118	[-.595, .764]
$\rho_{a,g}$	.060	[-.652, .751]

the absence of an evaluative inference were generated.

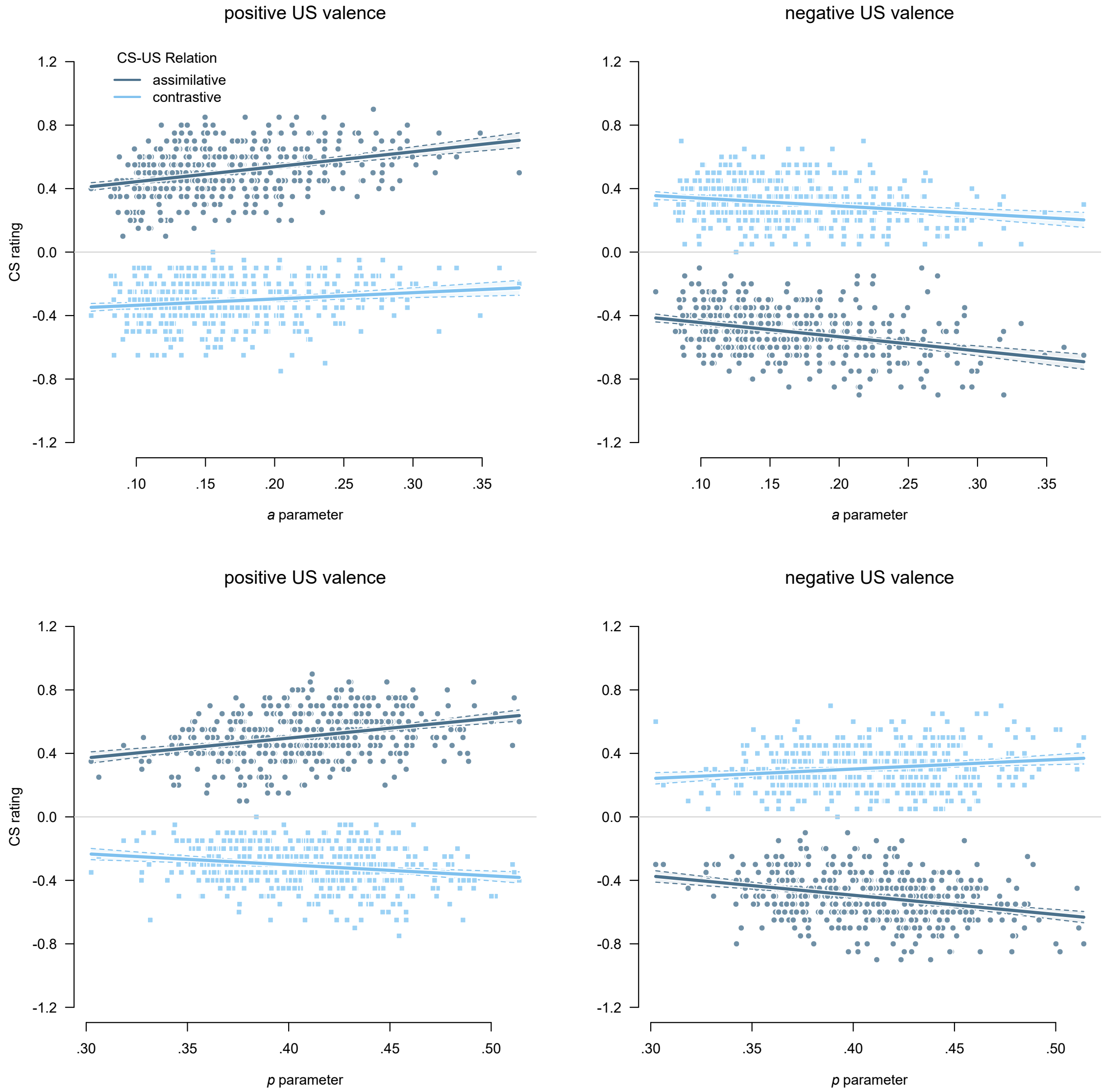

Figure 4.5. Simulation study 2: CS ratings as a function of MPT parameters. Solid lines represent linear regression slopes, gray areas represent 95% confidence intervals.

Correlations between MPT parameters and CS ratings. As expected, the correlations between the a parameter and the (mean) CS ratings followed an "associative" pattern (see upper row in Figure 4.5): for CSs "presented" with positive USs, larger a parameters were associated with more positive CS ratings in the assimilative, $r = .35$, 95% CI $[.27, .42]$, $t(498) = 8.34$, $p < .001$, as well as in the contrastive condition, $r = .16$, 95% CI $[.08, .25]$, $t(498) = 3.68$, $p < .001$; equivalently, for the CSs "presented" with negative USs, larger a parameters came with more negative CS ratings in the assimilative, $r = -.32$, 95% CI $[-.40, -.24]$, $t(498) = -7.64$, $p < .001$, as well as in the contrastive condition, $r = -.20$, 95% CI $[-.29, -.12]$, $t(498) = -4.65$, $p < .001$.

The correlations between the p parameter and the (mean) CS ratings also matched our expectations: for all four US valence × CS-US relation conditions alike, "participants" with larger p parameters gave more propositionally correct CS ratings (see lower row in Figure 4.5). For CSs "presented" with positive USs and an assimilative relation, larger p parameters were associated with more positive CS ratings, $r = .32$, 95% CI $[.24, .40]$, $t(498) = 7.54$, $p < .001$. By contrast, for CSs also "presented" with positive USs but a contrastive relation, larger p parameters came with more negative CS ratings, $r = -.20$, 95% CI $[-.28, -.11]$, $t(498) = -4.49$, $p < .001$. For CSs "presented" with negative USs, the pattern was reversed: here, larger p parameters were associated with more negative CS ratings in the assimilative condition, $r = -.30$, 95% CI $[-.38, -.22]$, $t(498) = -7.13$, $p < .001$, but with more positive CS ratings in the contrastive condition, $r = .17$, 95% CI $[.08, .25]$, $t(498) = 3.85$, $p < .001$.

Discussion

Simulation study 2 demonstrated that the "associative" correlations between the a parameter and evaluative CS ratings are compatible with an a parameter brought about by an EI asymmetry between the assimilative vs. contrastive conditions. Furthermore, the data-generating mechanism driven by propositional valence and unbiased guessing also produced the "propositional" correlations between the p parameter and evaluative CS ratings reported by Kukken et al. Taken together, these findings suggest that, while indicating that evaluative CS classifications and ratings are informed by overlapping sets of mental processes, the correlations between the MPT parameters and evaluative CS ratings cannot clarify whether one of these processes actually derives evaluative responses from US valence or not. Consequently, distinguishing between an a parameter driven by a systematic influence of US valence (whether due to CS-US associations or partial retrieval) and an a parameter based on an EI asymmetry will have to rely on other investigative approaches.

One such approach was pursued by Heycke et al. who experimentally manipulated several factors related to the encoding and retrieval of memory for the CS-US propositions, and observed their impact on the size of the p as well as the a parameter. For the p parameter, the manipulations produced a consistent effect reflecting its propositional underpinnings: for all four factors alike, the p parameter was substantially larger in the experimental condition for which more memory for CS-US propositions should be available; that is, when participants were given more time for encoding the CS-US-relation triplets during the learning procedure (Exp. 2), when the triplets were presented more often during the learning procedure (Exp. 3), when participants had more time for generating their response in the CS classification task (Exp. 4), and when the time delay between the learning procedure and the CS classification was comparatively brief (Exp. 5). By contrast, the effects of the four memory-related factors on the a parameter were less consistent: whereas all four manipulations produced descriptively larger a parameters in the condition with higher memory availability, only the effect of giving participants more vs. less response time during the classification task reached statistical significance. As discussed by Heycke et al., the direction of this effect poses a serious challenge to both the *CS-US association* and *partial retrieval* accounts of the a parameter. Given that the retrieval of CS-US associations is assumed to be highly efficient and resource-independent, an a parameter driven by CS-US associations should be unaffected by the manipulation of time availability during the CS classification task. In turn, as more time for response generation should raise the likelihood of full memory retrieval, an a parameter reflecting partial retrieval of CS-US propositions should be decreased when participants are given more time for determining their response. By contrast, the inference asymmetry account predicts the reported increase in the a parameter whenever more memory for the CS-US propositions can be retrieved. In the next section, we first introduce the mechanism underlying this claim and then present a simulation study demonstrating its viability in explaining the surprising findings reported by Heycke et al.

Simulation study 3

Simulation study 1 demonstrated that the larger-than-zero a parameter can be explained by a mental process that derives propositionally correct CS classifications from CS-US propositions, albeit with a lower probability in the contrastive than in the assimilative condition. Such a data-generating mechanism has two crucial components: on the one hand, the probability with which an evaluative inference is drawn from a given CS-US proposition; and on the other hand, the probability with which such a CS-US proposition is available in memory during the evaluative classification task[25]. Since the (asymmetrical) inference probabilities are the crucial force behind the emergence of a larger-than-zero a parameter, Simulation study 1 focused on the first component while assuming perfect memory availability of CS-US propositions in all four *US valence* $\times$ *CS-US relation* conditions. However, this is obviously not a realistic scenario as memory for CS-US propositions should vary greatly depending on the circumstances of its formation, storage and retrieval. Crucially, the degree to which such memory is available has a strong impact on both the resulting p and a parameters. This is due to the fact that the evaluative inferences underlying the classification responses (and therefore the MPT parameters) can only be drawn whenever a CS-US proposition is available. Consequently, a large p parameter will only emerge when a) the mean inference probability (across the two *CS-US relation* conditions) is high, and b) the probability of memory for the CS-US propositions is high; by contrast, a small p parameter will emerge when at least one of the two aforementioned probabilities is low. An equivalent logic applies to the a parameter: whenever memory for the CS-US propositions is low, a large share of classification trials will be driven by a guessing process that is identical for all four *US valence* $\times$ *CS-US relation* condition and will therefore result in relatively more comparable shares of propositionally correct classifications across the two

[25] Alternatively, one could assume that (at least some of) the evaluative inferences are drawn already during the learning procedure and therefore retrieved directly during the evaluative classification task. In this case, availability of CS-US propositions refers to the degree to which participants are aware of the propositional information connected to a given CS during the learning procedure.

CS-US relation conditions. By contrast, whenever memory for CS-US propositions is readily available, a larger share of classification trials will be affected by the unequal inference probabilities between the assimilative vs. contrastive conditions resulting in relatively less comparable shares of propositionally correct classifications for the two *CS-US relation* conditions. Therefore, the a parameter's increase under conditions of improved memory retrieval (i.e., when participants are given more time for generating a classification response) is easily accounted for by a data-generating mechanism driven by (asymmetrical) inference probabilities and guessing.

In order to corroborate this EI probability account of the previously mentioned finding, we conducted a third simulation study in which we compared the a parameters in two samples. For both samples alike, we simulated a systematic inference asymmetry between the assimilative and contrastive conditions which was comparable across the two samples. Additionally, we simulated individual memory retrieval probabilities which were systematically higher in the "high memory" sample than in the "low memory" sample. Based on the previously introduced mechanism, we expected the "high memory" sample to produce a substantially larger a parameter than the "low memory" sample. Given that a higher probability of retrieving a CS-US proposition allows for higher shares of propositionally correct CS classifications in both *CS-US relation* conditions alike, we also expected to find a larger p parameter in the sample with higher probabilities for memory retrieval.

Method

Sample size and design. For each level of "Memory retrieval", 500 participants were simulated. The data simulation followed a 2 (*Memory retrieval*: "low" vs. "high") × 2 (*US Valence*: "positive" vs. "negative") × 2 (*CS-US relation*: "assimilative" vs. "contrastive") mixed design with the first factor varying between "participants".

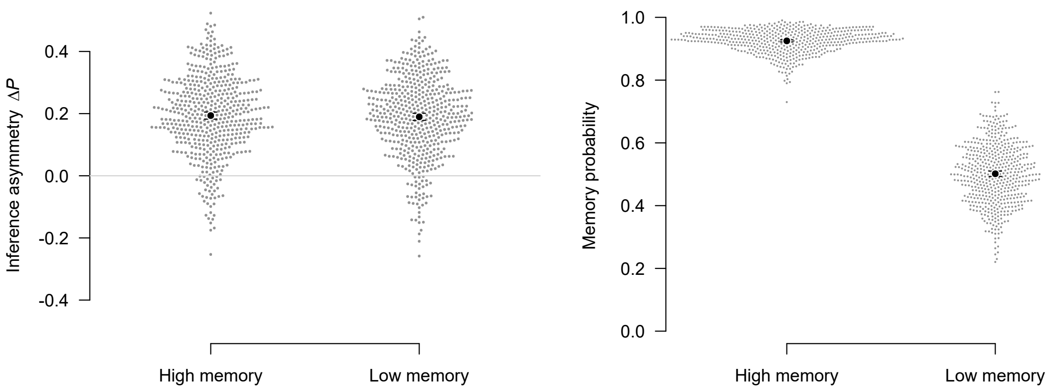

Figure 4.6. Simulation study 3: Individual inference asymmetries (left panel) and memory probabilities (right panel) in the "high memory" and "low memory" conditions.

Simulation of individual inference probabilities. The individual inference probabilities for the assimilative and contrastive conditions were again generated by drawing a random value from a normal distribution and transforming it into its corresponding cumulative probability For both levels of *Memory retrieval*, we simulated an "asymmetrical" sample using the same parameters as in the previous simulation studies: $\mu_{\text{ass}} = 0$, $\mu_{\text{con}} = -0.5$ $\sigma_{\text{ass}} = \sigma_{\text{con}} = 0.25$; $\rho_{\text{ass,con}} = 0$. For both samples alike, the sample mean of the individual inference probabilities was higher in the assimilative (*high memory*: $P_{\text{ass}} = .504$; *low memory*: $P_{\text{ass}} = .499$) than in contrastive condition (*high memory*: $P_{\text{con}} = .311$; *low memory*: $P_{\text{con}} = .309$) resulting in a mean inference asymmetry of .194 and .190 in the samples with high vs. low memory, respectively (see Figure 4.6, left panel). As before, the individual inference probabilities in the assimilative and contrastive conditions were un-correlated with one another (*high memory*: $r = -.03$, 95% CI $[-.12, .05]$, $t(498) = -0.73$, $p = .463$; *low memory*: $r = .03$, 95% CI $[-.05, .12]$, $t(498) = 0.77$, $p = .440$). Finally, both samples featured a positive correlation between the individual inference asymmetries and the individual inference probabilities in the assimilative condition (*high memory*: $r = .74$, 95% CI $[.70, .78]$, $t(498) = 24.58$, $p < .001$; *low memory*: $r = .72$, 95% CI $[.68, .76]$, $t(498) = 23.28$, $p < .001$), and a negative correlation between the individual inference asymmetries and the individual inference probabilities in the contrastive condition (*high memory*: $r = -.70$, 95% CI $[-.74, -.65]$, $t(498) = -21.65$, $p < .001$; *low memory*: $r = -.67$, 95% CI $[-.71, -.61]$, $t(498) = -19.96$, $p < .001$).

Simulation of individual memory probabilities. The individual memory retrieval probabilities were also generated by drawing a random number from a normal distribution and transforming it into its corresponding cumulative probability. For the "high memory" condition, we sampled from a normal distribution with a mean of $\mu_{\text{high}} = 1.5$ and a standard deviation of $\sigma_{\text{high}} = 0.25$. The resulting individual memory retrieval probabilities had a sample mean of $P_{\text{high}} = .925$ (see Figure 4.6, right panel) and were un-correlated with the individual inference probabilities (assimilative condition: $r = .00$, 95% CI $[-.09, .09]$, $t(498) = 0.00$, $p = .997$; contrastive condition: $r = -.07$, 95% CI $[-.16, .02]$, $t(498) = -1.60$, $p = .111$) as well as with the individual inference asymmetries ($r = .05$, 95% CI $[-.04, .14]$, $t(498) = 1.07$, $p = .285$). For the "low memory" condition, we sampled from a normal distribution with a mean of $\mu_{\text{low}} = 0$ and a standard deviation of $\sigma_{\text{low}} = 0.25$. The resulting individual memory retrieval probabilities had a sample mean of $P_{\text{low}} = .502$ (which was markedly lower than in the "high memory" sample) and were un-correlated with the individual inference probabilities (assimilative condition: $r = .01$, 95% CI $[-.08, .09]$, $t(498) = 0.16$, $p = .876$; contrastive condition: $r = .01$, 95% CI $[-.08, .10]$, $t(498) = 0.19$, $p = .848$) as well as with the individual inference asymmetries ($r = .00$, 95% CI $[-.09, .09]$, $t(498) = -0.02$, $p = .987$).

Simulation of individual classification responses. For each participant, we simulated a total of 80 CS classifications (20 per US valence × CS-US relation condition) using the following algorithm. For each CS, we first sampled a memory status from a Bernoulli distribution with a *1* (*0*) indicating that a relational CS-US proposition was available

Table 7
Simulation study 3: Absolute (N) and relative (%) frequencies of propositionally correct CS classifications as a function of Memory formation, US valence and CS-US relation.

		High memory		Low memory	
US valence	CS-US relation	*N*	%	*N*	%
positive	assimilative	7,380	73.80	6,190	61.90
	contrastive	6,521	65.21	5,736	57.36
negative	assimilative	7,392	73.92	6,167	61.67
	contrastive	6,401	64.01	5,719	57.19

(unavailable) for a given CS. The hit probability of this Bernoulli distribution was independent of both *CS-US relation* and *US valence*; specifically, for each CS, the hit probability was set equal to the individual memory retrieval probability of a given "participant". In the next step, an inference status was generated: for each CS with an available CS-US proposition, we sampled from a Bernoulli distribution with a *1* (*0*) indicating that an evaluative inference was available (unavailable) for a given CS. As in the previous simulation studies, the hit probability of this Bernoulli distribution was set equal to the individual inference probability of a given CS' level of CS-US relation. For all CSs for which a relational CS-US proposition was unavailable, the inference status was set to *0*. Finally, the CS classifications were generated: all CSs with an available inference received a classification response according to their propositional valence; i.e., a *positive* classification for CSs presented with positive (negative) USs and an assimilative (contrastive) relation, and a *negative* classification for CSs presented with negative (positive) USs and an assimilative (contrastive) relation. For each CS without an available inference, we simulated an unbiased guessing process by sampling a positive (*1*) vs. negative (*0*) response from a Bernoulli distribution with a hit probability of .5. As intended, the (overall) share of propositionally correct CS classifications as well as the classification asymmetry between the assimilative vs. contrastive conditions was larger in the "high memory" sample than in the "low memory" sample (see *Table 5*).

Model analysis. The model analysis and assessment of model fit were conducted using the same methods and settings as in the previous two simulation studies. For the comparisons of MPT parameters, we used the *betweenSubjectMPT* function of the *TreeBUGS* package (Heck et al., 2018).

Results

The response frequencies in the two samples were well accounted for by the MPT model ("high memory" condition: $T_1^{obs} = 0.01$, $T_1^{pred} = 0.01$, $p = 0.43$; "low memory" condition: $T_1^{obs} = 0.00$, $T_1^{pred} = 0.01$, $p = 0.60$). For both samples alike, the p parameter was substantially larger than zero indicating response selection based on propositional valence in 38.4% and 18.6% of classification trials in the "high memory" vs. "low memory" sample

Table 8
Simulation study 3: Posterior summaries of MPT parameter estimates.

	High memory		Low memory	
Parameter	M	95% CI	M	95% CI
Means				
p	.384	[.373, .394]	.186	[.175, .197]
a	.132	[.111, .152]	.047	[.030, .063]
g	.494	[.486, .503]	.501	[.494, .507]
Correlations				
$\rho_{p,a}$	.219	[-.154, .572]	.311	[-.427, .807]
$\rho_{p,g}$	.073	[-.677, .787]	.222	[-.568, .824]
$\rho_{a,g}$	.102	[-.702, .796]	.253	[-.670, .869]

respectively (see *Table 6*). As expected, the p parameter was substantially larger in the "high memory" sample than in the "low memory" sample, $\Delta_p = .197$, 95% CI $[.182, .213]$. Furthermore, the a parameters were also substantially larger than zero indicating response selection based on US valence in 13.2% and 4.7% of classification trials (for which response selection was not driven by propositional valence) in the "high memory" vs. "low memory" sample respectively. Again according to expectation, the a parameter was substantially larger in the "high memory" sample than in the "low memory" sample, $\Delta_a = .084$, 95% CI $[.058, .111]$. Finally, the g parameters of the two samples were both close to .5 and did not differ from one another, $\Delta_a = -.006$, 95% CI $[-.017, .005]$.

Discussion

Simulation study 3 demonstrated that a higher availability of memory for the CS-US propositions amplifies the effect of the asymmetrical inference probabilities on response generation in the evaluative classification task. In contrast to the correlations reported by Kukken et al., the a parameter's increase under conditions of improved memory retrieval (reported by Heycke et al.) can therefore distinguish between the inference asymmetry account and explanations of the a parameter that posit an independent effect of US valence: while said finding is inconsistent with the dual-process and partial retrieval accounts, it is predicted by and therefore consistent with the inference asymmetry account of the a parameter. However, time availability during response generation is not the only critical factor that may distinguish between the competing accounts of the a parameter. In fact, any manipulation which raises the availability of memory for the CS-US propositions should produce an increase in an a parameter based on an inference asymmetry. By contrast, the dual-process and partial retrieval accounts each make a specific set of predictions that differ from one manipulation to the other: while an a parameter based on CS-US associations should be increased by presenting the CS-US-relation triplets more often during the learning procedure and unaffected by giving participants more time for encoding the triplets as well as by a shorter delay between the learning procedure and the classification task, an *a* parameter driven by partial retrieval should be unaffected by all of these three factors[26]. As previously stated, Heycke et al. did not find any significant effects of these factors on the size of the *a* parameter. However, it is worth noting that the descriptive changes in the size of the *a* parameter again matched the predictions of the inference asymmetry account: that is, said MPT parameter was larger in all experimental conditions with a (presumably) higher availability of memory for the CS-US propositions. Furthermore, the obtained pattern of a significant increase of the *p* parameter combined with a non-significant increase of the *a* parameter is exactly what we found whenever we simulated a smaller difference in memory availability between the “high” vs. “low memory” samples (otherwise using the exact same setup as in Simulation study 3). This is due to the fact that an increase in memory availability has a relatively larger effect on the overall share of propositionally correct classification reflected by the *p* parameter, and a relatively smaller effect on the classification asymmetry between the assimilative vs. contrastive conditions reflected by the *a* parameter. Consequently, the insignificant effects of the tested memory manipulations on the *a* parameter may simply reflect their relatively lower impact on memory strength which may be compensated for by using stronger manipulations or more statistical power in future validation studies of the *a* parameter. In the course of the following section, we will expand on this and other avenues of future research opened up by the results of the three simulation studies; however, we will first summarize our results and relate them to current theoretical and methodological debates in research on (relational) evaluative conditioning.

General Discussion

Simulation study 1 demonstrated that a larger-than-zero *a* parameter can be explained by a data-generating mechanism driven by evaluative inferences and guessing: whenever the probability of drawing such an inference is lower in the contrastive than in the assimilative condition a classification asymmetry between the two *CS-US relation* conditions and consequently a larger-than-zero *a* parameter will emerge. Moreover, Simulation study 2 showed that a pattern of seemingly “associative” correlations between the *a* parameter and evaluative CS ratings are equally compatible with an *a* parameter based on an inference asymmetry between the two *CS-US relation* conditions: replicating the pattern reported by Kukken et al., we found positive (negative) correlations between evaluative CS ratings for positively (negatively) paired CSs and an *a* parameter fuelled by said inference asymmetry. Finally, Simulation study 3 showed that a previously inexplicable finding reported by Heycke et al. is easily accounted for by an *a* parameter based on an inference asymmetry: whenever the CS classification task allows for better retrieval of CS-US propositions from memory the effect of the asymmetrical inference probabilities on the CS classifications will be amplified and a larger *a* parameter will emerge.

[26]Note that these predictions are based on a specific understanding of partial retrieval as presented by Heycke et al. (2020) and that other versions of a partial retrieval account may allow for diverging predictions.

Before we discuss the theoretical and methodological implications of these simulation studies, we want to address a potential criticism concerning their generalizability. While we used a very specific set of parameter values for the simulation of the individual inference probabilities underlying the CS classifications in all three simulation studies, it is worth noting that the very same pattern of results can be produced by a wide variety of probability distributions and parameter values. In the course of our work on this topic, we have consistently been able produce the three critical findings using many different normal, truncated normal and beta distributions. In all of these simulations, the emergence of a larger-than-zero a parameter was solely dependent on the size of the simulated inference asymmetry, but independent of the standard deviations and correlations of the underlying distributions. Furthermore, the associative correlations between the a parameter and the CS ratings were found whenever we allowed for sufficient variance on the individual inference asymmetries and whenever the correlation between the inference probabilities in the assimilative and contrastive conditions were not too strong. Finally, the moderating effect of memory availability on the a parameter was obtained whenever the underlying inference asymmetry as well as the difference in memory availability for the "high" vs. "low memory" samples were sufficiently large. Taken together, we therefore conclude that the parameter settings of the three simulation studies represent more than a specific mathematical scenario which happens to produce the three critical findings; rather, they constitute one of many possible implementations of a simple, yet powerful mechanism that can explain the whole range of currently available findings connected to the a parameter.

Theoretical implications

The results of the three simulation studies have important implications for the debate over the mental underpinnings of the a parameter. First of all, as the implemented data-generating mechanism gave rise to a larger-than-zero a parameter without ever being driven by the US valence of a given CS, said MPT parameter cannot be equated with a direct effect of US valence on CS evaluation. While such an associative effect can explain both the classification asymmetry underlying the a parameter and the attenuated reversed EC effect on evaluative CS ratings, these two findings (as well as their correlations) are equally well accounted for by differential effects of assimilative vs. contrastive relations on CS evaluation. Furthermore, based on the current evidence, an inference asymmetry (reflecting the weaker impact of contrastive CS-US relations) actually poses the most coherent account of the findings brought about by the MPT approach to relational EC learning. This claim is based on the fact that the inference asymmetry account not only explains all of the findings that a dual-process or partial retrieval perspective can account for, but also those that are inconsistent with these two explanations of the a parameter (i.e., the effect of more vs. less time or weak vs. strong memory retrieval during the CS classification task). While we are confident that the findings so far obtained in MPT analyses of relational EC learning are best explained by an inference asymmetry between the two *CS-US relation* conditions, we do not want to imply that an additional contribution of an effect of US valence (due to CS-US associations and/or partial retrieval) to the classification asymmetry can be excluded. However, based on the moderating effect of memory availability during the classification task, we expect such a contribution to be relatively small in comparison to that of an inference asymmetry (if it exists at all).

The implications for theoretical accounts of (relational) EC effects mirror the previously raised implications for the different accounts of the *a* parameter. Due to the fact that the inference asymmetry account poses the most coherent explanation of the *a* parameter's characteristics, we conclude that the findings of the MPT approach to relational EC learning so far favor a purely propositional over a dual-process perspective on (relation) EC effects. Admittedly though, the explanatory advantage of a propositional account is currently based on a single finding and future validation studies may yet demonstrate moderating influences on the *a* parameter that are best explained by an effect of CS-US associations. However, we would like to emphasize that none of the current findings connected to the *a* parameter lend unique support to the notion of an independent effect of US valence on CS evaluation, and that any future finding implying such an effect can only do so to the extent that it makes unique predictions that do not follow from an inference asymmetry account.

Methodological implications

MPT analyses of relational EC learning pose a new and promising approach to disentangling the respective contributions of proposition vs. association formation to evaluative learning. Nevertheless, the results of the three simulation studies illustrate the particular challenge that comes with this approach: specifically, in order to provide strong evidence for an impact of CS-US associations, future demonstrations of a larger-than-zero *a* parameter will have to rigorously control for a potential inference asymmetry between assimilative vs. contrastive CS-US relations.

To the extent to which the inference asymmetry is driven by characteristics of the learning materials, one potential control strategy may consist in presenting participants with "symmetrical" material, i.e., CS-US-relation triplets that induce equally strong CS evaluations in the assimilative vs. contrastive conditions. Such a strategy would require extensive pre-testing ensuring that valence induction for the two *CS-US relation* conditions is not only comparable under the conditions implemented by the pre-test, but also during the actual learning experience in the relation EC procedure.

Furthermore, future MPT analyses of relational EC learning may also attempt to estimate an MPT model with two *p* parameters (one for the assimilative and one for the contrastive condition) as well as an *a* parameter (as a separate indicator of a potential influence of US valence). As already discussed by Kukken et al., the stable estimation of such an MPT model would require additional experimental conditions allowing to separate the effect of an inference asymmetry from that of CS-US associations. At this point, however, it is unfortunately not clear how such an extension to the experimental set-up would have to look like.

Future research

Our findings suggest a number of potentially attractive avenues for future research. First of all, the inference asymmetry account rests on the assumption of a weaker impact of contrastive CS-US relations on CS evaluation which is so far a purely theoretical claim. In order to provide an empirical test of this underlying assumption, future research may attempt to to identify material sets for which the CS evaluations induced by the two *CS-US relation* conditions are more vs. less comparable in strength. These material sets could then be compared with regard to the size of their respective a parameters emerging in an evaluative classification task of the CSs presented together with the (more vs. less symmetrical) US-relation combinations.

Another empirical test of the inference asymmetry account may consist of re-examining the effects of the other three memory-related factors tested by Heycke et al. (i.e., encoding time during the learning procedure, number of repetitions of the CS-US-relation triplets, time delay between the learning procedure and the classification task). As previously stated, their effects on the availability of memory for the CS-US propositions may be comparatively small. Additionally, the impact of these moderators on the size of the a parameter may have been restrained by the brief response window implemented in all studies other than the one explicitly testing the effect of more vs. less time for response generation. We therefore suggest that future studies seeking to re-evaluate the effects of said three moderators should implement a longer response window in the evaluative classification task: under such conditions, the potential effects of encoding time, of the frequency of triplet repetition and of the time delay should be able to yield a stronger influence on the size of an a parameter based on an inference asymmetry.

On a different note, future research may also focus on the nature and determinants of the (potentially) asymmetrical effects of assimilative vs. contrastive CS-US relations on CS evaluation. As mentioned earlier, such an asymmetry may represent idiosyncratic features of the materials typically employed in relational EC studies, and may therefore (partly) disappear whenever the relational EC procedure incorporates more balanced CS-US-relation triplets. In certain constellations, it may also be possible to obtain a reversed asymmetry between the two *CS-US relation* conditions; that is, CS evaluation may be more strongly affected by certain contrastive relations in comparison to certain assimilative relations. For example, some experimental designs may trigger the human tendency for loss aversion, i.e., the prioritization of avoiding losses over acquiring equivalent gains (Kahneman & Tversky, 1979): accordingly, certain pairs of assimilative and contrastive relations tapping into the concepts of gain and loss respectively may produce stronger EC effects in the contrastive than in the assimilative conditions (at least for positively paired CSs).

Additionally, prospective studies on the asymmetrical regular and reversed EC effects may also receive inspiration from research on human causal learning. For example, (Wu & Cheng, 1999) demonstrated that for preventive causal learning to emerge, the to-be-prevented effect must be present at least sometimes when the candidate preventive cause is absent. By extension, the size of the reversed EC effect for creatures stopping a positive vs. negative sound may depend on whether the learning procedure includes trials in which the stopping creature does not appear and sound duration is markedly longer (in comparison to trials where the stopping creatures do appear). In a similar vein, EC effects for CSs preventing a positive vs. negative health state may be affected by manipulations of participants' beliefs that a given health state is likely to befall them. Finally, the suggested similarities between causal and relation EC learning can also be taken to suggest that the recently introduced MPT approach to studying the asymmetrical effects of assimilative vs. contrastive relations will also prove useful in future examinations of generative vs. preventive causal learning.

Conclusions

MPT modeling of relational EC learning represents an undoubtedly promising approach to investigating the processes underlying evaluative learning and has quite a few advantages over earlier analytical approaches. Specifically, the recently introduced MPT model is capable of separating the potential effects of associations vs. propositions on a single measure of evaluation, provides estimators of their respective contributions to (relational) EC effects and also allows for a more direct test of possible moderators of EC effects mediated by CS-US associations vs. propositions. However, the findings of the three simulation studies illustrate the particular challenges posed by the MPT approach in gathering unambiguous evidence for association formation in (relational) EC learning. Crucially, future demonstrations of a larger-than-zero a parameter supposedly indicating response generation based on US valence will have to ensure that the underlying classification asymmetry is not in fact driven by a weaker effect of contrastive CS-US relations on CS evaluation. We believe that the construction of symmetrical CS-US-relation triplets as well as the development of statistical control strategies represent promising approaches in the pursuit of this aim.

Chapter 5

General Discussion

The mental processes underlying evaluative conditioning have been studied extensively and through numerous methodological approaches. These approaches have been highly effective in demonstrating, unambiguously, the involvement of propositional processes in mediating the effects of CS-US co-occurrences on CS evaluations. By contrast, a convincing demonstration of an additional, associative contribution to evaluative conditioning is still lacking, and appears to require innovative methodologies in order to ever be attained. In an attempt to promote this methodological evolution, the present thesis therefore examined the potentials and limitations of three promising approaches to demonstrating the involvement of association formation in evaluative conditioning. The results of these examinations will now be summarized and discussed with regard to their theoretical as well as methodological implications.

Summary of results

Chapter 2 reported a series of studies implementing a novel approach to investigating EC in the absence of awareness for the CS-US co-occurrences. In these studies, CS-US awareness was manipulated through a well established psycho-physical technique: as in several previous investigations, the CSs were shown only briefly and in close temporal succession with other stimuli masking their presentation. Departing from earlier research, these studies did however not rely on a CS identification task in order to ensure the effectiveness of this technique (at suppressing conscious visibility of the CSs and, by implication, awareness for the CS-US co-occurrences). Instead, my co-author and I sought to exclude conscious processing of the CS presentations through self-reported CS (in-)visibility on a customized perceptual awareness scale. By relying on such a subjective criterion of CS unawareness, the studies reported in Chapter 2 allowed for longer presentation durations and, accordingly, higher levels of (unconscious) CS processing than previous research using the subliminal CS presentation approach. Despite these improved conditions for CS-US co-activation (and, therefore, association formation), my co-author and I did not obtain significant EC effects in the absence of (subjective) awareness for the CS-US co-occurrences. Rather, conditioned changes in CS evaluation were found only when objective neural processing of the CSs (as indicated by above-chance identification performance) was accompanied by substantial CS visibility as measured by the perceptual awareness scale. Demonstrating EC's dependence on (subjective) awareness for the CS-US co-occurrences, the studies reported in Chapter 2 are well in line with numerous studies reporting significant EC effects only for consciously visible CS presentations (Dedonder et al., 2013; Heycke et al., 2017; Heycke & Stahl, 2018; Högden et al., 2018; Stahl et al., 2016), and do therefore not support an additional, associative contribution to EC.

Chapter 3 focused on a second methodological approach to demonstrating association formation in EC. In two studies, my co-authors and I tested a non-associative alternative account of the unqualified EC effect reported by Moran and Bar-Anan (2013). In these studies, Moran and Bar-Anan's learning procedure (presenting cartoon characters that either started or stopped positive vs. negative sounds) was combined with novel IAT variants. These IATs were specifically designed so that some of them allowed for diverging predictions of the original "associative" and the alternative "similarity construction" account, while others elicited identical predictions by the two accounts. Across these more or less discriminatory IAT variants, a consistent and highly informative pattern emerged: while the predictions of the "similarity construction" account were confirmed in the discriminatory as well as the non-discriminatory IATs, the predictions of the "associative" account were confirmed only in the non-discriminatory IATs (for which the "similarity construction" account makes an identical prediction). As such, this pattern demonstrates, unambiguously, the involvement of similarity construction in the experimental setting created by Moran and Bar-Anan (2013) and therefore corroborates a non-associative explanation of the unqualified EC effect in the original "stopping" IAT (in terms of a similarity construction based on propositional knowledge about the co-occurring US valences). Similar to other non-informative demonstrations of unqualified EC on indirect evaluation measures (Gawronski et al., 2015; Hu et al., 2017), the un-reversed preference in the "stopping" IAT can therefore not provide unique support for association formation in EC.

Chapter 4 addressed a recently introduced process dissociation approach to studying association formation in EC. Through a series of simulation studies, my co-authors and I explored the empirical implications of a non-associative alternative account of the findings obtained by Hütter and Sweldens (2018), Heycke and Gawronski (2020), and Kukken et al. (2020). In these studies, (simulated) classification asymmetries driven solely by a propositionally disadvantaged exclusion condition were found to result in MPT parameter estimates that mimicked an unqualified influence of US valence on the (simulated) CS classifications. Moreover, correlations between these parameter estimates and (simulated) CS ratings were found to reproduce a seemingly associative pattern which has so far been viewed as providing unique support for an associative account of the classification asymmetry and its corresponding MPT parameter (Hütter & Sweldens, 2018; Kukken et al., 2020). Finally, a propositionally disadvantaged exclusion condition was also found to predict a previously inexplicable increase of the supposedly associative MPT parameter under conditions of improved memory retrieval which contravenes the predictions of an associative explanation of its underlying classification asymmetry (Heycke & Gawronski, 2020). Taken together, the simulation studies reported in Chapter 4 thus show that the results of the novel process dissociation approach have so far failed to provide unique support for association formation in EC, and are to date better explained by a non-associative alternative account in terms of a propositionally disadvantaged exclusion condition.

Theoretical implications

Evaluative conditioning has featured heavily in the long-standing debate over single- vs. dual-process theories of human associative learning (e.g., Shanks, 2007; Mitchell et al., 2009). As previously explained, however, this prominent role was gained largely through initial evidence in favor of association formation that proved to be equally well explained by a (single-process) propositional account of evaluative conditioning. Later methodological approaches have either failed to confirm the unique predictions of associative (or, more precisely, dual-process) accounts, or —quite regularly— produced non-informative findings that are compatible with both purely propositional and dual-process theories of evaluative conditioning. As apparent by the preceding summary, the studies reported in this thesis continue this empirical trend. Addressing the currently most promising approaches to demonstrating an associative contribution to EC, unique evidence for association formation was either not found (Chapter 2), or shown to be in fact comprehensively explained by the constructive and inferential processes proposed by propositional accounts (Chapter 3 and 4). The main theoretical implication of this thesis is therefore well aligned with a growing consensus that, though impossible to refute, the assumption of a second, associative mediator of EC is simply unnecessary based on the currently available evidence (e.g., Corneille & Stahl, 2019).

Aside from providing general support for purely propositional accounts of evaluative conditioning, the studies presented in this thesis also bear implications for the specific contents of these theories. In response to recent reports of seemingly unqualified EC effects (Heycke & Gawronski, 2020; Hütter & Sweldens, 2018; Kukken et al., 2020; Moran & Bar-Anan, 2013), prominent proponents of propositional accounts have considered the possibility of a given EC effect being driven not just by one, but several propositions about the relation between the co-occurring stimuli (e.g., De Houwer, 2018). Importantly, some of these mediating propositions may not encode the specific relation between CS and US, and could therefore produce EC effects that run counter to what is implied by the omitted CS-US relation. To illustrate this scenario, consider once again the introductory example of “ESREN” and the rose graffito. Given a certain spatio-temporal configuration of the co-occurring graffiti, their joint sight may give rise to the previously mentioned “erasure” interpretation of “ESREN”, and —at the same time— a mere co-occurrence proposition simply stating that the two graffiti co-occurred. To the extent that these two propositions are differentially available across contexts, the “ESREN” tag could then elicit negative evaluations based on the “erasure” proposition in some situations, and positive evaluations derived from its co-occurrence with the beautiful rose graffito in others. Obviously, this notion of (sometimes) unqualified EC effects based on additional mere co-occurrence propositions bears clear resemblance to a second, associative contribution to EC, and therefore blurs —to some degree— the distinction between dual-process and purely propositional accounts of evaluative conditioning. As revealed by Chapters 3 and 4 of this thesis, however, such a blurring does not seem to be necessary based on the currently available demonstrations of unqualified EC. Specifically, neither Moran and Bar-Anan's "stopping" IAT effect nor the MPT modeling results reported by the previously mentioned authors teams appear to be based on genuinely unqualified sources of CS evaluation, and do therefore not require a propositional explanation in terms of additional, mere co-occurrence propositions.

A third and final implication of this thesis addresses the popular idea of propositionally mediated evaluations being based —necessarily— on a normatively rational use of available information. Given the meager information presented in a traditional EC procedure (i.e., mere stimulus pairings), it has thus been argued repeatedly that EC effects, as such, indicate non-rational and, by implication, associative mediators of human learning (e.g., Shanks & Dickinson, 1990; Shanks, 2007). As illustrated by the manifold IAT effects reported in Chapter 3 however, propositional use of information is highly flexible as well as context-dependent, and —to some degree— detached from normative conceptions of consistency and rationality. As such, this thesis therefore corroborates previous claims that propositional reasoning can explain both rational and (seemingly) irrational evaluations as well as other forms of associative learning (e.g., Mitchell et al., 2009).

Methodological implications and future research

As illustrated by the previous sections, the currently most prominent approaches to demonstrating an associative contribution to EC have so far failed to achieve their empirical goal. Continued research on this topic will therefore require, once again, novel (or modified) methodologies for investigating association formation in EC. In order to foster these methodological developments, the central findings of this thesis will now be discussed with regard to future research into the mental processes underlying evaluative conditioning.

To date, experimental suppression of CS visibility (and, by implication, CS-US awareness) represents the only convincing approach to precluding a propositional contribution to a given instance of EC, and therefore possesses great potential for demonstrating, unambiguously, an additional, associative contribution to this evaluative learning effect. Accordingly, the experimental approach to investigating EC without CS-US awareness featured heavily in recent research on association formation in EC, and has been implemented through several psycho-physical techniques. Despite such methodological diversity, the results of this methodological approach have been sobering: so far, neither continuous flash suppression nor subliminal or parafoveal stimulus presentations have allowed for significant EC effects in the absence of conscious processing of the CSs. In Chapter 2, a potential explanation for this lack of unaware EC in terms of an overly strict objective criterion of CS awareness was tested — and discarded. As previously explained, and of immediate theoretical importance, my co-author and I obtained significant EC effects only under learning conditions that allowed for substantial levels of objective as well as subjective CS awareness. Moreover, and more methodologically relevant, we did not find any dissociations between the two criteria of CS awareness: as indicated by correlational analyses, the presence (or absence) of objective awareness for the CS presentations was always accompanied by the presence (or absence) of subjective CS awareness. As such, the studies reported in Chapter 2 not only corroborate previous findings based on an objective awareness criterion, but also justify its continued use in future research on the role of CS-US awareness for the emergence of EC. At the same time, these studies also alleviate previous concerns over the potential non-exhaustiveness of subjective awareness measures (with regard to capturing all consciously processed information), and therefore identify subjective awareness criteria as a possible alternative (or supplement) to objective criteria of stimulus awareness.

As explained in the previous section, failures to obtain EC for visually suppressed CSs seem to be unrelated the specific methodology by which unawareness for the suppressed stimulus presentations is established. Future research adopting this experimental approach should therefore explore alternative methodological avenues for increasing the chances of demonstrating EC in the absence of awareness for the CS-US co-occurrences. For example, proponents of S-R models of EC have repeatedly argued that the learning conditions implemented in Olson and Fazio's surveillance procedure elicit incidental processing of the CS-US pairings, and thereby produce improved conditions for the formation of S-R associations (e.g., Olson & Fazio, 2001; Jones et al., 2009). As mentioned earlier, using the surveillance procedure cannot, by itself, exclude a propositional contribution to a given instance of EC. However, combining this modified EC procedure with an (established or novel) experimental suppression of CS visibility may promote the formation of mental associations, while at the same time precluding CS-US awareness and, by implication, propositionally mediated EC effects.

Even though future research (on association formation in EC) is well-advised in exhausting the full potential of the experimental approach to investigating EC without CS-US awareness, it seems equally sensible for such research to consider its possible limitations (and perhaps turn to more promising methodologies). For one thing, the subliminal presentation approach has often been criticized for its weak ecological validity (e.g., Bargh & Morsella, 2008): brief and sandwich-masked stimulus presentations do not occur naturally and may therefore not represent appropriate input for a mental process likely to have evolved in tandem with the regular strength stimuli that are part of the natural environment. While this critique (in terms of low ecological validity) is less applicable to parafoveal stimulus presentations as well as continuous flash suppression, the usefulness of these (and other) psycho-physical techniques in investigating association formation can still be criticized on more principled grounds. Specifically, precluding awareness of the CS-US co-occurrences through any visual suppression of the CS presentations may be taken to introduce an unduly expansive understanding of CS-US unawareness — entailing not only an absence of awareness for the fact that two stimuli co-occur, but also a lack of conscious processing of (one of) the co-occurring stimuli as such. Although this conceptual distinction is not explicitly addressed in many associative accounts, its importance seems to be (indirectly) acknowledged by certain S-R models emphasizing source confusability between clearly visible CS and US presentations (with regard to the affective UR) as the prime determinant of S-R link formation (Jones et al., 2009). However, since a disruption of conscious CS processing is currently indispensible for precluding a propositional contribution to a given instance of EC, its potentially undermining effect on association formation cannot be resolved within the experimental approach to investigating EC without CS-US awareness (at least, in its current form). Accordingly, future attempts at demonstrating an associative contribution to EC should also focus on other methodologies that do not require visually suppressed stimulus presentations in order to provide unique support for association formation in EC.

Whenever an EC procedure allows for conscious processing of the CS-US co-occurrences, propositional and (potential) associative contributions to EC need to be disentangled during the measurement of the conditioned CS evaluations. As demonstrated by Chapters 3 and 4 of this thesis, such disentangling of propositional and associative sources of CS evaluation has not yet been achieved, and seems to be far more demanding than originally thought. For the task dissociation approach (seeking to demonstrate association formation through unqualified EC effects on indirect evaluation measures), the biggest challenge stems from the fact that, contrary to their founding idea, indirect measures do not provide preferential access to mental associations and are in fact highly receptive to the (evaluative) outcomes of propositional processes (for a discussion, see Corneille & Hütter, 2020). In the search for a convincing demonstration of association formation in EC, this receptivity can have several detrimental effects. First of all, sensitivity to propositionally mediated CS evaluations may mask additional associative influences (rendering a shift from direct to indirect evaluation measures rather pointless). Secondly, propositional processes are highly flexible and attuned to situational demands, and may therefore mimick associative sources of CS evaluation whenever such mimicry is incentivized by an indirect measurement procedure. As illustrated by Chapter 3, this possibility is particularly likely in the IAT (and related measures such as the sorting paired features task, Bar-Anan, Nosek, & Vianello, 2009) in which active CS classification may be improved through strategic use of (propositional) information. Finally, the findings reported in Chapter 3 can also be framed in terms of a third methodological disadvantage implied by the propositional receptivity of indirectly measured CS evaluations. Assuming for example that the unqualified EC effect in the stopping IAT reflects a genuinely associative source of CS evaluation, similarity construction in other IAT variants (or similar propositional influences on indirect evaluation measures) may also undermine actual evidence in favor of association formation in EC. In response to these methodological weaknesses, future demonstrations of unqualified EC on indirect evaluation measures will have to be accompanied not only by a significantly reversed EC effect on direct measures of CS evaluation, but also by appropriate control conditions excluding potential (propositionally mediated) artefacts of the indirect measurement procedure. Due to its susceptibility to similarity construction, the IAT might not be able to meet the latter requirement without casting doubt on actual demonstrations of associatively mediated EC, and may therefore be replaced by other indirect measurement procedures in future implementations of the task dissociation approach. However, most of these alternative indirect measures have already been used in previous research based on this approach, and —without exception— failed to produce convincing demonstrations of unqualified EC (e.g., Peters & Gawronski, 2011; Hu et al., 2017; Zanon, De Houwer, & Gast, 2012). Based on these sobering results, future investigations of association formation in EC should also explore other methodological approaches to demonstrating unqualified EC effects.

Avoiding both visually suppressed stimulus presentations and potential artefacts of indirect evaluation measures, the process dissociation approach to investigating unqualified EC effects represents an undoubtedly promising methodology for demonstrating association formation in EC. However, as illustrated by Chapter 4 of this thesis, the crucial classification asymmetry (and its corresponding MPT parameter) can also be brought about by a propositionally disadvantaged exclusion condition, and does therefore not, by itself, demonstrate unqualified sources of CS evaluation. As indicated by the significant increase in the supposedly associative MPT parameter under conditions conducive to memory retrieval, the findings of Heycke and Gawronski (2020) seem to be driven —at least, in part— by this non-associative mechanism. Moreover, Hütter and Sweldens (2018) reported two experimental manipulations that produced equivalent (yet non-significant) increases in the associative MPT parameter, and also found the seemingly associative correlations between the supposedly associative MPT parameter and evaluative CS ratings that are predicted (albeit non-uniquely) by a propositionally disadvantaged exclusion condition. Finally, due to the non-informative nature of these correlations (with regard to the two accounts of the classification asymmetry), and a lack of other, more informative validation studies, a propositional disadvantage in the exclusion conditions implemented by Kukken et al. (2020) can neither be confirmed nor rejected. Taken together, and in contrast to a widespread assumption (e.g., Corneille & Hütter, 2020; De Houwer, Van Dessel, & Moran, 2020; Gawronski, Brannon, & Luke, 2021), the process dissociation approach has therefore not yet achieved an unambiguous demonstration of unqualified EC. To accomplish this goal, the following methodological guidelines should be observed. Firstly, future implementations of the process dissociation approach should treat a propositionally disadvantaged exclusion condition as a likely scenario, and therefore make explicit efforts to avoid its actual occurrence. However, depending on its exact (and yet unknown) origins, a propositional disadvantage in the exclusion condition may not always be completely preventable despite considerable expenditures. Accordingly, future studies adopting the process dissociation approach should always include at least one highly-powered test of a unique prediction implied by a propositionally disadvantaged exclusion condition. Finally, these studies should also identify and test unique predictions of genuinely unqualified sources of a classification asymmetry between their implemented inclusion and exclusion conditions. Observing these methodological guidelines, future implementations of the process dissociation approach may very well produce convincing demonstrations of unqualified EC. However, due to the previously mentioned possibility of unqualified EC based on mere co-occurrence propositions, the potential successes of the process dissociation approach may not necessarily count as unique evidence for an additional, associative contribution to EC. Demonstrating association formation via unqualified EC (whether based on the task or process dissociation approach) will therefore not only require methodological rigor in addressing the previously discussed artefacts, but also depend on future theorizing on the evaluative influence of mere co-occurrence propositions.

References

Avneon, M., & Lamy, D. (2018). Reexamining unconscious response priming: A liminal-prime paradigm. *Consciousness and Cognition*, *59*, 87–103. https://doi.org/10.1016/j.concog.2017.12.006

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*(4), 179–199. https://doi.org/10.1016/0146-6402(88)90014-8

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*(2), 133–142. https://doi.org/10.1016/0005-7967(92)90136-5

Baeyens, F., Hermans, D., & Eelen, P. (1993). The role of CS-US contingency in human evaluative conditioning. *Behaviour Research and Therapy*, *31*(8), 731–737. https://doi.org/10.1016/0005-7967(93)90003-D

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task: A measure of association strengths. *Experimental Psychology*, *56*(5), 329–343.

Bargh, J. A., & Morsella, E. (2008). The Unconscious Mind. *Perspectives on Psychological Science*, *3*(1), 73–79. https://doi.org/10.1111/j.1745-6916.2008.00064.x

Bengson, J. J., & Hutchison, K. A. (2007). Variability in Response Criteria Affects Estimates of Conscious Identification and Unconscious Semantic Priming. *Consciousness and Cognition*, *16*(4), 785–796. https://doi.org/10.1016/j.concog.2006.12.002

Biegler, P., & Vargas, P. (2013). Ban the Sunset? Nonpropositional Content and Regulation of Pharmaceutical Advertising. *The American Journal of Bioethics*, *13*(5), 3–13. https://doi.org/10.1080/15265161.2013.776127

Biegler, P., & Vargas, P. (2016). Feeling Is Believing: Evaluative Conditioning and the Ethics of Pharmaceutical Advertising. *Journal of Bioethical Inquiry*, 1–9. https://doi.org/10.1007/s11673-016-9702-8

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*(3), 242–261. https://doi.org/10.1207/S15327957PSPR0603_8

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 19681987. *Psychological Bulletin*, *106*(2), 265–289. https://doi.org/10.1037/0033-

2909.106.2.265

Breitmeyer, B. G. (2015). Psychophysical "blinding" methods reveal a functional hierarchy of unconscious visual processing. *Consciousness and Cognition*, *35*, 234–250. https://doi.org/10.1016/j.concog.2015.01.012

Brendl, M. C., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*(5), 760–773. https://doi.org/10.1037/0022-3514.81.5.760

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics*, *36*(4), 387–395.

Corneille, O., & Hütter, M. (2020). Implicit? What Do You Mean? A Comprehensive Review of the Delusive Implicitness Construct in Attitude Research. *Personality and Social Psychology Review*, 1088868320911325. https://doi.org/10.1177/1088868320911325

Corneille, O., & Stahl, C. (2019). Associative attitude learning: A closer look at evidence and how it relates to attitude models. *Personality and Social Psychology Review*, *23*(2), 161–189. https://doi.org/10.1177/1088868318763261

Corneille, O., Yzerbyt, V., Pleyers, G., & Mussweiler, T. (2009). Beyond awareness and resources: Evaluative conditioning may be sensitive to processing goals. *Journal of Experimental Social Psychology*, *45*(1), 279–282. https://doi.org/10.1016/j.jesp.2008.08.020

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. In *Seminars in the Neurosciences* (Vol. 2, pp. 263–275). Saunders Scientific Publications.

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, *6*(2), 119.

Dawson, M. E., & Schell, A. M. (1985). Information Processing and Human Autonomic Classical Conditioning. *Advances in Psychophysiology*, *1*, 89–165.

Dedonder, J., Corneille, O., Bertinchamps, D., & Yzerbyt, V. (2013). Overcoming Correlational Pitfalls Experimental Evidence Suggests That Evaluative Conditioning Occurs for Explicit But Not Implicit Encoding of CSUS Pairings. *Social Psychological and Personality Science*. 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*(6), 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*(5), 204–211. https://doi.org/10.1016/j.tics.2006.03.007

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, *37*(6), 443–451.

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation*, *37*(2), 176–187. https://doi.org/10.1016/j.lmot.2005.12.002

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*(2), 230–241. https://doi.org/10.1017/S1138741600006491

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, *37*(1), 1–20.

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, *13*(3), 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*(6), 629–633.

De Houwer, J., Geldof, T., & De Bruycker, E. (2005). The Implicit Association Test as a general measure of similarity. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *59*(4), 228.

De Houwer, J., Hendrickx, H., & Baeyens, F. (1997). Evaluative learning with "subliminally" presented stimuli. *Consciousness and Cognition*, *6*(1), 87–107. https://doi.org/10.1006/ccog.1996.0281

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*(6), 853–869. https://doi.org/10.1037/0033-2909.127.6.853

De Houwer, J., Van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. In *Advances in experimental social psychology* (pp. 127–183). Academic Press.

Dienes, Z., & Seth, A. (2010). Gambling on the Unconscious: A Comparison of Wagering and Confidence Ratings as Measures of Awareness in an Artificial Grammar Task. *Consciousness and Cognition*, *19*(2), 674–681. https://doi.org/10.1016/j.concog.2009.09.009

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Duncker & Humblot.

Eder, A. B., Rothermund, K., & Houwer, J. D. (2013). Affective compatibility between stimuli and response goals: A primer for a new implicit measure of attitudes. *PLOS ONE*, *8*(11), e79210. https://doi.org/10.1371/journal.pone.0079210

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215.

Fazekas, P., & Overgaard, M. (2018). A Multi-Factor Account of Degrees of Awareness. *Cognitive Science*, *42*(6), 1833–1859. https://doi.org/10.1111/cogs.12478

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229–238. https://doi.org/10.1037/0022-3514.50.2.229

Fiedler, K., & Unkelbach, C. (2011). Evaluative Conditioning Depends on Higher Order Encoding Processes. *Cognition & Emotion*, *25*(4), 639–656. https://doi.org/10.1080/02699931.2010.513497

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*(3), 534–540. https://doi.org/10.1080/02699931.2011.588687

Gast, A. (2018). A Declarative Memory Model of Evaluative Conditioning. *Social Psychological Bulletin*, *13(3)*, e28590. https://doi.org/10.5964/spb.v13i3.28590

Gast, A., Gawronski, B., & De Houwer, J. (2012). Evaluative conditioning: Recent developments and future directions, *43*(3), 79–88. https://doi.org/10.1016/j.lmot.2012.06.004

Gast, A., & Rothermund, K. (2010). When old and frail is not the same: Dissociating category and stimulus effects in four implicit attitude measurement methods. *Quarterly Journal of Experimental Psychology*, *63*(3), 479–498.

Gast, A., & Rothermund, K. (2011). What you see is what will change: Evaluative conditioning effects depend on a focus on valence. *Cognition and Emotion*, *25*(1), 89–110. https://doi.org/10.1080/02699931003696380

Gawronski, B., Balas, R., & Creighton, L. (2014). Can the formation of conditioned attitudes be intentionally controlled? *Personality and Social Psychology Bulletin*, *40*(4), 419–432.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2014). The associative-propositional evaluation model: Operating principles and operating conditions of evaluation. In *Dual-process theories of the social mind* (pp. 188–203). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the associative-propositional evaluation model. *Social Psychological Bulletin*, *13*(3), 1–33. https://doi.org/10.5964/spb.v13i3.28024

Gawronski, B., Brannon, S. M., & Luke, D. M. (2021). Unintentional influences in intentional impression formation. In E. Balcetis & G. B. Moskowitz (Eds.), *The handbook of impression formation: A social psychological approach*. New York: Routledge.

Gawronski, B., Mitchell, D. G., & Balas, R. (2015). Is evaluative conditioning really uncontrollable? A comparative test of three emotion-focused strategies to prevent the acquisition of conditioned preferences. *Emotion*, *15*(5), 556.

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*(3), 617–623. https://doi.org/10.1016/j.jesp.2012.01.002

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Greenwald, A. G., & De Houwer, J. (2017). Unconscious conditioning: Demonstration of existence and difference from conscious conditioning. *Journal of Experimental Psychology: General*, *146*(12), 1705–1721. https://doi.org/10.1037/xge0000371

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

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

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. https://doi.org/10.3758/s13428-017-0869-7

Henley, S. H. A. (1984). Unconscious perception re-revisited: A comment on Merikle's (1982) paper. *Bulletin of the Psychonomic Society*, *22*(2), 121–124. https://doi.org/10.3758/BF03333780

Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy*, *40*(3), 217–234.

Heycke, T., Aust, F., & Stahl, C. (2017). Subliminal influence on preferences? A test of evaluative conditioning for brief visual conditioned stimuli using auditory unconditioned stimuli. *Royal Society Open Science*, *4*(9), 160935. https://doi.org/10.1098/

rsos.160935

Heycke, T., & Gawronski, B. (2020). Co-occurrence and relational information in evaluative learning: A multinomial modeling approach. *Journal of Experimental Psychology: General*, *149*(1), 104–124. https://doi.org/10.1037/xge0000620

Heycke, T., Gehrmann, S., Haaf, J. M., & Stahl, C. (2018). Of two minds or one? A registered replication of Rydell et al. (2006). *Cognition and Emotion*, *0*(0), 1–20. https://doi.org/10.1080/02699931.2018.1429389

Heycke, T., & Stahl, C. (2018). No evaluative conditioning effects with briefly presented stimuli. *Psychological Research*, 1–8. https://doi.org/10.1007/s00426-018-1109-1

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2019). A conceptual introduction to Bayesian Model Averaging. *PsyArXiv*. https://doi.org/10.31234/osf.io/wgb64

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

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*(10), 1641–1657. https://doi.org/10.1037/xlm0000533

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional versus dual-process accounts of evaluative conditioning: The effects of co-occurrence and relational information on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *43*(1), 17–32.

Hughes, S., Ye, Y., Van Dessel, P., & De Houwer, J. (2019). When people co-occur with good or bad events: Graded effects of relational qualifiers on evaluative conditioning. *Personality and Social Psychology Bulletin*, *45*(2), 196–208. https://doi.org/10.1177/0146167218781340

Hütter, M., & Sweldens, S. (2018). Dissociating Controllable and Uncontrollable Effects of Affective Stimuli on Attitudes and Consumption. *Journal of Consumer Research*. 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*(3), 539–557. https://doi.org/10.1037/a0026477

Jachs, B., Blanco, M. J., Grantham-Hill, S., & Soto, D. (2015). On the independence of visual awareness and metacognition: A signal detection theoretic analysis. *Journal* *of Experimental Psychology: Human Perception and Performance*, *41*(2), 269–276. https://doi.org/10.1037/xhp0000026

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*(5), 933–948. https://doi.org/10.1037/a0014747

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kattner, F. (2012). Revisiting the Relation between Contingency Awareness and Attention: Evaluative Conditioning Relies on a Contingency Focus. *Cognition & Emotion*, *26*(1), 166–175. https://doi.org/10.1080/02699931.2011.565036

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Konstantinidis, E., & Shanks, D. R. (2014). Don't bet on it! Wagering as a measure of awareness in decision making under uncertainty. *Journal of Experimental Psychology: General*, *143*(6), 2111–2134. https://doi.org/10.1037/a0037977

Kouider, S., & Dupoux, E. (2004). Partial Awareness Creates the "Illusion" of Subliminal Semantic Priming. *Psychological Science*, *15*(2), 75–81. https://doi.org/10.1111/j.0963-7214.2004.01502001.x

Kukken, N., Hütter, M., & Holland, R. W. (2020). Are there two independent evaluative conditioning effects in relational paradigms? Dissociating the effects of CS-US pairings and their meaning. *Cognition and Emotion*, *34*(1), 170–187. https://doi.org/10.1080/02699931.2019.1617112

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition*, *10*(3), 294–340. https://doi.org/10.1006/ccog.2000.0494

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Gainesville, FL: University of Florida.

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*(1), 3–26. https://doi.org/10.1037//0097-7403.28.1.3

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A* *model comparison perspective*. Belmont, CA, US: Wadsworth/Thomson Learning.

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology*, *104*(1), 45–69. https://doi.org/10.1037/a0030734

Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift Für Experimentelle Psychologie*.

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*, 1948550616687093. 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*(02), 183. https://doi.org/10.1017/S0140525X09000855

Moran, T., & Bar-Anan, Y. (2013). The effect of object-valence relations on automatic evaluation. *Cognition & Emotion*, *27*(4), 743–752. https://doi.org/10.1080/02699931.2012.732040

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2016). The assimilative effect of co-occurrence on evaluation above and beyond the effect of relational qualifiers. *Social Cognition*, *34*(5), 435–461. https://doi.org/10.1521/soco.2016.34.5.435

Moran, T., Bar-Anan, Y., & Nosek, B. A. (2017). The effect of the validity of co-occurrence on automatic and deliberate evaluations. *European Journal of Social Psychology*, *47*(6), 708–723.

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4), 435–450. https://doi.org/10.2307/2183914

Nieuwenstein, M., & Kromm, M. (2017). Visual Memory: The Price of Encoding Details. In *58th Annual Meeting of the Psychonomic Society*. Vancouver, Canada.

Odegaard, B., Chang, M. Y., Lau, H., & Cheung, S.-H. (2018). Inflation versus filling-in: Why we feel we see more than we actually do in peripheral vision. *Phil. Trans. R. Soc. B*, *373*(1755), 20170345. https://doi.org/10.1098/rstb.2017.0345

Olson, M. A., & Fazio, R. H. (2001). Implicit Attitude Formation through Classical Conditioning. *Psychological Science*, *12*(5), 413–417. https://doi.org/10.1111/1467-9280.00376

Pavlov, I. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford University Press.

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*(3), 277–293. https://doi.org/10.1037/0022-3514.89.3.277

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience*, *10*(2), 257–261. https://doi.org/10.1038/nn1840

Peters, K. R., & Gawronski, B. (2011). Are We Puppets on a String? Comparing the Impact of Contingency and Validity on Implicit and Explicit Evaluations. *Personality and Social Psychology Bulletin*, *37*(4), 557–569. https://doi.org/10.1177/0146167211400423

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*(1), 130–144. https://doi.org/10.1037/0278-7393.33.1.130

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, *3*, 1–23.

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

Reingold, E. M., & Merikle, P. M. (1990). On the Inter-relatedness of Theory and Measurement in the Study of Unconscious Processes. *Mind & Language*, *5*(1), 9–28. https://doi.org/10.1111/j.1468-0017.1990.tb00150.x

Rescorla, 1. (1968). Probability of shock in the presence and absence of cs in fear conditioning. *Journal of Comparative and Physiological Psychology*, *66*(1), 1.

Rothermund, K., Teige-Mocigemba, S., Gast, A., & Wentura, D. (2009). Minimizing the influence of recoding in the implicit association test: The recoding-free implicit association test (IAT-RF). *Quarterly Journal of Experimental Psychology*, *62*(1), 84–98.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the implicit association test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*(2), 139.

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*(6), 995. 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 ahanging valence-inconsistent implicit and explicit attitudes. *Psychological Science*, *17*(11), 954–958. https://doi.org/10.1111/j.1467-9280.2006.01811.x

Sandberg, K., Bibby, B. M., & Overgaard, M. (2013). Measuring and testing awareness of emotional face expressions. *Consciousness and Cognition*, *22*(3), 806–809. https://doi.org/10.1016/j.concog.2013.04.015

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring Consciousness: Is One Measure Better than the Other? *Consciousness and Cognition*, *19*(4), 1069–1078.

Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *The Quarterly Journal of Experimental Psychology*, *60*(3), 291–309. https://doi.org/10.1080/17470210601000581

Shanks, D. R., & Dickinson, A. (1990). Contingency awareness in evaluative conditioning: A comment on baeyens, eelen, and van den bergh. *Cognition & Emotion*, *4*(1), 19–30. https://doi.org/10.1080/02699939008406761

Shanks, D. R., & St. John, M. F. (1994). Characteristics of Dissociable Human Learning Systems. *Behavioral and Brain Sciences*, *17*(3), 367–395. https://doi.org/10.1017/S0140525X00035032

Stahl, C., & Aust, F. (2018). Evaluative Conditioning as Memory-Based Judgment. *Social Psychological Bulletin*, *13(3)*, 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*(5), 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*(2), 286–291.

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*(3), 404–420. https://doi.org/10.1037/a0016196

Sweldens, S., Corneille, O., & Yzerbyt, V. (2014). The Role of Awareness in Attitude Formation Through Evaluative Conditioning. *Personality and Social Psychology Review*, *18*(2), 187–209. https://doi.org/10.1177/1088868314527832

Sweldens, S., Van Osselaer, S. M., & Janiszewski, C. (2010). Evaluative Conditioning Procedures and the Resilience of Conditioned Brand Attitudes. *Journal of Consumer*

Research, *37*(3), 473–489. https://doi.org/10.1086/653656

Timmermans, B., & Cleeremans, A. (2015). How can we measure awareness? An overview of current methods. In M. Overgaard (Ed.), *Behavioral Methods in Consciousness Research* (pp. 21–46). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199688890.003.0003

Timmermans, B., Sandberg, K., Cleeremans, A., & Overgaard, M. (2010). Partial awareness distinguishes between measuring conscious perception and conscious content: Reply to Dienes and Seth. *Consciousness and Cognition*, *19*(4), 1081–1083. https://doi.org/10.1016/j.concog.2010.05.006

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A Meta-Analysis. *Psychological Bulletin*, *135*(3), 452–477. https://doi.org/10.1037/a0015329

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*, 919–934.

Wentura, D., & Rothermund, K. (2007). Paradigms we live by. A plea for more basic research on the IAT. *Implicit Measures of Attitudes*, 195–215.

Wernicke, M., & Mattler, U. (2019). Masking procedures can influence priming effects besides their effects on conscious perception. *Consciousness and Cognition*, *71*, 92–108. https://doi.org/10.1016/j.concog.2019.03.009

Wierzchoń, M., Paulewicz, B., Asanowicz, D., Timmermans, B., & Cleeremans, A. (2014). Different subjective awareness measures demonstrate the influence of visual identification on perceptual awareness ratings. *Consciousness and Cognition*, *27*, 109–120. https://doi.org/10.1016/j.concog.2014.04.009

Wu, M., & Cheng, P. W. (1999). Why causation need not follow from statistical association: Boundary conditions for the evaluation of generative and preventive causal powers. *Psychological Science*, *10*(2), 92–97. https://doi.org/10.1111/1467-9280.00114

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

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*, 724.

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation*, *43*(3), 155–165.

Zehetleitner, M., & Rausch, M. (2013). Being Confident without Seeing: What Subjective Measures of Visual Consciousness Are About. *Attention, Perception, &*

Psychophysics, *75*(7), 1406–1426. https://doi.org/10.3758/s13414-013-0505-2

Appendices

Appendix A
Pilot studies

Three pilot studies were conducted to identify presentation duration settings, assess subjective awareness scales and their possible interactions with the identification task, and develop a 6-point perceptual awareness scale suitable for the present materials.

Pilot Study 1

The first pilot aimed to identify three different CS presentation durations yielding low, intermediate, and high visibility conditions, respectively. We assessed both objective visibility (i.e., identification accuracy) and subjective visibility (using 4-point scales assessing either confidence or perceptual awareness; this was varied between participants). The procedure adapted CS durations such that the subjective awareness ratings for the *low* condition were at the lowest scale point (i.e., 1 out of 4); for the *intermediate* condition, the second-lowest scale point was targeted; and for the *high* condition, subjective awareness ratings were adjusted to between scale points 3 and 4 (i.e., ratings should reach the highest level, but avoid ceiling effects). For a realistic test, this study realized an EC procedure and repeatedly presented each CS together with USs (IAPS images) of either positive or negative valence.

Participants and design. A sample of $N = 10$ University of Cologne students from different majors completed the experiment in exchange for either a monetary compensation or partial course credit. The study followed a 2 (*Rating Scale*: confidence vs. perceptual awareness) × 3 (*Material*: face vs. product vs. nonword) × 2 (*US Valence*: positive vs. negative) × 3 (*CS Visibility*: low, intermediate, high) mixed design with the first factor varying between participants.

Materials. The CS pool comprised 30 pictures of human faces (14 of them showing males), 30 pictures of mundane products (e.g., toothbrush), and 30 pictures of nonwords made up of 6 or 7 vowels and consonants. All pictures were black-and-white and had a size of 200px × 200px. For each participant, 18 faces, 18 products, and 18 nonwords were randomly drawn. For each level of *Material*, the 18 pictures were randomly assigned to the levels of *US Valence* and *CS Visibility Level.* The remaining faces, products, and nonwords were used as additional options in the objective identification task (see *Procedure*). As USs, we used 36 positive and 36 negative IAPS pictures (350px × 263px; Lang, Bradley, & Cuthbert, 2008; for a list of IAPS pictures used in this study). As forward and backward masks, we used 12 pictures of black-and-white patterns (240px × 205px).

Procedure. The study was administered on a personal computer with a CRT monitor set at 100 Hz refresh rate and controlled by the OpenSesame software (Mathôt et al., 2012). Participants were seated in a cubicle and were told that the study was about attention and consisted of a series of trials during each of which various stimuli would be presented. They were instructed to attend carefully to all presented stimuli, and to perform two tasks after each trial. Subsequently, participants in both *Rating Scale* conditions were introduced to the objective identification task which required them to select the previously presented CS from a list of 15 options (i.e., 5 from each of the three material sets).

Next, participants were introduced to the second task. Participants in the "confidence" group were asked to rate their subjective confidence of having chosen the correct CS in the preceding identification task on a 4-point scale. The lowest scale point was labeled "completely unsure", and the highest scale point was labeled "completely sure". The two intermediate scale points did not have labels; participants were instructed to use them to reflect their gradual increase in confidence. Participants in the "perceptual awareness" group were asked to rate their subjective clarity of perception during CS presentation. The lowest scale point was labeled "I did not recognize anything", and the highest scale point was labeled "I recognized the picture clearly". Once again, the two intermediate scale points were not labelled, and participants were instructed to use them to reflect their gradual increase in clarity of perception.

After the instructions, 10 blocks of 54 trials were adminstered (with short breaks between blocks). A trial proceeded as follows: First, a blank screen was presented for 500ms. Then a forward mask was presented centrally for 500ms. Next, the CS-US pair was presented, with the US shown in the upper half of the screen and the CS shown centrally, replacing the forward mask and slightly overlapping with the much bigger US. The CS-US pair was presented for the assigned CS duration, after which the CS was replaced by the backward mask. Postmask and US remained in place for another (1000 - *[CS duration]*) ms, after which the screen was cleared. Hence, US duration and total trial duration were 1000ms and 2000ms, respectively. US valence as well as CS duration varied on a trial-by-trial basis, and trial order was determined randomly for each participant anew. Each CS appeared once in each block. Following each CS-US presentation, participants of both *Rating Scale* groups were required to perform the identification task. For every combination of *US Valence* and *CS Visibility Level*, there was a different list of response options. Each list consisted of all 9 faces, products, and nonwords pertaining to the given factor combination (3 per material) as well as 6 additional options (2 per material). Finally, participants rated their confidence or perceptual awareness on the 4-point scale.

In order to identify presentation conditions for the three aspired visibility levels, we created three CS categories which differed by their initial presentation durations and the rules that governed the intra-individual adjustment of these presentation durations across the experiment. In the first block, *low*-visibility CSs were presented for 30ms, *intermediate* CSs were presented for 40ms, and *high* CSs were presented for 100ms. In the following blocks, presentation durations were adjusted depending on participants' confidence/perceptual awareness ratings, and for the three materials separately. For *low* CSs, we aimed for responses at the lowest scale point (almost) exclusively. Hence, presentation duration for a given material was increased (decreased) by 10ms if the participant had (had not) responded with the lowest scale point in all *low* CS trials of the respective material in the preceding block. For *intermediate* CSs, we aimed at responses at the second scale point. Hence, presentation duration for a given material was increased (decreased) by 10ms if the participant had used the lowest scale point more often (less often) than he/she had used the third or fourth scale point. Whenever the lowest option was used equally often as the third or fourth scale point, the presentation duration was held constant. For *high* CSs, we aimed at respones at the highest scale point in 50 to 66.6% of trials. Hence, presentation duration for a given material was increased by 10ms if the participant had used scale points 1,2, or 3 more often than the highest scale point. If a participant had responded with the highest scale point in at least 83.3% of trials, the presentation duration was decreased by 10ms. Presentation duration remained unchanged when the fourth scale point was used in 50 to 66.6% of trials. After the last block, evaluative ratings were collected for all 90 CSs. Participants rated the CSs on a 20-point rating scale with endpoints labeled as *unpleasant* and *pleasant*. Finally, participants were thanked, debriefed and compensated.

Results. First, we analyzed the CS presentation durations resulting from the adaptive procedure. A 2 (*Rating Scale*: confidence vs. perceptual awareness) $\times$ 3 (*Material*: face vs. product vs. nonword) $\times$ 3 (*CS Visibility*: low, intermediate, high) mixed-design ANOVA revealed significant main effects of *CS Visibility*, $F(1.28, 10.20) = 70.76$, $MSE = 545.75$, $p < .001$, $\hat{\eta}^2_G = .834$, and of *Material*, $F(1.85, 14.79) = 9.42$, $MSE = 27.00$, $p = .003$, $\hat{\eta}^2_G = .046$, as well as a significant three-way interaction of *Rating Scale*, *Material*, and *CS Visibility*, $F(3.01, 24.06) = 3.45$, $MSE = 30.52$, $p = .032$, $\hat{\eta}^2_G = .031$. This three-way interaction was irrelevant for the purpose of the present study and is therefore not discussed further. All other main effects and interactions did not reach significance, all ps $\geq .073$. The main effect of *Material* reflected lower CS presentation durations for faces than for nonwords, $F(1, 9) = 10.87$, $MSE = 12.27$, $p = .009$, $\hat{\eta}^2_G = .130$, as well as for products, $F(1, 9) = 7.39$, $MSE = 8.29$, $p = .024$, $\hat{\eta}^2_G = .057$. CS presentation durations did not differ between nonwords and products, $F(1, 9) = 1.32$, $MSE = 10.52$, $p = .280$, $\hat{\eta}^2_G = .014$. The main effect of *CS Visibility* reflected higher CS presentation durations for higher levels of *CS Visibility*. For *low* CSs, resulting presentation durations ranged between 10 and 60 ms with a mean of 24.7 ms. For *intermediate* CSs, presentation durations ranged between 10 and 80 ms with a mean of 45.3 ms. For *high* CSs, presentation durations ranged between 40 and 150 ms with a mean of 81.5 ms.

For the nonword materials that were selected for the main studies, results indicated CS durations of 30-40 ms for the *low* category; of 50-60 ms for the *intermediate* condition; and of 70-80 ms for the *high* condition. These ranges are consistent with the subjective rating results obtained across *CS Visibility* categories: The lowest category on the 4-point subjective rating scales (i.e., 1 out of 4) was dominant for nonwords presented for up to 40 ms (i.e., the *low* category); subjective ratings of 2 were most common for nonwords presented for 50 and 60 ms (i.e., the *intermediate* condition); and subjective ratings between

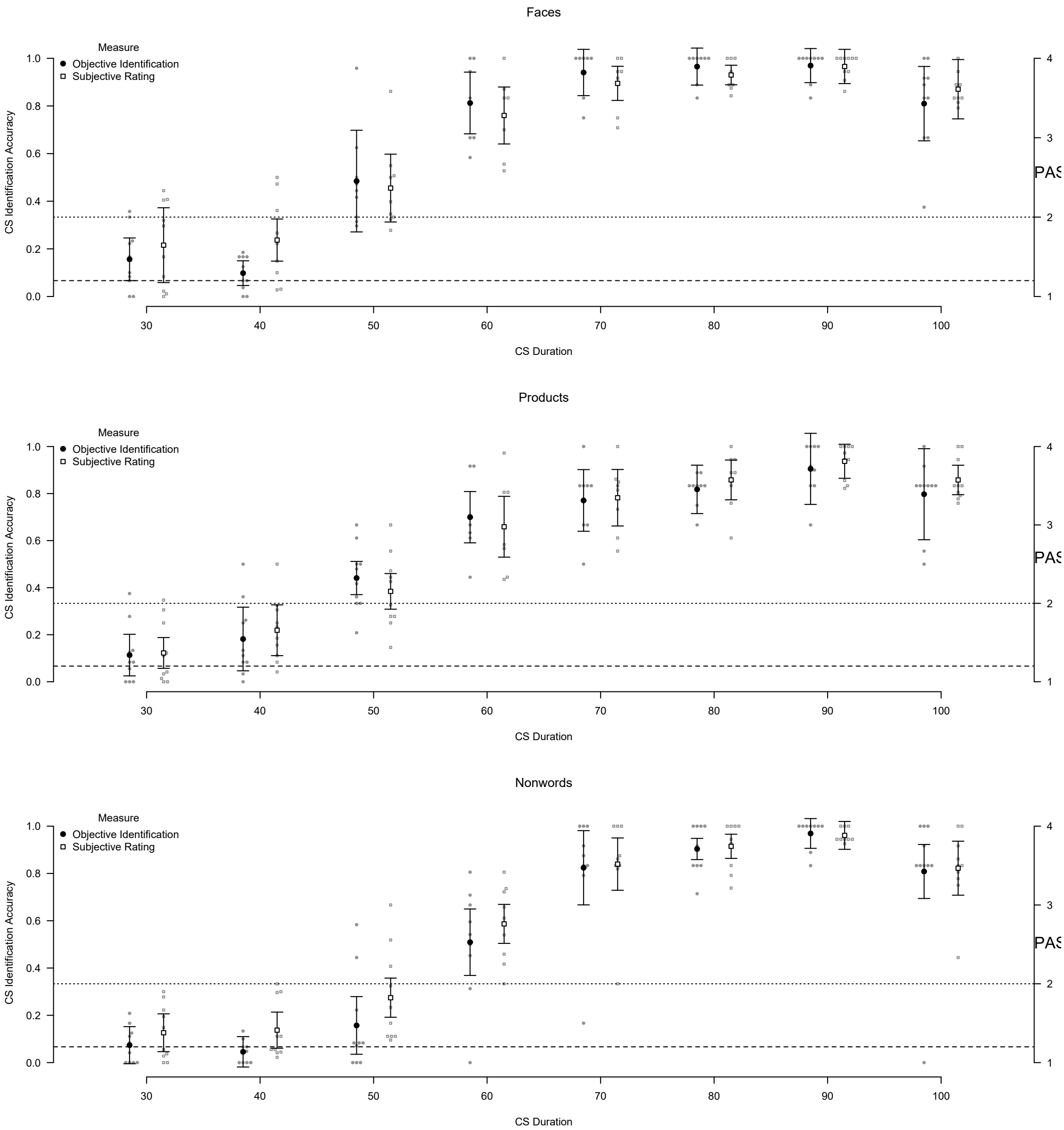

Figure A1. Pilot 1: Objective identification performance and subjective ratings (and 95% CI) as a function of CS Duration and Material. The dashed line represents the overall chance level (1/15). The dotted line represents the chance level per level of Material (1/3).

3 and 4 were obtained for nonwords presented for 70 ms or longer (i.e., the *high* condition). Objective identification performance was significantly above chance for nonwords presented for 50 ms or higher, all ps $< .001$, but did not differ from chance for nonwords presented for less than 50 ms, all ps $\geq .432$. Objective identification performance and subjective ratings are illustrated as a function of presentation duration in Figure A1.

Pilot Study 2

The second pilot study again used both confidence and perceptual awareness ratings and aimed to identify potential interference effects between tasks. All participants performed both the objective and subjective tasks on some of the CSs at some point during the procedure; however, the tasks were combined in different ways, and we tested whether the combination and/or order of tasks affected performance.

On a given trial, some participants performed both the subjective and objective tasks (high load); others were administered only one of the two tasks on a given trial (low load), but the type of task changed halfway through the experiment (i.e., after 5 blocks of trials). Orthogonally, we manipulated task order: Some participants performed the objective task first (*objective-first*), while others started with the subjective task. For high-load participants, the order condition indicates which of the two tasks was administered first within each trial; for low-load participants, it refers to the task administered during the first half of the study.

Thus, we investigated the effect of load (two vs. one task per trial, i.e., high vs. low load) and order (objective-first vs. subjective-first) on identification performance and subjective ratings by comparing four conditions: In the first two, both objective and subjective tasks were performed on each trial. The third group first performed the objective task for five blocks, then switched to the subjective task; task order was switched for the fourth group (subjective first, then objective).

Participants and design. A sample of $N = 49$ University of Cologne students from different majors completed the experiment in exchange for either a monetary compensation or partial course credit. Due to technical problems, two participants could not finish the experiment; their data were excluded from the analysis (i.e., final $N = 47$). The study followed a 2 (*Rating Scale*: confidence vs. perceptual awareness) $\times$ 2 (*Load*: one task vs. two tasks) $\times$ 2 (*Order*: objective vs. subjective task first) $\times$ 3 (*Material*: face vs. product vs. nonword) $\times$ 2 (*US Valence*: positive vs. negative) $\times$ 3 (*CS Visibility*: low, intermediate, high) mixed design with the first three factors varying between participants.

Materials and Procedure. We used the same materials as in the first pilot study. The procedure was by and large the same as in the previous study. Replicating the previous procedure, $N = 11$ participants (confidence: 4, perceptual awareness: 7) performed the identification task and the subjective rating after each CS-US presentation, with the former task always preceding the latter (*high-load/objective-first* group). Another $N =$ 12 participants (confidence: 6, perceptual awareness: 6) also performed both tasks after each trial, but the subjective rating was performed first (*high-load/subjective-first* group).[27] Another $N = 12$ participants (confidence: 5, perceptual awareness: 7) performed the identification task in the first five blocks, and the subjective rating task in blocks 6 to 10 (*low-load/objective-first* group). Finally, $N = 12$ participants (confidence: 6, perceptual awareness: 6) performed the subjective rating task in the first blocks and the identification task in blocks 6 to 10 (*load-load/subjective-first* group). CS durations differed as a function of *CS Visibility* and *Block*: For all levels of *Material*, *low* (*intermediate*) CSs were presented for 30ms (60ms) in blocks 1,2,6, and 7, for 40ms (50ms) in blocks 3,4,8, and 9, and for 50ms (40ms) in blocks 5 and 10; *high* CSs were presented for 100ms in blocks 1 and 6, for 90ms in blocks 2 and 7, for 80ms in blocks 3 and 8, for 70ms in blocks 4 and 9, and for 60ms in blocks 5 and 10.

Results. Objective and subjective performance are illustrated in Figure A2. Objective performance was a function of *CS duration* and *Material* but was unaffected by *Order*, *Load* and *Rating Scale*. Accordingly, a 2 (*Rating Scale*) × 2 (*Load*) × 2 (*Order*) × (*Material*) × 8 (*CS Duration*) mixed-design ANOVA revealed significant main effects of *CS Duration*, $F(2.99, 122.39) = 319.55$, $MSE = 0.11$, $p < .001$, $\hat{\eta}^2_G = .706$, and of *Material*, $F(1.67, 68.35) = 42.23$, $MSE = 0.08$, $p < .001$, $\hat{\eta}^2_G = .113$, as well as a significant interaction between the two factors, $F(5.59, 229.27) = 9.21$, $MSE = 0.04$, $p < .001$, $\hat{\eta}^2_G = .048$. All other main effects and interactions did not reach significance, all ps $\geq .059$. The main effect of CS Duration reflected higher objective identification performances for higher CS Duration levels. The main effect of Material reflected higher identification performances for faces, followed by products and - with a greater gap - nonwords. The interaction of *CS Duration* and *Material* reflected the fact that the simple main effect of *Material* was significant for all levels of *CS Duration*, all ps $\leq .002$, except for 30 ms where objective identification performances did not differ between faces, products and nonwords, $p = .208$.

Similarly, PAS ratings were a function of *CS Duration* and *Material*, and unaffected by *Order* and *Load*. Accordingly, a 2 (*Load*) × 2 (*Order*) × (*Material*) × 8 (*CS Duration*) mixed-design ANOVA revealed significant main effects of *CS Duration*, $F(1.37, 30.04) = 107.85$, $MSE = 0.28$, $p < .001$, $\hat{\eta}^2_G = .641$, and of *Material*, $F(1.65, 36.29) = 38.03$, $MSE = 0.06$, $p < .001$, $\hat{\eta}^2_G = .138$, as well as a significant interaction between the two factors, $F(4.78, 105.07) = 6.85$, $MSE = 0.03$, $p < .001$, $\hat{\eta}^2_G = .037$. All other main effects and interactions did not reach significance, all ps $\geq .140$.

In contrast, in addition to *CS Duration* and *Material*, confidence ratings were also affected by *Order* and *Load*. A 2 (*Load*) × 2 (*Order*) × (*Material*) × 8 (*CS Duration*) mixed-design ANOVA revealed significant main effects of *CS Duration*, $F(2.74, 49.30) = 282.67$, $MSE = 0.06$, $p < .001$, $\hat{\eta}^2_G = .805$, of *Material*, $F(1.73, 31.09) = 38.18$, $MSE = 0.05$, $p < .001$, $\hat{\eta}^2_G = .246$, and of *Order*, $F(1, 18) = 11.12$, $MSE = 0.20$, $p = .004$, $\hat{\eta}^2_G = .175$, as

[27] In this group as well as the *low-load* groups, confidence scale phrasing was slightly changed: Instead of asking how sure they were to have had identified the correct picture, participants were asked how sure they were to be able to identify the correct picture.

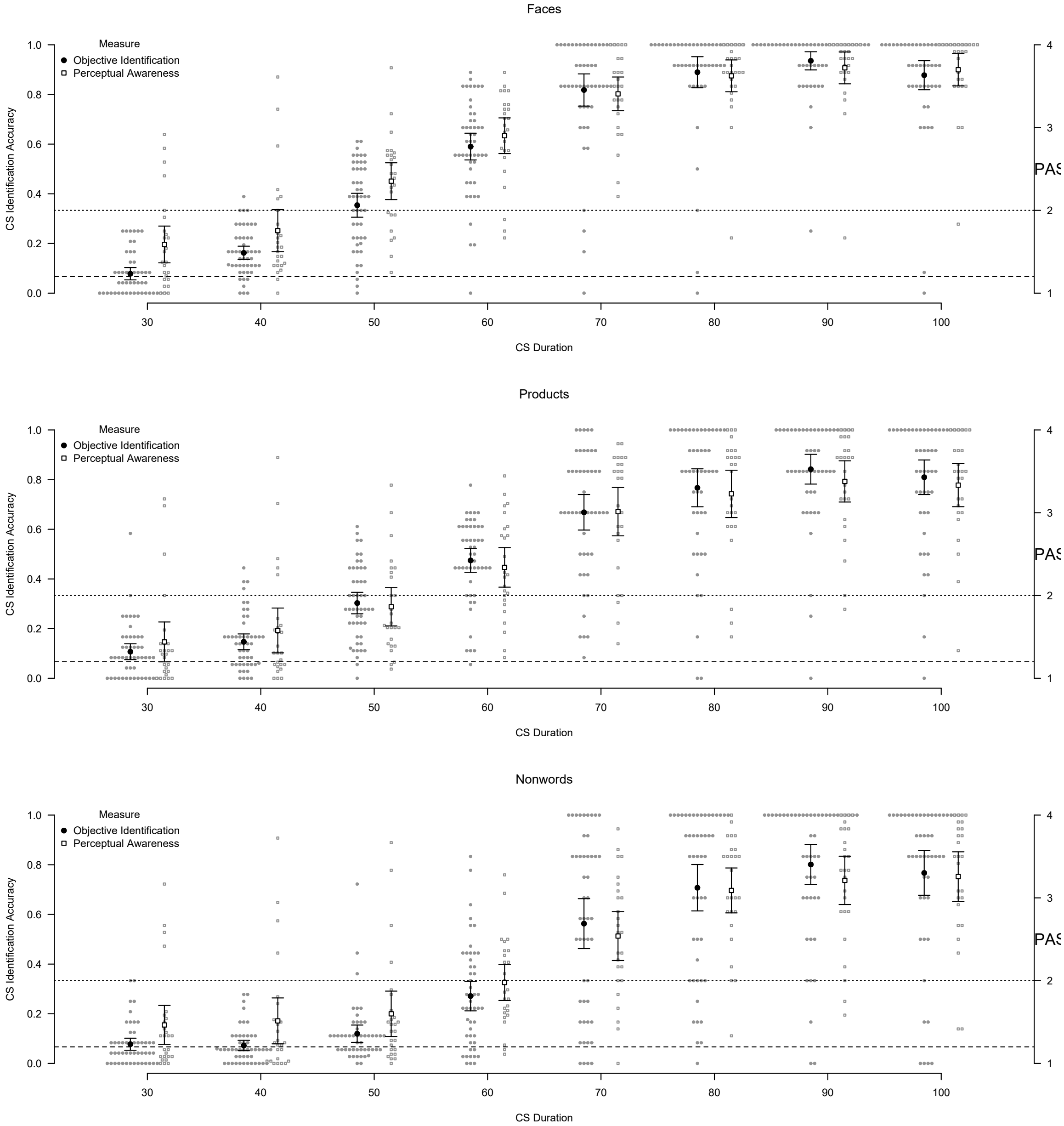

Figure A2. Pilot 2: Identification performance (left axis) and perceptual awareness ratings (right axis) as a function of CS Duration and Material (error bars represent 95% CIs). The dashed line represents the overall chance level (1/15). The dotted line represents the chance level per level of Material (1/3).

well as significant two-way interactions of *CS Duration* and *Material*, $F(4.77, 85.79) = 7.05$, $MSE = 0.03$, $p < .001$, $\hat{\eta}^2_G = .086$, and of *CS Duration* and *Order*, $F(2.74, 49.30) = 6.89$, $MSE = 0.06$, $p = .001$, $\hat{\eta}^2_G = .091$. Furthermore, significant three-way interactions of *Load*, *Order* and *CS Duration*,$F(2.74, 49.30) = 5.05$, $MSE = 0.06$, $p = .005$, $\hat{\eta}^2_G = .068$, and of *Order*, *CS Duration* and *Material*,$F(4.77, 85.79) = 2.69$, $MSE = 0.03$, $p = .028$, $\hat{\eta}^2_G = .035$, were found. All other main effects and interactions did not reach significance, all *p*s $\geq .099$.

Crucially, the main effect of *Order* indicated that confidence was higher when the objective task was performed first, irrespective of *Load*.[28] This effect may be an artefact of the additional CS presentation during the identification task (i.e., this presentation could further consolidate the percept). Alternatively, it may also reflect an influence of performing the identification task on confidence (or vice versa). In any case, it suggests that confidence ratings were affected by (or interfered with) the identification measure, rendering them unsuitable for the present purposes. Notably, there were no effects of order or load on PAS ratings, suggesting that collecting PAS ratings in addition to the idenficiation task on each trial does not interfere with either task.

Finally, the results of the first two pilot studies showed that a 4-point perceptual awareness scale was not fine-grained enough for our purposes of distinguishing between different (low, intermediate, and high) visibility conditions in terms of perceptual experience: Given that the *low* condition did not exceed chance level accuracy but showed subjective ratings between 1 and 2 (i.e., could not be restricted to a mean close to 1 even with CS durations as brief as 10 ms), and given that a rating of 3 was associated with the *high* condition (i.e., because it allowed an almost clear perception—and hence, conscious processing—of the CS), both of these categories could not be clearly distinguished from the *intermediate* condition. Thus, the 4-point scale would not allow us to distinguish between the conditions in terms of perceptual awareness (i.e., participants' experience in both the *low* and *intermediate* condition would be described as essentially identical, or at least largely overlapping, as having seen a "brief glimpse"). This may be largely a consequence of our use of more complex materials (e.g., whereas the original scale was developed for singular simple geometric figures, nonwords consist of multiple complex figures), which may allow for a wider range of perceptual experiences. To capture more fine-grained distinctions, we developed a 6-point PAS scale for the present materials in a third pilot study.

Pilot Study 3

The third pilot developed and tested a 6-point perceptual awareness scale adapted for the present materials (faces, nonwords, products). The scale was developed by presenting several student assistants with the materials at different durations, and having them describe their perceptual experience. Their responses were grouped and abstracted, which yielded

[28] A 2 (*Order*) × 2 (*Load*) between-subjects ANOVA on the confidence ratings yielded a main effect of *Order*, $F(1, 19) = 9.25$, $MSE = 0.09$, $p = .007$, $\hat{\eta}^2_G = .328$. All other main effects and interactions did not reach significance, all $p \geq .418$.

the 6 scale points (described below in the *Procedure* section). The resulting scale was then tested on a larger sample.

To calibrate subjective ratings, we also included a few filler trials in which an empty grey rectangle (pre- and post-masked) was presented (for 30, 40, or 50 ms). For a realistic test, this study also realized an EC procedure and repeatedly presented each CS together with USs (IAPS images) of either positive or negative valence.

Participants and design. A sample of $N = 76$ University of Cologne students from different majors completed the experiment in exchange for either a monetary compensation or partial course credit. Due to technical problems, one participant could not finish the experiment. Their data were excluded from the analysis (final $N = 75$). The study followed a 2 (*Visibility Check*: none vs. subjective) × 3 (*Material*: face vs. product vs. nonword) × 2 (*US Valence*: positive vs. negative) × 3 (*CS Duration*: 30ms vs. 40ms vs. 50ms vs. 60ms vs. 70ms vs. 80ms vs. 90ms vs. 100ms) mixed design with the first factor varying between participants.

Materials and Procedure. We used the same materials as in the first two pilot studies. Participants were informed that the study consisted of a series of trials during each of which a stimulus pair would be presented. They were instructed to attend carefully to all presented stimuli. Subsequently, participants in both *Load* conditions were introduced to the valence focus task which required them to indicate their impression of the stimulus pair by clicking on one of two buttons ("pleasant" vs. "unpleasant"). Afterwards, participants in the "one task" group started the learning procedure, whereas participants in the "two tasks" group were familiarized with the 6-point scale of subjective awareness. To this end, they were first shown all six scale points. Secondly, each scale point was presented separately, together with instructions on the circumstances under which to use the given scale point. Participants were told to use the first scale point ("I could not perceive anything.") when they had perceived only the jitter of the pattern masks, and did not feel like they had seen something behind those masks. The second scale point ("I had a vague impression that I am not sure of") was supposed to be used when participants spontaneously felt that had seen something behind the pattern masks, when they could vaguely describe how this impression of having seen something was brought was about (e.g. "The image briefly appeared to be darker."), when they were unsure of their impression, and could not describe the image they had seen. Participants were told to use the third scale point ("I had a vague impression that I am relatively sure of.") when they could describe at least one feature of the briefly presented image (e.g., the shape), and felt sure or relatively sure about the accuracy of their description, but could not name the specific entity that was depicted. The fourth scale point ("I perceived several features, but I am not sure about my perception.") was supposed to be used when participants were able to describe several features of the presented image, and could name the entity that was depicted, but were not entirely sure of their description, and could not describe the depicted entity in its entirety. Participants were instructed to use the fifth scale point ("I perceived several features, and am sure about my perception.") when they were able to describe several features of the image, could name the entity depicted in the image, and felt sure or relatively sure of their description, but could not describe the depicted entity in its entirety. The sixth, and highest scale point was supposed to be used when the participants had perceived the entire image, and were sure of their perception. After having read these instruction, participants practiced the use of this scale in two training blocks that had the same trial structure as the upcoming learning procedure, but used different CSs. After each trial of the training blocks, participants described their perception of the briefly presented stimulus to a student assistent who then helped to select the appropriate scale point reflecting this perception. After the training, the student assistant left the cubicle, and participants in the "two tasks" group started the actual learning procedure. In total, 8 blocks of 54 trials were adminstered (with short breaks between blocks). A trial proceeded as follows: First, a blank screen was presented for 500ms. Then a fixation dot was presented centrally for 500ms. The fixation dot was replaced by a forwards mask presented for 200ms. Next, the CS-US pair was presented, with the US shown in the upper half of the screen and the CS shown centrally, replacing the forward mask and slightly overlapping with the much bigger US. The CS-US pair was presented for the assigned CS duration, after which the CS was replaced by the backward mask. The backward mask disappeared after 50ms, and the US remained on screen for another (1450 - *[CS duration]*) ms, after which the screen was cleared. Hence, US duration and total trial duration were 1500ms and 2200ms, respectively. US valence as well as CS duration varied on a trial-by-trial basis, and trial order was determined randomly for each participant anew. Each CS appeared once in each block. Following each CS-US presentation, participants of both *Load* groups were required to perform the valence focus task. Afterwards, participants in the "one task" group rated their subjective awareness on the 6-point scale. The rating task was followed by the beginning of the next trial. After the last block, evaluative ratings were collected for all 48 CSs that had been presented during the learning phase. Participants rated the CSs on a 20-point rating scale with endpoints labeled as *unpleasant* and *pleasant*. Finally, participants were thanked, debriefed and compensated.

Results. Figure A3 shows the perceptual awareness ratings as a function of material and CS duration. First, the PAS scale worked well: Participants had no problems to understand the instructions and scale usage; they made use of all of the 6 levels (with the exception that 2 participants never used PAS level 1), suggesting that the scale was used in a gradual (vs. binary) manner as intended; and PAS ratings monotonically increased as a function of CS duration.[29] The PAS ratings also covered a wider range: Compared to the restricted range (between 1.5 and 3.5) of the original four-point scale, means on the more fine-grained 6-point subjective perceptual awareness scale ranged from 1.5 to above 5.

An 8 (*CS Duration*) $\times$ 3 (*Material*) within-subjects ANOVA on the subjective ratings

[29] Across all 34 subjects in that condition and all 7 level transitions there were only 13 cases (i.e., 5.5%) in which a higher CS duration level yielded a descriptively lower mean PAS rating for the nonword material we selected for the main studies; most of these occurred in the 30-50 ms range of chance-level identification accuracy.

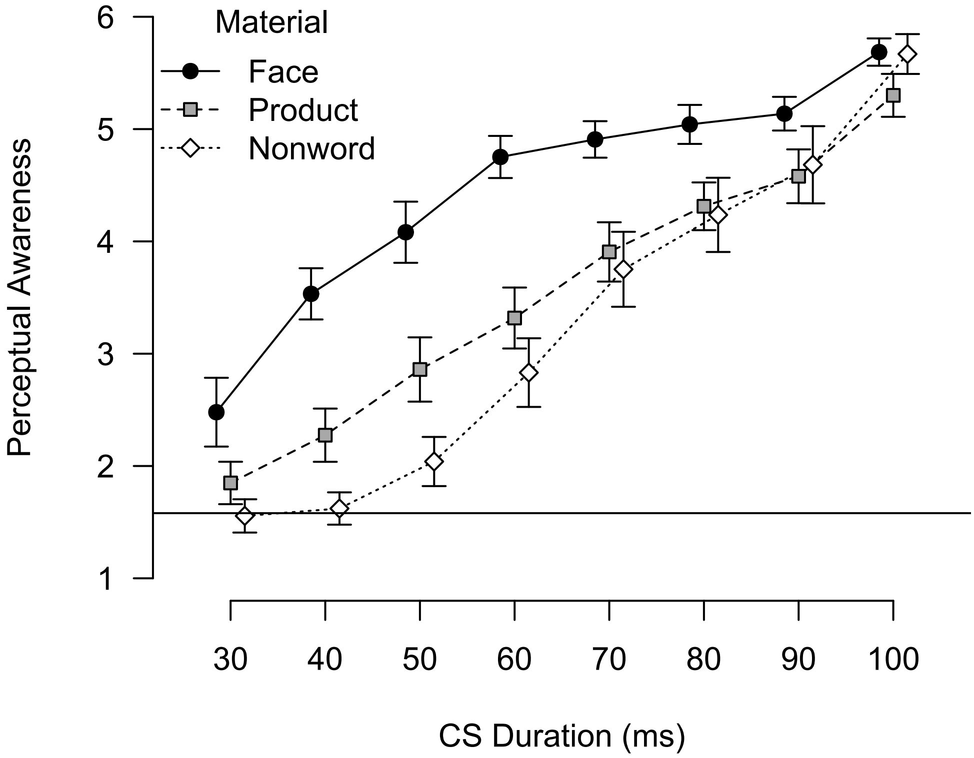

Figure A3. Pilot 3: Perceptual awareness ratings (with 95% CIs) as a function of CS Duration and Material (horizontal line indicates the mean PAS rating for empty filler trials).

revealed significant main effects of *CS Duration* (i.e., higher subjective awareness ratings for higher CS durations), $F(4.00, 139.88) = 455.59$, $MSE = 0.63$, $p < .001$, $\hat{\eta}^2_G = .733$, and of *Material*, $F(1.97, 68.83) = 150.70$, $MSE = 0.72$, $p < .001$, $\hat{\eta}^2_G = .338$, as well as a significant interaction of the two factors, $F(5.38, 188.27) = 19.19$, $MSE = 0.70$, $p < .001$, $\hat{\eta}^2_G = .148$. The main effect of *Material* reflected higher subjective awareness ratings for faces, followed by products, with nonwords receiving the lowest ratings on average. The interaction reflected the fact that the size of the simple main effect of *Material* differed across levels of *CS Duration* (see also Figure A3).

Mean subjective awareness for empty catch trials was $M_{empty} = 1.58$ (i.e., participants perceived either nothing or a *flicker*). This finding suggests that participants cannot reliably distinguish between the presence versus absence of a CS stimulus even at a PAS level of 2; it is consistent with an interpretation of PAS level 2 as reflecting lack of perceptual awareness. For nonwords presented for 30 or 40 ms, subjective ratings did not differ from the mean of subjective ratings on empty catch trials, one-tailed ps $\geq .071$, indicating that participants were subjectively unaware of the nonwords for these presentation durations. For nonwords presented for higher durations as well as for faces and products of levels of *CS Duration*, subjective ratings were significantly higher than the mean of the subjective ratings, all ps $\leq .001$. The ability to reliably separate face, product, and nonword stimuli at individual CS duration levels suggests that the scale was well able to discriminate between low and medium awareness levels.

Finally, we also analyzed evaluative ratings. A 2 (*US Valence*) $\times$ 8 (*CS Duration*) $\times$ 3 (*Material*) $\times$ 2 (*Visibility Check*) on the evaluative ratings revealed significant main effects of *Visibility Check*, $F(1, 72) = 6.84$, $MSE = 134.77$, $p = .011$, $\hat{\eta}^2_G = .020$, of *Material*, $F(1.93, 139.10) = 6.10$, $MSE = 74.55$, $p = .003$, $\hat{\eta}^2_G = .019$, and of *CS Duration*, $F(5.52, 397.33) = 2.34$, $MSE = 12.62$, $p = .036$, $\hat{\eta}^2_G = .004$, but no significant effects involving *US Valence*, all ps $\geq .119$. All other interactions not involving *US Valence* were insignificant as well, all ps $\geq .101$.

The fact that, despite the application of a valence focus, we did not observe EC may be explained by several factors. For one, the strong focus on CS visibility may have distracted from processing the USs and/or their valence. This interpretation is supported by the observation of reduced EC under CS identification instructions (Stahl et al., 2016). Yet, the fact that the lack of EC held for both *Visibility Check* groups seems to suggest that other factors, such as the high number of CSs and the mix of different materials may have played a role as well: In a previous study that used a large number of CSs and mixed different materials (Stahl et al., Exp. 5), EC effects were restricted to the 1000 ms condition but absent for masked CSs presented for 100 ms. Another relevant factor that may help explain the absence of an EC effect is the relatively brief US presentation duration realized in the pilot studies (i.e., 1 second). In previous work we have observed EC effects with similar brief US durations (Stahl et al., Exp. 6), but most of our previous studies presented the USs for longer (i.e., 3-4 seconds; EC effects in these studies were generally larger than in Exp. 6). A smaller number of CSs and longer US presentations were used in the main studies which consistently obtained EC effects.

Appendix B
Additional Results

Experiment 1

Evaluative Ratings. The evaluative ratings of the four families were submitted to a 2 (*US Valence*) $\times$ 2 (*CS-US Relation*) $\times$ 2 (*IAT Type*) mixed-design ANOVA. The between-subjects factor *IAT Type* yielded a marginally significant main effect, $F(1, 62) = 3.63$, $MSE = 3.50$, $p = .061$, $\hat{\eta}^2_G = .015$, but did not enter any significant interaction, all ps $\geq .166$. The main effect reflected higher evaluative ratings in the “valence of action” group ($M = 4.01$, $SD = 2.41$) than in the “valence of sound” group ($M = 3.56$, $SD = 2.48$). Furthermore,

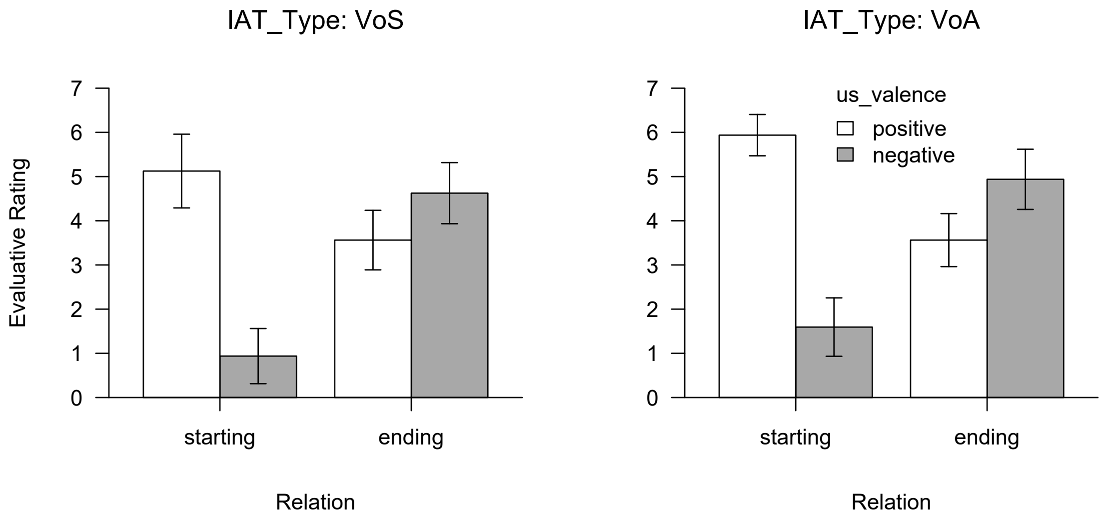

Figure B1. Experiment 1: Mean evaluative ratings as a function of *US Valence*,*CS-US Relation*, and *IAT Type*. Error bars represent 95% within-subjects confidence intervals.

we found a replication of the results by Moran and Bar-Anan (2013), i.e. significant main effects of *US Valence*, $F(1, 62) = 66.44$, $MSE = 2.24$, $p < .001$, $\hat{\eta}^2_G = .150$, and of *CS-US Relation*, $F(1, 62) = 14.08$, $MSE = 2.72$, $p < .001$, $\hat{\eta}^2_G = .043$, as well as a significant interaction of the two factors, $F(1, 62) = 93.79$, $MSE = 5.13$, $p < .001$, $\hat{\eta}^2_G = .364$. This interaction reflected the fact that the direction of the simple main effect of *US Valence* depended on the level of *CS-US Relation.* Among "starting" families, the "melody" family ($M = 5.53$, $SD = 1.77$) was preferred over the "scream" family ($M = 1.27$, $SD = 1.78$), $F(1, 63) = 161.42$, $MSE = 3.61$, $p < .001$, $\hat{\eta}^2_G = .594$. Among "ending" families, the "scream" family ($M = 4.78$, $SD = 1.99$) was preferred over the "melody" family ($M = 3.56$, $SD = 1.84$), $F(1, 63) = 12.99$, $MSE = 3.66$, $p = .001$, $\hat{\eta}^2_G = .093$. Furthermore, the "starting melody" family was preferred over the "ending melody" family, $F(1, 63) = 31.90$, $MSE = 3.89$, $p < .001$, $\hat{\eta}^2_G = .232$, while the "ending scream" family was preferred over the "starting scream" family, $F(1, 63) = 100.47$, $MSE = 3.94$, $p < .001$, $\hat{\eta}^2_G = .468$ (see Figure B1).

Overall IAT effects.

"VoS"-group.

For participants in the "VoS" group, we computed D_o scores in such a way that positive values reflected faster stimulus classification when the "melody" family shared a key with positive attributes, and the "scream" family shared a key with negative attributes (i.e., a preference for the "melody" family over the "scream" family).

The mean D_o score for the "starting" IAT was significantly greater than zero, $t(31) = 4.10$, $p < .001$ (one-tailed), indicating a preference for the "starting melody" family over the "starting scream" family ($M = 0.19$, $SD = 0.27$). The same was true for the "ending" IAT: the mean D_o score was also significantly greater than zero, $t(31) = 1.98$, $p = .028$ (one-tailed), indicating a preference for the "ending melody" family over the "ending scream" family ($M = 0.15$, $SD = 0.43$). A paired-samples t-test found no difference between the mean D_o scores of the two IATs, $M_d = 0.04$, 95% CI $[-0.15, 0.24]$, $t(31) = 0.45$, $p = .658$.

"VoA"-group.

For participants in the "VoA" group, we computed D_o scores in such a way that positive values reflected faster stimulus classification when the family performing the *positive* action ("starting melody" family in the "melody" IAT, "ending scream" family in the "scream" IAT) shared a key with positive attributes, and the family performing the *negative* action ("ending melody" family in the "melody" IAT, "starting scream" family in the "scream IAT") shared a key with negative attributes (i.e., a preference for the "positive action" family over the "negative action" family). The mean D_o score for the "melody" IAT as well as the mean D_o score for the "scream" IAT were significantly greater than zero. The positive mean D_o score indicates a preference for the "starting melody" family over the "ending melody" family in the "melody" IAT ($M = 0.24$, $SD = 0.51$), $t(31) = 2.63$, $p = .007$ (one-tailed), and a preference for the "ending scream" family over the "starting scream" family in the "scream" IAT ($M = 0.16$, $SD = 0.41$), $t(31) = 2.29$, $p = .015$ (one-tailed). A paired-samples t-test found no difference between the mean D_o scores of the two IATs, $M_d = 0.07$, 95% CI $[-0.11, 0.25]$, $t(31) = 0.83$, $p = .411$.

Experiment 2

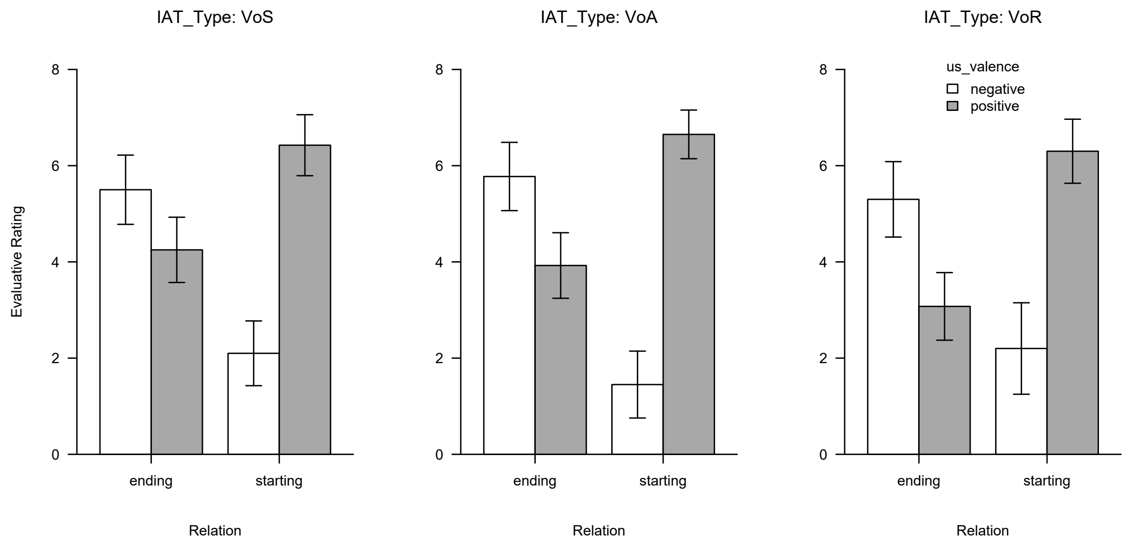

Figure B2. Experiment 2: Mean evaluative ratings as a function of *US Valence* and *CS-US Relation*, and *IAT Type*. Error bars represent 95% within-subjects confidence intervals.

Evaluative Ratings. The evaluative ratings of the four creature families were submitted to a 2 (*US Valence*) $\times$ 2 (*CS-US Relation*) $\times$ 2 (*IAT Type*) mixed-design ANOVA. Once again, significant main effects of *US Valence*, $F(1, 117) = 71.92$, $MSE = 3.19$, $p < .001$, $\hat{\eta}^2_G = .096$, and of *CS-US Relation*, $F(1, 117) = 5.44$, $MSE = 4.47$, $p = .021$, $\hat{\eta}^2_G = .011$, as well as a significant interaction of the two factors, $F(1, 62) = 93.79$, $MSE = 5.13$, $p < .001$, $\hat{\eta}^2_G = .364$, were found. The factor *IAT Type* did not yield a main effect or enter any significant two- or three-way interactions, all ps $\geq .150$. The interaction of *US Valence* and *CS-US Relation* followed the same pattern as in experiment 1. Among "starting" families, the "melody" family ($M = 6.46$, $SD = 1.83$) was preferred over the "scream" family ($M = 1.92$, $SD = 2.32$), $F(1, 119) = 248.40$, $MSE = 4.98$, $p < .001$, $\hat{\eta}^2_G = .543$. Among "ending" families, the "scream" family ($M = 5.53$, $SD = 2.22$) was preferred over the "melody" family ($M = 3.75$, $SD = 2.21$), $F(1, 119) = 36.20$, $MSE = 5.22$, $p < .001$, $\hat{\eta}^2_G = .139$. Furthermore, the "starting melody" family was preferred over the "ending melody" family, $F(1, 119) = 96.20$, $MSE = 4.57$, $p < .001$, $\hat{\eta}^2_G = .309$, while the "ending scream" family was preferred over the "starting scream" family, $F(1, 119) = 112.92$, $MSE = 6.92$, $p < .001$, $\hat{\eta}^2_G = .389$ (see Figure B2).

Overall IAT effects.

"VoS"-group.

For participants in the "VoS" group, we computed D_o scores in such a way that positive values reflected faster stimulus classification when the "melody" families shared a key with positive attributes, and the "scream" families shared a key with negative attributes (i.e., a preference for the "melody" families over the "scream" families).

The mean D_o score in the VoS group was significantly greater than zero, $t(39) = 4.47$, $p < .001$ (one-tailed), indicating a preference for the "melody" families over the "scream" families ($M = 0.38$, $SD = 0.54$).

"VoA"-group.

For participants in the "VoA" group, we computed D_o scores in such a way that positive values reflected faster stimulus classification when the families performing the *positive* actions ("starting melody" and"ending scream") shared a key with positive attributes, and the families performing the *negative* actions ("ending melody" and"starting scream") shared a key with negative attributes (i.e., a preference for the "positive action" families over the "negative action" families).

The mean D_o score in the VoA group was significantly greater than zero, $t(39) = 3.50$, $p = .001$ (one-tailed), indicating a preference for the "positive action" families over the "negative action" families ($M = 0.24$, $SD = 0.44$).

"VoR"-group.

For participants in the "VoR" group, we computed D_o scores in such a way that positive values reflected faster stimulus classification when the "starting" families shared a key with positive attributes, and the "ending" families shared a key with negative attributes (i.e., a preference for the "starting" over the "ending" families).

The mean D_o score in the VoR group was significantly greater than zero, $t(39) = 1.97$, $p = .028$ (one-tailed), indicating a preference for the "starting" families over the "ending" families ($M = 0.13$, $SD = 0.43$).