Contingency Awareness in Evaluative Conditioning: Investigations Using Subliminal Stimulus Presentations

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Tobias Heycke

aus

Bergisch Gladbach

Köln 2018

1. Berichterstatter: Prof. Dr. Christoph Stahl (Köln)
2. Berichterstatter: Prof. Dr. Christian Unkelbach (Köln)
Diese Dissertation wurde von der Humanwissenschaftlichen Fakultät der Universität zu Köln im Juli 2018 angenommen.
Tag der mündlichen Prüfung: 18.07.2018

Erklärung

Chapter 2 beruht auf folgendem Manuskript:

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. doi:10.1080/02699931.2018.1429389

Ich habe Studie 1 zusammen mit SG & CS und Studie 2 gemeinsam mit JH & CS entwickelt. Die Datenerhebung von Studie 1 fand in Köln statt und wurde von SG durchgeführt. Die Erhebung von Studie 2 fand in Columbia, MO, USA statt und wurde von JH überwacht. Ich habe die Daten analysiert und das Manuskript geschrieben, wobei SG, JH und CS zu allen Schritten wertvolle Beiträge geliefert haben.

Chapter 4 beruht auf folgendem (unveröffentlichten) Manuskript:

Heycke, T., Verwijmeren, T., & Stahl, C. (2015). Relevance effects in evaluative conditioning: A pre-registered replication and extension of Verwijmeren, Karremans, Stroebe, & Wigboldus (2012).

Gemeinsam mit CS habe ich die Studie entwickelt, die Daten analysiert, sowie das Manuskript geschrieben. Ich habe die Datenerhebung in Köln überwacht. TV hat bei der Konzeption und Interpretation der Studie, sowie beim Schreiben des Manuskripts wertvolle Beiträge geliefert.

Chapter 6 beruht auf folgendem Manuskript:

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, 160935. doi: 10.1098/rsos.160935

Ich habe die Idee für diese Studien entwickelt und die Datenerhebung aller Studien überwacht. FA hat Beiträge zu den statistischen Analysen geliefert. CS hat zu allen Schritten Beiträge geliefert. Ich habe das Manuskript geschrieben und von Beiträgen von FA und CS profitiert.

Chapter 7 beruht auf folgendem (unveröffentlichten) Manuskript:

Heycke, T. & Stahl, C. (2018). No evaluative conditioning effects with briefly presented stimuli.

Ich habe die Idee für diese Studien entwickelt, die Datenerhebung der Studien überwacht, die statistische Analyse durchgeführt und das Manuskript geschrieben. CS hat zu allen Schritten wertvolle Beiträge geliefert.

Tobias Heycke, Köln, April 2018

Danksagung

Mein besonderer Dank gilt meinem Betreuer Christoph Stahl für die wunderbare Betreuung während der letzten 4 Jahre, dem schönen Arbeitsumfeld und die produktive aber trotzdem freundschaftliche Zusammenarbeit. Mein weiterer Dank gilt meinen Kolleginnen und Kollegen (und Freunden) Roscoe Araujo, Frederik Aust, Karoline Bading, Marius Barth, Julia Haaf, Fabia Högden, Anita Jain, Joachim Radt und Philine Thomasius für viele anregende Gespräche, die gute Zusammenarbeit, Hilfe bei dieser Arbeit (inhaltlich sowie emotional) und schönen Jahren in einem wunderbaren Arbeitsumfeld. Auch meinem Zweitbetreuer Christian Unkelbach möchte ich herzlich für das hilfreiche Feedback danken. Für die Datenerhebung sowie der Hilfe bei vielen kleineren und größeren Aufgaben danke ich herzlich unseren Hilfskräften: Pauline Anton, Nora Becker, Hannuh Frings, Julia Haaf, Laura Henze, Fabia Högden, Julia Krasko, Katharina Mattonet, Lucy Moerschbacher, Philipp Musfeld, Marius Schmitz, Lisa Spitzer und Margret Vössing. Auch ihr habt einen großen Anteil an dem Gelingen dieser Arbeit. Zuletzt möchte ich mich bei meinen Freunden, meinen Eltern, meiner Schwester und dir, Jana, für eure Unterstützung bedanken ohne die diese Arbeit nicht möglich gewesen wäre.

Contents

Summary

Evaluative conditioning (EC) describes a change in preference towards a formerly neutral stimulus (Conditioned Stimulus; CS) after this stimulus is paired with a valent stimulus (Unconditioned Stimulus; US), in the direction of the valence of the US. Evaluative conditioning is proposed as a mechanism of automatic preference acquisition in dual-process theories of attitudes (Gawronski & Bodenhausen, 2006). An automatic route to preference acquisition entails that people do not need to be aware of the CS-US contingency in order for EC effects to develop. To experimentally investigate whether EC effects persist without peoples' awareness, stimuli (e.g., the CS) are presented subliminally (i.e., too briefly to be consciously perceived). When the CS is presented subliminally, contingency awareness between CS and US can be ruled out. Hence, EC effects with subliminal CSs would support theories claiming that contingency awareness is not necessary for EC effects to occur. Studies demonstrating an EC effect with subliminally presented stimuli are reviewed and the stability of the described subliminal EC effect is evaluated. A series of replication studies and additional experiments were conducted to investigate possible boundary conditions for EC effects to occur with subliminally presented stimuli. Specifically, it was tested whether the following features would be beneficial for EC effects with subliminally presented stimuli: (I) goal-relevance during the learning procedure and a relation between CS and US (II) subliminal US presentation with additional counter-attitudinal information (III) choice between CSs after the learning phase as a potentially more sensitive measure of conditioning effects and (IV) a simultaneous presentation of CS and US. Across the experiments, indications for the absence of EC effects with subliminal stimuli were discovered. Additional findings hint to the possibility that subliminal EC effects that were previously found, might have been based on briefly presented—but visible—stimulus presentations. The findings lead to the conclusion that EC does not form automatically. Theoretical implications about the nature of EC are discussed.

Zusammenfassung

Evaluatives Konditionieren (EC) beschreibt die Veränderung in der Evaluation eines ursprünglich neutralen Objekts (CS, engl. Conditioned Stimulus) nachdem das Objekt zusammen mit einem valenten Stimulus (US, engl. Unconditioned Stimulus) gezeigt wurde, in Richtung der Valenz des valenten Stimulus. Evaluatives Konditionieren wird—in 2 Prozess Theorien der Präferenzentwicklung—als ein Mechanismus der automatischen Präferenzentwicklung angesehen (Gawronski & Bodenhausen, 2006). Wenn Präferenzen automatisch entstehen könnten, hätte dies zur Folge, dass Personen kein CS-US Kontingenzbewusstsein besitzen müssen damit evaluative Konditionierungseffekte entstehen können. Um experimentell zu testen ob evaluative Konditionierungseffekte ohne Kontingenzbewusstsein entstehen können kann einer der beiden Stimuli (z.B. der CS) subliminal dargeboten werden. Bei der subliminalen Präsentation wird der CS so kurz dargeboten, dass er nicht bewusst wahrgenommen wird, aber unbewusst verarbeitet werden sollte. Wenn bei solch einer Präsentation ein evaluativer Konditionierungseffekt entsteht, kann angenommen werden, dass evaluatives Konditionieren ohne Kontingenzbewusstsein möglich ist. Dieser Befund würde Theorien der automatischen Präferenzentwicklung unterstützen. In dieser Arbeit wurden Studien die einen subliminalen EC Effekt gefunden haben untersucht und methodische Probleme dieser Arbeiten wurden herausgestellt. Um diese potentiellen Probleme empirisch zu evaluieren, wurden mehrere Studien durchgeführt und folgende, potentiell nutzbringende, Situationen geschaffen um die Möglichkeit eines subliminalen EC Effektes zu erforschen:
(I) Ziel-Relevanz während der Lernphase, sowie eine inhaltliche Beziehung zwischen CS und US (II) subliminale Präsentation des USs mit zusätzlicher, gegenläufiger, expliziter Information während der Lernphase (III) eine Auswahl zwischen CSen nach der Lernphase als potentiell sensiblere Messung des Effektes, sowie (IV) eine gleichzeitige Präsentation des CSs und des USs. Über mehrere Studien hinweg wurden statistische Beweise für die Abwesenheit eines Konditionierungseffektes gefunden, wenn die Stimuli tatsächlich subliminal präsentiert wurden. Ein zusätzlicher Befund der Studien könnte als Indiz angesehen werden, dass vorherige Studien einen Effekt gefunden haben weil Stimuli überzufällig sichtbar waren. Die Ergebnisse der empirischen Studien deuten darauf hin, dass EC Effekte nicht automatische entstehen und die Auswirkung dieser Befunde auf Theorien der Präferenzentwicklung wird diskutiert.

Chapter 1: General Introduction

"*Alle unsere Erkenntnis hebt von den Sinnen an, geht von da zum Verstande, und endigt bei der Vernunft, über welche nichts Höheres in uns angetroffen wird, [...]*"

"*All our knowledge begins with sense, proceeds thence to understanding, and ends with reason, beyond which nothing higher can be discovered in the human mind, [...]*"
(Kant, 1787, p. 228)

Imagine you are in a new country and you want to buy bottled water. You do not know any of the brands, the prices are equal, and they look similar. How do you choose? How could your preferences for a specific water bottle be influenced? Preferences are generally considered an important determinant of behavior (Allport, 1935; Martin & Levey, 1978). Therefore, understanding how preferences are formed could have important implications for consumer science and important life choices (Petty & Wegener, 1998) and might give insights into learning theories in general (De Houwer, Baeyens, & Field, 2005). Of course, it should be noted that preferences are not the only influential factor in determining behavior (Kruglanski et al., 2015). Factors such as the strength of the preference (Krosnick & Petty, 1995) and a desire to act upon the preference (i.e., "wanting") need to be present (Berridge, 2009). Additionally, a behavioral outcome should appear attainable and all factors need to be translated into a goal (for a comprehensive review see Kruglanski et al., 2015). One way in which a preference towards an object, for example, one of the water bottle brands, might be changed is through evaluative conditioning. In the example above, one might have a positive evaluation of one specific brand because we previously saw a commercial poster showing the brand together with cute puppies, while we might have a negative evaluation of another brand because we saw this brand in a dirty trash can before we entered the store. Evaluative conditioning effects describe a change of evaluation of a stimulus (for example a water brand) due to its co-occurrence with a valent stimulus (for example the puppies or a dirty trash can; De Houwer, 2007). Research on evaluative conditioning might give insight into the processes that underlie learning and memory and could help understand phenomena in social psychology and consumer science (De Houwer et al., 2005).

Evaluative conditioning

In an article from 1957, Staats and Staats describe an experimental procedure in which meaning was established by means of classical conditioning. Participants saw neutral nonsense syllables and the experimenter read out valent words, which had to be repeated by the participants. Each syllable was presented 16 times and a different word with the same valence for each syllable was read out by the experimenter. When the nonsense syllables were rated after this learning phase, syllables that were presented together with words of positive valence were evaluated as more positive than syllables presented together with words of negative valence (C. Staats & Staats, 1957). This effect, later coined evaluative conditioning (EC), can be defined as a change in the valence of a neutral stimulus (conditioned stimulus, CS) that results from pairing the stimulus with another valent stimulus (unconditioned stimulus, US; De Houwer, 2007) and can therefore be considered to be a subclass of Pavlovian conditioning (Pavlov, 1927). As in Pavlovian conditioning, stimuli are paired, but in EC the focus is primarily on changes in valence compared to other changes in (physical) responses in Pavlovian conditioning (De Houwer, 2007). The definition of EC by De Houwer (2007) is explicitly agnostic to the processes underlying EC and defining EC as an effect—instead of a process—therefore does not exclude any possible explanations.

The effect of evaluative conditioning was replicated numerous times since the publication of Staats and Staats: Levey and Martin (1975) introduced the term evaluative conditioning with a study pairing postcards of neutral and positive/negative valence. In the past decades, many other studies showed EC effects with images and words (e.g., Baeyens, Eelen, & Van den Bergh, 1990; Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Hammerl & Grabitz, 1993). Other studies extended these findings, showing EC effects with smells (Baeyens, Wrzesniewski, De Houwer, & Eelen, 1996; Todrank, Byrnes, Wrzesniewski, & Rozin, 1995), flavors (Baeyens, Eelen, Van den Bergh, & Crombez, 1990; Dickinson & Brown, 2007) and sounds (Gast, Langer, & Sengewald, 2016; Gorn, 1982; Schemer, Matthes, Wirth, & Textor, 2008). However, until recently researchers were not uniformly convinced that evaluative conditioning is a real phenomenon, as some researchers argued that the EC effect might be due to a stimulus assignment artifact and a lack of control groups (Field & Davey, 1997; Shanks & Dickinson, 1990). Studies addressing these methodological critiques, however, still reported EC effects (Díaz, Ruiz, & Baeyens, 2005; Field & Moore, 2005) which led De Houwer et al. (2005) to state that "this should eliminate all doubts about whether genuine EC effects do exist" (p. 163). Furthermore, in a large scale meta-analysis using 253 independent studies from 145 publications, Hofmann, De Houwer, Perugini, Baeyens, and Crombez (2010) found a mean effect size of $d = 0.524$ and concluded that "beyond any doubt [...] EC is a genuine phenomenon" (Hofmann et al., 2010, p. 410).

Two phenomena were—and still are—central in the discussion whether EC should be considered a distinct form of Pavlovian conditioning: extinction and contingency awareness (Davey, 1994). Extinction is the phenomenon that an EC effect is reduced after the CS is shown alone—without the US—after the initial learning trials. Many studies reported that no extinction could be found in EC (e.g., Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Díaz et al., 2005). However, the debate whether EC is resistant to extinction is ongoing with the meta-analysis by Hofmann et al. (2010) showing that a reduced EC effect could be found after an extinction phase (cf. Gawronski, Gast, & De Houwer, 2014). Other studies support the notion that EC is not resistant to extinction (Aust, Haaf, & Stahl, 2017; Lipp, Oughton, & LeLievre, 2003). The discussion on the absence or presence of extinction effects in EC is therefore still ongoing.

In the previously mentioned research by Staats and Staats (1957) the authors found that the EC effect was still present when excluding participants who reported to be aware of a relationship between the syllables (CSs) and valent words (USs). The authors therefore concluded that the "present results indicate that the meaning of stimuli may be learned without awareness" (C. Staats & Staats, 1957, p. 79). This exact conclusion is still debated 60 years later (e.g., Sweldens, Corneille, & Yzerbyt, 2014). The awareness of the CS-US contingency during encoding, or put differently the "knowledge that a specific CS predicts a specific US", is called contingency awareness in the context of evaluative conditioning (Lovibond & Shanks, 2002, p. 4). Evaluative conditioning is just one of two learning mechanisms where indications for unaware conditioning has been reported (Lovibond & Shanks, 2002). The other learning mechanism is the effect discovered by Perruchet (1985), in which a dissociation between the expectancy of an airpuff and the physical response (i.e., eyeblinking) was shown. The findings by Staats and Staats (1957) and others who reported EC without contingency awareness (e.g., Baeyens et al., 1990, 1992; De Houwer, Baeyens, & Eelen, 1994) supported the view that EC is based on a purely associative, automatic, bottom-up, and low-level process (De Houwer, 2007). However, the meta-analysis by

Hofmann et al. (2010) clearly showed that the most important statistical predictor of EC effects is contingency awareness. While this result does not imply that EC effects cannot be produced by associative processes, it is unlikely that they can only be caused by these processes and not by propositional, higher-order processes. Therefore, ideas and theories proposing that EC is a purely associative and automatic process have not received much attention since this finding (Gast, Gawronski, & De Houwer, 2012).

As preferences are important determinants of behavior, it is vital to understand how they are formed (Hofmann et al., 2010). It is therefore important to understand the mechanisms that underlie EC and theories explaining these should be validated. The central goal of the work presented in the next chapters will be the investigation of theories of EC by addressing the question whether contingency awareness is necessary in EC. In the next paragraph, current models explaining the underlying processes of EC effects are therefore discussed.

Current theories in evaluative conditioning

A scientific theory summarizes single effects found through empirical work and creates something novel that goes beyond single effects and phenomena (Fiedler, 2004) and might, therefore, help to uncover new predictions (De Houwer, 2014b). Specific predictions of boundary conditions of EC effects by scientific theories are especially important as EC might be useful for consumer science and learning theories. Generally, a lack of specific theories might delay the scientific progress in a given field as alternative explanations for an effect can be constructed, leading to possibly unnecessary discussions (Fiedler, 2004). As one will see in the next paragraph, currently discussed theories in evaluative conditioning have different predictions on the circumstances necessary for an EC effect to occur. Considering the possible implications for learning theories, consumer science (De Houwer et al., 2005), and psychological therapies (Lipp & Purkis, 2005) it is of particular interest to test these divergent predictions. Currently two different—and partially opposing—views on possible working mechanisms underlying evaluative conditioning are discussed: the propositional single-process view and the dual-process view.

Propositional single-process models

Even before Hofmann and colleagues (2010) found that contingency awareness is the most important statistical predictor of EC effects, propositional single-process models have been proposed to explain EC effects (Lovibond & Shanks, 2002). Multiple other publications have contributed to describe characteristics of a propositional single-process view[1] which will be discussed in the next paragraph (De Houwer, 2009, 2014b, 2014c, 2014a; Mitchell, De Houwer, & Lovibond, 2009).

Propositions can be understood as mental links that qualify how two events are related (Mitchell et al., 2009) and they can therefore specify the strength and structure of relations in the world (De Houwer, 2009). Mitchell et al. (2009) give the following example when describing a propositional account of the classical Pavlovian experiment where a bell and food were presented to dogs: "There is no mental link between the bell and food, but a proposition of the form: When I hear a bell, I will receive food" (p. 186; *colon added*). From this description, it can be seen that propositions are part of a controlled reasoning process which leads to a belief about the relationship of events (Mitchell et al., 2009). One possible mechanism that could underlie these propositional processes is an episodic memory account in which the CS-US pairs are stored in memory during the learning phase (see Stahl & Heycke, 2016). When asked for an evaluation of the CS a person might use the memory traces of the learning phase (i.e., the CS-US pairs) and an evaluation is constructed on the spot (Schwarz, 2007; Schwarz & Bohner, 2001; Stahl & Heycke, 2016). Of course, other mechanisms, such as a storing of attitudes (e.g., Fazio, Chen, McDonel, & Sherman, 1982), are also plausible under a propositional process assumption. From the definition of propositions three characteristics can be derived: (1) the relationship between events is important, (2) propositions depend on a truth evaluation, and (3) propositions depend on contingency awareness.

(1) Propositions can contain information about the type of relation between events and therefore might qualify a relation between a CS and US in an EC paradigm (De Houwer, 2014c). If, for example, a health product (CS) prevents a negative

[1] As no publication describes a specific theory or model for the propositional single-process view, researchers have suggested referring to the ideas surrounding single propositional working mechanisms in EC as "propositional single-process views". However, as some researchers have argued that the predictions of dual-process models are similarly (un-)specific as those of propositional single-process views (Corneille & Stahl, 2018) I will treat them as equal theories competing in describing the underlying processes in evaluative conditioning (see also Chapter 8).

health outcome (US) it should be evaluated positively according to the propositional account (Hu, Gawronski, & Balas, 2017). Even though the CS is paired with a negative outcome it is evaluated as more positive (for example compared to a CS causing a negative health outcome), which demonstrates that the CS-US pair relies on relational qualifiers.

(2) One additional aspect of propositions is that they are dependent on a truth evaluation (De Houwer, 2009). While the truth evaluation might not be conscious, a proposition will be accepted and acted upon only if it is in line with other propositions (De Houwer, 2014c). We could imagine a situation where a person sees a health product which is liked by this person (for example because it helped this person previously) but the product is shown together with a negative image, which could imply that the product and the negative image are related. It is likely that the possible proposition "the product causes something negative" is evaluated as false due to existing opposing propositions and therefore ignored.

(3) A third central claim by the propositional approach is that evaluative learning always "requires controlled processes" and that "it is not possible to have learned about a relationship between two events in the environment without being, or having been, aware of that relationship" (Mitchell et al., 2009, p. 187). Put differently, only when a participant possesses CS-US contingency awareness an EC effect can be obtained. As awareness of the CS-US relation is necessary for EC effects to arise—according to the propositional single-process view—the process itself can be considered effortful and therefore depends on cognitive resources (De Houwer, 2014c; Mitchell et al., 2009). Under a propositional single-process view, an EC effect can only be found if CS, US, and their co-occurrence are processed consciously. Therefore finding EC effects without knowledge of the relationship between CS and US during the learning phase would contradict the predictions made by the propositional single-process view. This prediction will be investigated in detail in the next chapters. It is important to note that the requirement of awareness discussed above is limited to the acquisition of the proposition and the propositional single-process view does not entail that a person is aware of all steps involved in the acquisition of preferences (i.e., people might be unaware of the process itself or the effect of the process on their behavior; De Houwer, 2009, 2014c).

Lovibond and Shanks (2002) propose two different propositional single-process models: the strong single-process model and the weak single-process model. Contingency awareness also plays a central role in the difference between these postulated models: In the strong single-process model a propositional learning process leads to contingency awareness, which in turn leads to an EC effect. In the weak single-process model the propositional learning process leads to both contingency awareness and an EC effect. Awareness therefore plays no causal role in the weak model, however, it is still an important part in both single-process models proposed by Lovibond and Shanks (2002). It is important to note that in both the strong and the weak model a time delay between the learning phase and a CS-US memory test could result in a decrease in CS-US memory. Finding a difference between a CS evaluation and reported CS-US contingency awareness might therefore be the result of forgetting processes and not evidence for learning in the absence of contingency awareness. A person might, for example, have a positive evaluation for a CS but forgot that the CS was shown with a US of positive valence. In this framework, one additional factor might explain differences between contingency awareness and an evaluation: When asking a participant to evaluate a CS after the learning phase, the task cannot be answered incorrectly and any evaluation is "correct". However, when asking which US (valence) was presented with a given CS, participants are aware that there is a correct answer. Especially, when participants can answer "don't know" in the contingency awareness test, a participant using a conservative decision criterion might reply that he is unaware of the CS-US contingency but might still use the information in an evaluation.

Summarizing the propositional single-process view, a proposition is formed provided that the CS-US contingency is consciously perceived, that sufficient cognitive resources are available, and a proposition is used provided that it is evaluated as truthful.

Dual-process models

As discussed previously, the meta-analysis by Hoffman and colleagues (2010) showed that contingency awareness of the CS-US pairs was the most important predictor, which strongly indicates that EC is based (at least partially) on propositional processes. However, as will be discussed in the following segments, a second process—next to the propositional process described above—might be necessary to explain all instances of evaluative conditioning effects (e.g., Gawronski, Balas, & Creighton, 2014; Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012; Sweldens, Van Osselaer, & Janiszewski, 2010).

A variety of dual-process models have been introduced in the past decades explaining processes underlying preferences (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006). While all models differ, they generally have in common that they consist of explicit (controlled) and implicit (automatic) processes. One dual-process model, which could be considered the most prominent representative of dual-process models explaining preferences, is the associative-propositional evaluation model, or APE model (Gawronski & Bodenhausen, 2006, 2011, 2014a). In the next paragraph, I will outline the APE model, its assumptions, and its predictions that can be derived from it.

As mentioned above, and as can be deduced from the name, the associative-propositional evaluation model proposes two distinct learning processes: propositional and associative processes. The propositional process described in the APE model is highly similar to the propositional process described in the single-process view (Mitchell et al., 2009). The second process, the associative process, can be defined in the context of evaluative conditioning in that "observed co-occurrences between objects and events result in a co-activation of their corresponding mental concepts, which in turn creates an associative link between the two" (Gawronski & Bodenhausen, 2014b, p. 191). The associations, therefore, are understood as mental links that grow stronger when activated (e.g., by a repeated visual presentation of a CS-US pair; Gawronski, Brannon, & Bodenhausen, 2017). One important characteristic of the APE model is that it proposes that associations are generally captured by implicit measures, while propositional processes are generally captured by explicit measures (Gawronski & Bodenhausen, 2006, 2014a). The associative reaction is also described as a spontaneous gut reaction, based on the valence of a given association that is activated automatically when a relevant stimulus is encountered (Gawronski & Bodenhausen, 2006, 2014b).

The APE model differs from other dual-process models (e.g., Rydell, McConnell, & Mackie, 2008) as it allows for a causal influence of associative on propositional processes and vice versa. It is therefore possible that an EC effect is acquired through an associative process which subsequently has a direct influence on the propositional process. The learned EC effect might therefore be found in implicit as well as explicit measures (Case 1, Gawronski & Bodenhausen, 2006). All other interactions between the processes are also possible according to the APE model (for a detailed overview see case 1 to case 8; Gawronski & Bodenhausen, 2006). The interaction between the two processes in the APE model can therefore have the effect that implicit and explicit measures can be influenced by both associative or propositional processes respectively (Gawronski & Bodenhausen, 2006).

From the description of associative processes given above, the following characteristics of EC in the APE model are deduced by Gawronski and Bodenhausen (2006): As associations are defined as automatic activations when a CS-US pair is presented, a repeated pairing of CS and US should strengthen the link between them (Gawronski & Bodenhausen, 2014b). Additionally, contrary to propositions, associations are independent of truth evaluations and associative reactions therefore do not depend on previous beliefs or previous propositions (Gawronski & Bodenhausen, 2006). Additionally, associative processes “may indeed have some features of automaticity” (Gawronski & Bodenhausen, 2007, p. 969), which is of particular interest as the concept of automaticity entails a series of characteristics. Four characteristics of automaticity are described by Bargh (1994) and related to EC can be summarized as: “efficient”, “goal-independent”, “uncontrollable”, “absence of CS-US awareness” (for a detailed overview on automaticity in EC see Corneille & Stahl, 2018). Gawronski and Bodenhausen (2014b) argue that there is no one to one mapping of the proposed associative process in the APE model and the concept of automaticity and argues that both the associative and the propositional process have controlled and automatic aspects. However, the APE model specifies that the associative process can be efficient and resource independent, goal-independent, and associations in memory can indeed occur unintentionally (Gawronski & Bodenhausen, 2014b). Additionally, the APE model clearly states that EC effects can be formed independently of CS-US awareness (Gawronski & Bodenhausen, 2014b, Table 13.1) and states that “associative learning is commonly assumed to be independent of people’s awareness of the relevant contiguities that are responsible for the formation of new associative links. The APE model generally agrees with this contention.” (Gawronski & Bodenhausen, 2014b, p. 194). It is also explicitly stated that “associative learning may occur in the absence of conscious awareness to the extent that the relevant CS-US pairings are in the focus of attention” (Gawronski & Bodenhausen, 2014b, p. 200).

The APE model therefore predicts that an EC effect is possible when participants are not aware of the CS-US contingency during the learning phase. This prediction is in direct contrast to the previously discussed propositional single-process view which predicts that no EC effect is possible without contingency awareness (Mitchell et al., 2009).

Contingency awareness in EC

While multiple phenomena might be important to differentiate between the two views on EC (e.g., relational information as a qualifier of CS-US information), this work will focus on one specific phenomenon: contingency awareness of the CS-US pair. As discussed earlier, contingency awareness has been an important part of research on EC since the first publications on EC (C. Staats & Staats, 1957) and, as outlined above, contingency awareness is one of the central testable differences between propositional single-process views and dual-process models. Mitchell et al. (2009) also allocate a central role to contingency awareness and state: “If all learning is aware, there is less to be gained from postulating an automatic link-formation mechanism in addition to a propositional reasoning mechanism” (p. 189). The prediction of propositional single-process views that contingency awareness is necessary for EC effects to form allows for a clear falsification of the propositional single-process view proposed by Mitchell et al. (2009). It is therefore the main question of this dissertation whether evaluative conditioning requires contingency awareness, or whether EC effects can be found in the absence of contingency awareness. In the next paragraph, I will therefore give an overview of recent studies published on the questions whether contingency awareness is necessary for EC effects to form.

One way to investigate whether EC effects might arise without contingency awareness is to see whether an EC effect exists in people who report no contingency awareness. In multiple studies, EC effects were found for participants who could not identify whether a given CS was presented with a given US (Olson & Fazio, 2001). Other studies reported that participants who could not select a given US that was paired with a CS displayed EC effects for these CSs nonetheless (Baeyens et al., 1990; Walther & Nagengast, 2006). Another set of studies found that EC effects were found even though participants could not name the US valence for a given CS (Baeyens et al., 1990; Dickinson & Brown, 2007). Contrasting these findings, in a series of experiments an EC effect was observed only for participants who could report the US valence for a given CS but not for participants who were unable to report the US valence for a given CS (Stahl & Unkelbach, 2009; Stahl, Unkelbach, & Corneille, 2009). The authors argue that previous studies showing EC effects in the absence of contingency awareness might have been due to the fact that participants were aware of the US valence a CS was paired with (and therefore contingency aware) but not aware of the identity of the US. In another study, Pleyers, Corneille, Luminet, and Yzerbyt (2007) argue that contingency awareness should not be measured on a person level but instead on an item (CS) level. This argument seems plausible as one could consider that an EC effect is driven by the contingency awareness of merely one or two CS-US pairs. This person might be classified as "unaware" but the actual EC effects is driven by only a few (remembered) pairs. Indeed, Pleyers et al. (2007) found EC effects only for remembered CS-US pairs, but not for CS-US pairs that were not remembered, when running an item based statistical analysis.

However, studies showing EC effects only for remembered CS-US pairs might have overestimated contingency awareness in participants. Specifically, participants might have used their affective reaction towards the CS to inform their decision whether a CS was paired with a positive or negative US and did not possess actual contingency awareness (Hütter et al., 2012). This "affect-as-information" explanation (Schwarz & Clore, 1983) could result in an overestimation of actual knowledge of the CS-US valence pairings and would bias any analysis against EC without contingency awareness (Gawronski & Walther, 2012; but see Hütter & De Houwer, 2017 for empirical indications against this explanation).

The most promising attempt to detect EC effects in the absence of contingency awareness stems from Hütter et al. (2012) who used a process-dissociation (PD) approach to tackle the "affect-as-information" problem. PD procedures have previously been used to disentangle explicit and automatic memory (Jacoby, 1991). In the experiment by Hütter et al. (2012), after an EC learning phase, participants were asked to indicate whether a given CS was paired with a positive or negative US and instructed to guess based on their affective reaction when they did not know. Participants were assigned to two different conditions, inclusion or exclusion. In the inclusion condition participants were asked to respond "pleasant" when they remembered that a CS had been paired with a pleasant US, or if they had a pleasant evaluation of the CS. Participants were asked to respond "unpleasant" when they remembered that a CS had been paired with an unpleasant US, or if they had an unpleasant evaluation of the CS. The other half of participants were assigned to an exclusion condition. Here participants were asked to reverse their responses in the task (e.g., participants had to respond “unpleasant” when they remembered that the CS had been shown with a pleasant US or when they evaluated the CS as pleasant). Using a multinomial processing tree model (Batchelder & Riefer, 1999) this experimental design now allowed for a calculation of three parameters: a memory parameter which captured the memory for the CS-US valence contingency, a guessing parameter which showed bias towards “pleasant” or “unpleasant” responses and most importantly an attitude parameter which captured the attitude towards the CS. As the attitude parameter is contingent on the absence of the memory parameter (see Figure 2 - 4, Hütter et al., 2012) an attitude parameter larger than 0 would indicate that EC effects can exist in the absence of CS-US valence memory. In four experiments Hütter et al. (2012) repetitively showed a significant deviation from 0 of the attitude parameter.

While some studies were able to replicate the finding of an attitude effect without memory with the PD procedure (Hütter & De Houwer, 2017; Hütter & Sweldens, 2013), other studies were not able to replicate the finding (Mierop, Hütter, & Corneille, 2017; Mierop, Hütter, Stahl, & Corneille, 2018). Nevertheless, a small meta-analysis of the previously mentioned studies containing 12 experiments indicated the presence of a small effect (Mierop et al., 2018). The fact that the attitude effect discovered by the PD procedure appears to be unstable—even though in most studies the original material and experimental script were used—should call for additional research to ensure that the effect can be reproduced by independent research groups.

One additional study using the PD procedure is of particular interest as it demonstrated that an attitude EC parameter was found in the PD procedure, when participants were given only an instruction of CS-US pairs (i.e., participants were told that they would see a specific CS being presented with a US) and were not actually shown the CS-US pairs (Hütter & De Houwer, 2017). The results of the PD procedure indicated that an attitude EC effect without memory was present, even though we can be certain that the CS-US pair was consciously processed when it was instructed. These results indicate that the attitude parameter of the PD procedure should be interpreted with caution and might not reflect EC without contingency awareness.

Given the discussed studies on EC and contingency awareness, the evidence for EC effects in the absence of contingency awareness is mixed. The PD procedure, while appearing promising, does not deliver the final answer on the question of EC in the absence of contingency awareness. One large problem is that memory of CS-US or CS-valence pairs is used to investigate a property (contingency awareness) that is said to be important during the encoding of the CS-US pair. Memory can only serve as a proxy for contingency awareness during the encoding of the CS-US pair as it is prone to additional influences such as forgetting and false memory (Lovibond & Shanks, 2002). Additionally, all procedures discussed above make use of a correlational approach as contingency awareness is measured individually and correlated with an EC effect. This approach drastically limits possible causal explanations between EC effects and contingency awareness and researchers should therefore shift to manipulate contingency awareness experimentally (Gawronski & Walther, 2012). There has been broad agreement that progress in investigations on the necessity of contingency awareness in EC can only be made when experimentally manipulating CS-US contingency awareness during the learning phase (Corneille & Stahl, 2018; Gawronski & Bodenhausen, 2014b). In the next paragraph, the literature that attempted to experimentally manipulate contingency awareness in EC will therefore be discussed.

Subliminal EC. When investigating the neuronal processing of stimulus perception two factors appear to be important: stimulus strength and attention towards the stimulus (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006). When the stimulus strength and attention are high a stimulus can be processed consciously and can therefore be reported. When both factors are low, only very weak neuronal activations can be observed and no effect on behavior can be expected. The other two combinations of the factors are of greater interest for the research at hand: When stimulus strength is high but attention is low a stimulus will be processed preconsciously (i.e., while the stimulus is processed it cannot be reported consciously as long as the attention is elsewhere). When stimulus strength is low but attention is high the stimulus will be processed subliminally (i.e., below the threshold of consciousness).

One study that implemented a presentation schedule of CSs and USs that assembles the conditions for preconscious conditioning is the “surveillance paradigm” by Olson and Fazio (2001). In this paradigm, participants are asked to react to one specific stimulus item while being presented with a stream of images. Within this stream of images, CS-US pairs are “hidden” by additionally showing pairs of two USs, pairs of two CSs, or single images. In this paradigm, EC effects are found even though participants appear to be unaware of the CS-US contingencies (Olson & Fazio, 2001). The implicit misattribution model (IMM, Jones, Fazio, & Olson, 2009; Jones, Olson, & Fazio, 2010) gives an extensive explanation for the effects found in the surveillance paradigm and argues that the explanation might be suited for other methods as well.

The IMM postulates that the affective reaction towards the US is misattributed to the CS without participants' awareness (Jones et al., 2010). The optimal circumstances for implicit attribution—according to Jones et al. (2010)—are when (1) CS and US are presented simultaneously (which is considered the most important factor in the IMM) (2) CS and US have a spatial proximity and (3) the CS is salient relative to the US. As these factors might be important for automatic effects with subliminal presentations as well, we will pick up these arguments again in the following chapters.

Coming back to the model proposed by Dehaene et al. (2006), subliminal processing of stimuli is possible when the attention is high but the stimulus strength is low (e.g., by presenting the stimulus very briefly). While the presentation time it too short to consciously see the stimulus and experience it, the stimulus should be processed neuronally (Dehaene et al., 2006). As no higher-order thoughts about the stimulus can be formulated when it is presented subliminally, contingency awareness between the stimulus and another supraliminally presented stimulus can be ruled out. A brief presentation of stimuli does not automatically guarantee that a stimulus is indeed presented subliminally, as the stimulus might still be perceived consciously. I thus define a stimulus to be presented truly subliminally if it cannot be identified above chance after the presentation. Therefore showing an EC effect when one of the two stimuli is presented truly subliminally in an EC learning phase would show that EC does not depend on contingency awareness, which is not in line with propositional single-process views (Mitchell et al., 2009). The subliminal presentation of one stimulus is therefore one way to experimentally manipulate contingency awareness in an EC paradigm.

While some researchers have serious doubt whether the current empirical evidence is supportive of EC effects in the absence of contingency awareness (Corneille & Stahl, 2018; Stahl, Haaf, & Corneille, 2016; Sweldens et al., 2014), other researchers have already concluded that the question has been resolved by empirical evidence showing that EC can occur in the absence of awareness (Jones et al., 2010). Specifically, the basis for this statement stems from studies showing an EC effect with a subliminal presentation schedule. To evaluate the persuasiveness of these studies I will now review the current corpus of literature on evaluative conditioning with subliminal stimulus presentation to evaluate the strength of the literature reporting EC effects with subliminally presented stimuli (for an overview of the literature search see Appendix A).

Three methodological factors might be important when reviewing the current literature reporting EC effects with subliminally presented stimuli: (1) whether the CS or US had been presented subliminally, (2) whether US valence was manipulated within or between participants, and (3) whether the visibility of the subliminally presented stimuli was controlled for. All points will be discussed when reviewing the corresponding literature. I will first review studies that used a subliminal US, starting with studies manipulating US valence between participants followed by studies manipulating US valence within participants. Afterward, I will review the literature that showed EC effects with a subliminally presented CS.

Subliminal US - US valence between participants.

In a study by Niedenthal (1990) images of people looking disgusted or joyful (or disgusted/joyful/neutral in Experiment 2) were used as USs that were presented for 2 ms. Afterward, a comic figure was presented as the CS. In a visibility test after the learning phase, the CSs were shown again, this time preceded by a US in half of the trials. Participants were asked to indicate after every trial whether a US had been shown or not and results showed that they were not able to indicate US presence above chance. In Experiment 1 an affective priming effect was found (indicating a preference for CSs paired with a joyful face over CSs paired with a disgusted face), while explicit ratings yielded mixed results in Experiment 2: When asked to judge whether positive/negative traits were characteristic of CSs, a difference was found when positive USs were paired with the CS (i.e., the CS was judged to have more positive than negative traits). This effect was also found for CSs paired with a neutral US and no difference in the judgment was found for CSs paired with negative USs. In another dependent variable of the same experiment, it was found that CSs that were paired with negative USs were judged to be more similar to negative than positive unrelated stimuli. Interestingly, CSs paired with neutral stimuli were also judged to be more similar to negative than to positive unrelated stimuli and no difference was observed for CSs paired with positive USs. While the visibility test is well designed and simulated to look very similar to the learning phase, it is possible that a (small) subset of participants saw the valence of the USs in one trial and acted upon this knowledge. Two potential problems can be identified as US valence was manipulated between participants. First, a conditioning effect could be explained if participants recognized the valence of a US in a single trial and evaluated all stimuli accordingly after the learning phase. Second, even if valent stimuli were processed subliminally, we do not know whether an EC effect caused a difference in evaluations or a change in evaluations was due to a difference in mood between the groups (Sweldens et al., 2014). Considering that the results are mixed, showing effects in one measure but not another and showing diverging effects for neutral USs on different dependent variables, the results should be interpreted with caution.

In two experiments by Krosnick, Betz, Jussim, and Lynn (1992) an EC effect was observed with briefly presented valent images as USs (presented for 13 ms) and visible CSs depicting a person. Again, US valence was manipulated between participants. While the first experiment had an apparent weakness (the researcher in the room was not blind to the US valence condition; see Gilder & Heerey, 2018), in the second experiment this problem was eliminated by a double-blind design. In Experiment 2 a visibility test after the learning phase was administered that was similar to the learning phase. However, participants were shown words and images (compared to only images) presented prior to the target person and were not asked about the US valence, but instead had to indicate whether they saw a word or a picture. While this visibility test is not highly indicative for the question whether participants perceived the US valence during the learning phase, it gives an indication that the USs were not easy to perceive as participants were not able to differentiate between words and images. The results of Experiment 2 indicate that the USs influenced the attitude towards the target person and the perceived personality of the target person in line with the valence of the USs. The effect was only found when participants had to judge the target person (CS) on an attitude and personality index, but not on an attractiveness scale. A MANOVA with all three dependent variables indicated that there was no effect of the briefly presented USs on the different ratings. The results should additionally be interpreted with caution as the individual test statistics for the three dependent variables in Experiment 2 report different degrees of freedom which is not explained in the article and some items were left out when computing the scores for the dependent variables (Krosnick et al., 1992, p. 160, Footnote 4).

Subliminal US - US valence within participants.

One study claimed to have found that unconscious processes are more sensitive to facial expressions than to verbal information. Specifically, faces with either a happy or angry expression (CS) were paired with the words “happy” or “angry” as USs (N. Kawakami, Miura, & Yoshida, 2015). The happy face was paired with the word “angry” and the angry face with the word “happy” either for 10 ms or 500 ms (manipulated between participants). In the subliminal condition, both CS and US were presented subliminally. After the learning phase participants had to rate how happy/angry the person in the photo (which looked happy or angry) appeared. There appeared to be no difference between the evaluations of the faces when the photo/word pairs were presented for 500 ms. When both words and images were presented subliminally the face looking angry was rated as more angry/less happy than happy faces. The results merely show that when participants received information contradicting the facial expression supraliminally the evaluation of the face is mixed and the happy face is not evaluated more positively than the angry face ($N = 23$ in this group, so a null effect should not be over-interpreted). In the subliminal condition the participants likely did not see or process the words or faces and therefore evaluations are in line with the facial expression participants see when rating the image. As the study does not allow to test for the influence of the subliminal presentation on an EC effect, the study will be ignored in all further analyses and discussions.

Rydell and colleagues published multiple studies showing subliminal EC effects on implicit measures of attitudes (Rydell & McConnell, 2006; Rydell et al., 2008; Rydell, McConnell, Mackie, & Strain, 2006). One main difference to the previously discussed literature was that participants explicitly learned information about a target character, of whom an image was shown, while also being presented with briefly presented valent words of the opposite valence shown before the image of the target character (for a detailed description see Chapter 2). In attitude measures after the learning phase, the explicit attitude reflected the learned valence from the explicitly learned information about the target character, while the implicit attitude measures reflected the valence of the briefly presented valent words. A memory check after the learning phase served as a proxy for the visibility of the briefly presented words during the learning phase and indicated that they were not remembered above chance. As the explicitly learned behavior had always the opposite valence to the briefly presented words (Rydell et al., 2008, 2006) mood effects can be ruled out as an alternative explanation for differences in the IAT effects. The results by Rydell and colleagues are considered to be one of the most convincing demonstrations of subliminal EC effects (Sweldens et al., 2014).

In another study, Field and Moore (2005) presented a visible image CS, followed by a subliminal valent image as US (17 ms) and a backward pixel mask to participants. In an explicit evaluation, an EC effect was found, but only when participants were not distracted by a cognitively demanding task during the learning phase. The same pattern was found for subliminally and supraliminally presented USs. Notably, valence was manipulated within participants and a relatively thorough debriefing tested for the visibility of the briefly presented USs: In the subliminal awareness test participants were presented with one of the four different pixel masks and had to respond whether they noticed an image being presented before the mask (and if yes what valence it had). This test, however, does not test whether participants remembered the CS-US contingency, but the US-mask contingency and it neither measures the visibility of the USs during the learning phase. In an additional contingency awareness test, participants were shown learned and new CS-US pairs and had to indicate whether these pairs had been shown during the learning phase. However, the new CS-US pair always had the same valence as the actual CS-US pair. If participants therefore remembered the US valence that was shown with a given CS (but not the identity of the US), participants might be classified as unaware even though they might have been aware of the US valence during the learning phase.

In an experiment by De Houwer et al. (1994) a neutral word as CS was followed by a valenced word as US, which was shown for a brief moment (29 ms). The valence was manipulated within participants and mood effects can therefore be ruled out as an alternative explanation. The awareness check consisted of the questions whether the participant had "noticed anything strange during the experiment" and a more explicit question asking whether participants had seen something between the neutral word and the backward mask. Only one participant reported having seen something and was excluded from the analysis. In an explicit evaluation measure at the end of the experiment words paired with positive USs were evaluated as more positive than words presented with negative USs.

Importantly, De Houwer, Hendrickx, & Baeyens (1997) ran four replication studies of the experiment, adding a strict visibility test in three of the studies. At the end of each experiment, participants were told that a word with positive or negative valence would be briefly presented during each trial and participants were presented with the same trials as in the learning phase. After every trial participants were then asked to guess whether the trial included a positive or negative word. Participants were classified as aware of the US valence when they correctly named the valence in 16 out of 24 trials. While it is still possible that participants saw one of the stimuli in the learning phase (but did not report this during the additional debriefing when asked whether they saw something) the study indeed shows a better visibility test than the previously discussed studies. In two of the four experiments, no EC effect was found (with one study being a direct replication of De Houwer et al., 1994). The authors also note in the discussion of the second experiment, showing an EC effect with briefly presented USs, that "the observed evaluative learning effect was once again weak [. . .] despite the fact that various aspects of the procedure [. . .] were optimized" (p. 99). Taken together the studies by De Houwer et al. (1997) show the most robust method and intensive tests for the visibility of briefly presented USs compared to the previously discussed literature, but the results are mixed. Nevertheless, specific characteristics of this set of studies should be taken into account in future studies testing EC with briefly presented stimuli: an extensive visibility test, explicit evaluations, and within subjects manipulation of US valence.

In an experiment by Fulcher and Hammerl (2001) visible Japanese characters served as CSs and different smiling/frowning faces were used as USs, which were presented for a variety of presentation times (12.5 ms, 25 ms, 50 ms, 125 ms). US valence was manipulated within participants and a visibility check was employed during the learning phase. Participants were told that they would see frowning or smiling faces and had to press a corresponding key in every trial to judge whether the face was frowning or smiling. No EC effect was found when the USs were presented for 12.5 ms, nor when the USs were presented for 50 or 125 ms. An EC effect was found when USs were presented for 25 ms, but the results of the visibility check show that the facial expressions were correctly identified above chance in the 25 ms condition, therefore indicating that the found effect cannot be considered to be a subliminal EC effect. Interestingly, an EC effect in the 12.5 ms condition was found, when analyzing the data only for participants who scored low on reactance measures (i.e., participants who do not mind being controlled or regulated by others). Based on a number of characteristics of the study by Fulcher and Hammerl (2001), the results should be interpreted with caution. First, as discussed above, even though trial by trial visibility data was available, an analysis of visibility was only done over participants for each presentation time (i.e., classified a participant as either aware or unaware) and not on an item level (see Pleyers et al., 2007). Therefore a participant might be aware of one or two USs—especially as participants knew which kind of USs to expect—but classified as unaware. This subset of participants might show an EC effect which results in an overall EC effect for participants classified as unaware. Second, no EC effect was observed for clearly visible stimuli (i.e., in the 125 ms condition) which could be considered a failed manipulation check, as EC effects with visible stimuli have consistently been found (Hofmann et al., 2010). Third, the authors admit that the used reactance measure is generally used to measure conscious reactions towards control (one item of the scale is for example: “Regulations trigger a sense of resistance in me”) and it is therefore difficult to explain why results in the subliminal—but not supraliminal—presentation condition were moderated by a reactance measure. While the results should be interpreted with caution, the experimental paradigm used by Fulcher and Hammerl (2001) seems promising as valence was manipulated within participants and visibility was measured during the learning phase. The latter point is especially interesting, as contingency awareness is only necessary during the acquisition of the CS-US pair according to propositional single-process views (Mitchell et al., 2009). A visibility check during the learning phase is therefore a better proxy for contingency awareness than a CS-US memory check at the end of the experiment. These procedural characteristics should be taken into consideration when running additional experiments.

While the studies that present a subliminal US and manipulate US valence within participants show a higher methodological quality (e.g., better visibility checks and US valence manipulated within participants making sure that mood effects can be ruled out), the results are mixed. On the one hand, some studies appear to show a reliable EC effect (Rydell & McConnell, 2006; Rydell et al., 2008, 2006) but do not use a visibility check. On the other hand, studies that implemented a better visibility check do not show reliable EC effects (De Houwer et al., 1997; Fulcher & Hammerl, 2001). One additional possibly important problem of studies using a subliminal US could be that affective stimuli appear not to evoke an affective response in participants when participants are unaware of the stimuli (Lähteenmäki, Hyönä, Koivisto, & Nummenmaa, 2015). If it turns out to be true that US valence can only evoke affective reactions in participants once the US identity has been processed, EC paradigms claiming to have found an EC effect with truly subliminally presented USs should be interpreted with caution. I will therefore discuss studies that presented the CS subliminally in the next section.

Subliminal CS.

The only study showing subliminal EC effects with a subliminally presented CS was obtained by Gawronski and LeBel (2008) who presented the CSs “Europe” or “Asia” for 17 ms which were followed by either a supraliminally presented positive or a negative US word. Valence was manipulated within participants and the combination of the CS and US valence was counterbalanced over participants (i.e., half of the participants saw “Europe” and positive USs and “Asia” and negative USs and vice versa for the other half of participants). Participants were instructed to “take a moment to think about their feelings about Europe and Asia” or to “take a moment to think about what they know about Europe and Asia”. The results indicated an EC effect for implicit measures regardless of the introspection condition but an EC effect for explicit measures only when people were instructed to introspect about their feelings. While the result appears intriguing the study does not report any visibility test and one should therefore be cautious interpreting the results as the presentation might not have been truly subliminally.

Studies similar to EC.

As there are very few studies with a briefly presented CS in an EC learning phase, I will also discuss a few other studies that presented a CS subliminally but might not be considered evaluative conditioning, as they either do not use valent USs or do not use evaluative measures. These studies might nevertheless give an insight into automatic association formation as the procedure is highly similar to the procedures in the previously discussed literature and might even be considered to be EC by some (e.g., some studies were included in the meta-analysis on EC by Hofmann et al., 2010).

Dijksterhuis (2004) published a six study article in which he demonstrated that an evaluative conditioning procedure can raise (implicit) self-esteem. Participants were presented with the Dutch word for “I” and positively valenced (or neutral) words, which were always presented for 17 ms. The word for “I” was either presented supraliminally (study 1 and 2) or subliminally (all other experiments). In all studies, a raised self-esteem was observed with a variety of implicit measures (IAT, name letter effect, better mood). The results of the study should be interpreted with caution for a variety of reasons. First, some sample sizes were very small making it surprising that an EC effect was found (e.g., in Experiment 3, 16 participants were divided into two groups and a difference in IAT scores was found, see Simmons, Nelson, & Simonsohn, 2011). Second, in five of the six experiments, the valence was manipulated between participants, therefore allowing a mood effect to be an alternative explanation for any observed differences between groups. Third and most importantly, no actual visibility test was conducted, but participants were asked whether they “had seen anything unusual" and whether they "had seen words flashing on the screen" during the debriefing. No participant in any of the studies reported having seen flashes. While it is possible that the valent words were truly subliminal, the task seems inadequate to test for stimulus visibility, as participants might just reply "no" in the debriefing to finish the task quickly.

Two publications reported three experiments (Grumm, Nestler, & von Collani, 2009; Jraidi & Frasson, 2010) replicating the findings by Dijksterhuis (2004). Importantly, effects were generally found on implicit measures of self-esteem (IAT and name letter effect) but only for explicit measures when participants were instructed to introspect on their feelings (Grumm et al., 2009, Exp. 3). As in the studies of Dijksterhuis (2004), no visibility check was conducted in any of the three experiments, leaving open the question whether the found effect was due to truly subliminally presented CSs and USs. In a recent study by Versluis, Verkuil, and Brosschot (2017), the experiment of Dijksterhuis (2004) was replicated. However, no effects of the learning phase on explicit nor implicit measures of self-esteem nor on physiological measures associated with self-esteem were found. In these experiments, a visibility test was employed (which was not described in detail in the article) and the authors conclude that the CS-US presentation was indeed subliminal.

Another set of studies closely related to the experiments by Dijksterhuis were published by Riketta and Dauenheimer (2003; for a successful replication see Svaldi, Zimmermann, & Naumann, 2012). In four experiments "I" and positive/negative words were presented subliminally, but again, no visibility check was conducted. In contrast to the previously mentioned studies changes on explicit (and implicit) measures of self-esteem were found. Riketta and Dauenheimer (2003) also investigated whether the found effect might be caused by EC processes or priming processes. They therefore presented a first name ("Leo") paired with negative/positive adjectives to a subset of participants. In an evaluation of "Leo" no effect of the pairing condition (positive/negative adjective) was found and the authors conclude that the effect might be due to "selective activation of evaluative self-knowledge rather than from changes in self-valence associations" (Riketta & Dauenheimer, 2003, p. 689). Evidence for subliminal EC in relation to self-esteem is mixed: While some studies show support for explicit self-esteem changes, others show only support for implicit changes of self-esteem, while another set of experiments shows no influence of the procedure on any measures. The latter study was also the only study who reported to have measured the actual visibility of the subliminally presented CS-US pair. It is additionally important to consider the results of Riketta and Dauenheimer (2003) who did not find an effect when using a first name as a CS and valent words as USs, which indicates that these studies might not have investigated EC effects.

In a series of studies by Custers and Aarts (2005), a CS describing a behavior (e.g., "studying") was briefly presented (23 ms) before a supraliminally presented US (adjectives with positive/negative valence). Afterward, participants showed that they wanted to act more upon the behavior that was paired with positive words than the behavior paired with negative words (Exp. 1, 2a, 2c, 3) and also evaluated the behavior as more positive (Exp. 2b). However, in four out of six studies the US valence was manipulated between participants, allowing for the alternative explanation that the result was caused by mood effects. Importantly, no concrete measures to detect whether the briefly presented CSs were indeed perceived subliminally are discussed in the article. It is merely mentioned that "participants were thoroughly debriefed and checked for awareness of the subliminally presented behavioral states" (p. 132). Considering the lack of a visibility test and that valence was manipulated between groups in four out of the six studies the results should be interpreted cautiously.

Galli and Gorn (2011) tested whether semantic meaning can transfer subliminally. The words "black" and "white" were presented as primes for 26 ms before the letters "I" or "G" were shown (the letters were Chinese characters as the study took place in China). In a semantic priming task an increased reaction time was observed when the letters were incongruent to the color of a shown word (e.g., when "I" was shown with "black" in the learning phase, an increase in reaction time was shown when "I" was presented before "rice" compared to "soy sauce"). Similar, in a second experiment products that were named either "I" or "G" were evaluated as more positive when the name and the color of the product (Soy milk vs. Cola) matched (i.e., a cola named with a letter paired with "black" was evaluated as more positive than a cola named with a letter paired with "white"). To test whether the words "black" and "white" were indeed presented subliminally a visibility test was employed at the end of the study. Participants saw 30 trials taken from the learning phase but this time were instructed that something might be shown before the words "black" and "white". After every trial participants were asked whether they saw something and if yes what they saw. About one third of the participants reported at least one correct letter ("I" or "G") in this test phase. While Galli and Gorn argue that visibility might be higher in the test phase than during the actual learning phase—because participants were made aware of a possible subliminal presentation—it is remarkable that one third of the participants named the correct letter indicating that they consciously perceived the letter. One can therefore speculate whether the effect observed in the evaluation and priming task might be driven by the subset of participants who saw the briefly presented words.

In a setup more related to Pavlovian than evaluative conditioning, Both et al. (2008) showed sexually arousing images for 30 ms as CSs and administered genital vibrotactile stimulation as USs. Another sexually arousing image was not paired with anything during the learning phase but also shown for 30 ms. As no other images were displayed, participants were therefore supposed to only consciously perceive the backward masks during the learning phase. To check for the visibility, participants were shown additional learning trials with both CSs presented after every trial. Participants were on average no better than chance to select the correct CS, indicating that the images were shown subliminally. An evaluation (at the very end of the study) did not show any differences between the CSs. Yet, when only analyzing the first trial of a post-conditioning extinction procedure, a higher vaginal pulse reaction (i.e., sexual arousal) towards the CS paired with the stimulation was found compared to the CS without the stimulation. The data should be interpreted with caution, as the researchers found no evidence for a difference in skin conductance, differences in evaluation and only a difference on sexual arousal when running a one-sided t test on the first trial of the extinction phase. Additionally, both images were already sexually arousing before the conditioning, making it difficult to find a difference caused by the learning phase between the CSs with only 18 participants.

Subliminal EC effects.

What can be learned from the corpus of studies showing EC effects and effects highly similar to EC with subliminal stimulus presentation? One question would be whether a specific presentation schedule (e.g., subliminal US vs. subliminal CS) and measurement (e.g., implicit versus explicit) appears to be dominant in the literature which could indicate that it is easier to find subliminal EC effects with this setup. As can be seen in Table 1, there appears to be no combination of methods that stand out.

One observation that appears remarkable is that 15 out of the discussed 18 publications reporting subliminal EC were published by different authors. One could hypothesize that if a researcher would have found a reproducible effect it would be

Table 1
Experiments reporting subliminal EC effects split by presentation schedule of CS/US and their measurement

	Implicit Measures	Explicit Measures
Subliminal US	9	10
Subliminal CS	12	7

Note. Single Experiments might be counted twice if using both implicit and explicit measures (e.g., Gawronski & LeBel, 2008) or presenting both CS and US subliminally (e.g., Dijksterhuis, 2004).

used for a series of different studies testing for boundary conditions. The only two researchers who published multiple papers on EC are De Houwer and Rydell. The second set of studies by De Houwer and colleagues (1997) show mixed effects, but the effect in the papers by Rydell and colleagues (2006, 2006, 2008) seem to be stable. It seems to be promising to replicate the findings by Rydell and colleagues (2006, 2006, 2008) to investigate the boundary conditions of EC with briefly presented stimuli (see Chapter 2).

Another observation is the heterogeneity of visibility checks testing whether the "subliminally" presented stimuli were not consciously perceived. Some studies even show that the presentation was not subliminal (e.g., Galli & Gorn, 2011), while other studies did not employ any procedures testing whether the briefly presented stimuli were consciously perceived (e.g., Gawronski & LeBel, 2008). An additional problem in research using subliminal presentation is that subliminality checks appear to have less statistical power than the test of the dependent variable. It is therefore possible that statistical tests yield that stimuli were not visible above chance even though they were, while statistical analyses testing for a difference in the evaluation yield significant differences (Vadillo, Konstantinidis, & Shanks, 2015). As most theorists supporting a propositional single-process view would agree that EC effects with visible—but briefly presented—stimuli would be in accordance with a propositional single-process view, many of the previously discussed studies might not be suited to differentiate between single- and dual-process views because they cannot claim that stimuli were presented truly subliminally.

This problem has been previously discussed by researchers who argue that poor contingency awareness reported in EC studies could be the result of procedural rather than theoretical factors (Davey, 1994). In a similar vein, De Houwer states: "I do not exclude the possibility that evaluative conditioning effects are based exclusively on the operation of propositional processes. In that case, evidence for evaluative conditioning effects that are resistant to extinction and independent of contingency awareness might indeed result from methodological problems" (De Houwer, 2007, p. 238).

To further analyze the reviewed literature, a *p*-curve analysis and a meta-analysis were conducted to evaluate the evidential value and estimate a mean effect size of the reviewed literature.

A *p*-curve analysis. One aspect of the reviewed literature that seemed worthy to further investigate was the proportion of reported *p* values close to 0.05. The distribution of *p* values can give insights into the evidential value of the published literature because *p* values are distributed differently when a true or a null effect is investigated. Specifically, *p* values over many studies investigating a true effect should be right-skewed (i.e., more *p* values should be between .01 and .02 than between .04 and .05). If however, a series of studies report on a null effect, *p* values should be evenly distributed. As statistically significant results are more likely to be published (Greenwald, 1975; Sterling, Rosenbaum, & Weinkam, 1995) it would be possible that all reported significant results actually report on an effect that does not exist (i.e., an α-error occurred). A significant right-skewed distribution of *p* values can therefore indicate that the reviewed studies carry evidential value (Simonsohn, Nelson, & Simmons, 2014). There are two different ways the right-skewness is calculated in the *p*-curve method proposed by Simonsohn, Simmons, and Nelson (2015). First, all values are taken ("full") and skewness is tested and second, only *p* values smaller than .025 are used ("half"). The "half" method is said to be less affected by (ambitious) *p*-hacking (i.e., using researchers' degrees of freedom to re-analyze the data until a significant effect is found).

To calculate the *p*-curve, one statistical test reporting a significant result per study per measure showing the presence of a subliminal EC effect was selected from the above-discussed literature. In total 43 statistical tests from 32 experiments were selected and *p* values were re-computed using the reported F and t values. One-sided *p* values were calculated if this was mentioned in the corresponding article. Four statistical test values were excluded as the corresponding *p* value was larger than .05, resulting in 39 significant tests. The *p*-curve was calculated using the R code of the *p*-curve app 4.06 (Simonsohn et al., 2015).

Results show that the p-curve is right-skewed, indicating that the studies carry evidential value (full: $z = -2.22; p = .0131$, half: $z = -3.18; p = .0007$, see Figure 1). Additionally, Simonsohn et al. (2014) suggest comparing the observed p-curve to a theoretical p-curve that would be expected when running studies with 33 % power on average (and both analyses are reported per default by the p-curve app; Simonsohn et al., 2015). The observed p-curve was significantly flatter (i.e., significantly less right-skewed) than the expected p-curve with 33 % power (full: $z = -3.63; p = .0001$, see Figure 1), which—contrary to the skewness test—indicates the absence of evidential value by the reviewed studies (Simmons & Simonsohn, 2017). The analyses of the p-curve mirror the results of the review above: Some studies show clear evidence for EC (leaving the question of visibility aside), while other studies do not show convincing evidence. The results of the p-curve analysis can therefore be understood as a clear sign that more research is needed to test the possibility of EC effects without contingency awareness by means of subliminal presentation.

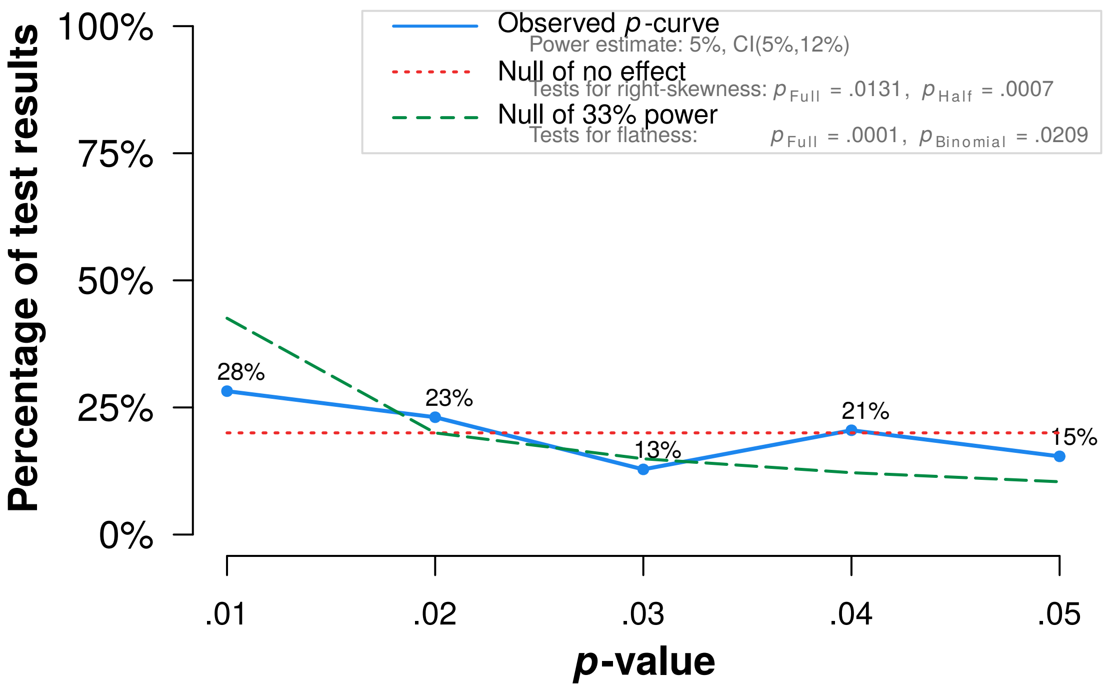

Figure 1. A p-curve with all 39 statistically significant p values observed in the literature review, of which 21 are $p < .025$.

Previous investigations have shown that the conclusion drawn from the p-curve can be influenced strongly by a single publication (Simonsohn et al., 2015). As the results by Rydell and colleagues appear important, because they are the only studies replicated systematically by the same lab, I tested the influence of the results reported by Rydell and colleagues on the p-curve. Therefore, the same p-curve analysis was conducted again, without the statistical values of the three publications by Rydell and colleagues. Testing the right skew of the p values the half p-curve is still significantly right-skewed (but less convincing than when the results of Rydell are included) while the full p-curve test does not indicate that the p-curve is skewed[2]: (full: $z = -1.01; p = .1551$, half: $z = -2.08; p = .0187$, see Figure 2). The observed p-curve is still significantly flatter than the theoretical p-curve with a power of 33 % (full: $z = -4.61; p < .0001$, see Figure 2). The results are again mixed and even weaker than the results of the analyses including all p values. The additional p-curve analysis demonstrates the importance of the research of Rydell and colleagues when assessing the evidential value of all studies published on subliminal EC. Note that the analyses do not allow any conclusions about the question whether the results of Rydell and colleagues are correct; they merely demonstrate that the results are important in the cumulative analysis of studies reporting subliminal EC. The study by Rydell et al. (2006) will therefore be central in the empirical investigation discussed in Chapter 2.

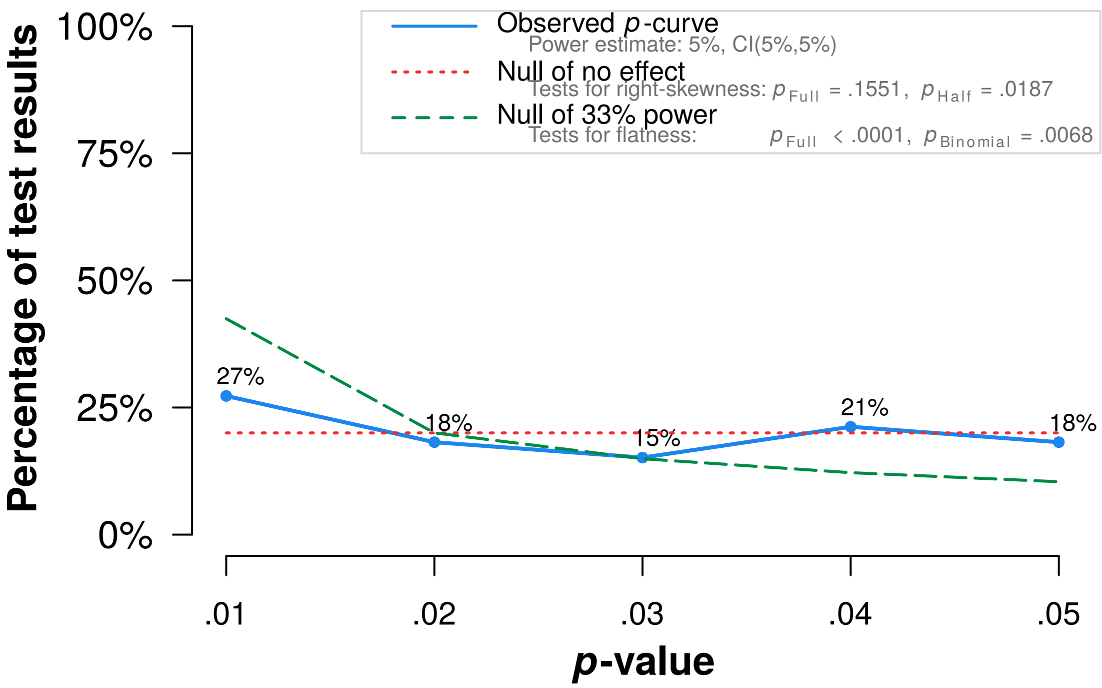

Figure 2. A p-curve leaving out the p values of the three experiments by Rydell and colleagues, resulting in a total of 33 significant p values, of which 16 are $p < .025$.

[2]The results of the p-curve analysis can still be interpreted as carrying evidential value, as the significance value of the half p-curve analysis is smaller than 5 % (Simonsohn et al., 2015). The result of the full p-curve analysis can therefore be ignored.

A meta-analysis. The data of the publications discussed above were also used to run a meta-analysis ($k = 37$). Note that the analysis here constitutes only as a first analysis quantifying the effects of the articles reviewed above and further analyses should take into account publication bias, the quality of the studies, and possible moderators (Hilgard, Engelhardt, & Rouder, 2017). Additional analyses should also attempt to acquire the data of the studies underlying the test statistics to get better predictors of effect size and variability. In a number of publications the degrees of freedom reported and the number of participants did not fit for example. The results of this analysis should therefore not be taken as a definitive answer for or against the absence/presence of an EC effect with briefly presented stimuli. Nevertheless, the analysis will give a quantified overview of the discussed literature and can be used to assess the added contribution of the studies conducted in the next chapters and other studies conducted in our lab to the corpus of literature on EC with briefly presented stimuli (see Chapter 8). Test statistics were only used in the meta-analysis when two mean values were compared in a statistical test to investigate whether an EC effect was present in a subliminal condition. The F and t statistics were taken and a correlation coefficient was calculated which was then transformed to Fisher's Z. For an overview of all undertaken steps to transform the values see Appendix B.

A random effects analysis revealed a mean effect size of $d = 0.45$ (Z = 0.22) with a 95% confidence interval from $d = 0.34$ to $d = 0.57$, see Figure 3. The analysis also yielded that the group was homogeneous $Q = 50.50$, $df = 36$, $p = .055$. All analyses were performed with R (Version 3.4.4; R Core Team, 2016) and the R-packages *MAd* (Version 0.8.2; Del Re & Hoyt, 2014), *metafor* (Version 2.0.0; Viechtbauer, 2010), *papaja* (Version 0.1.0.9735; Aust & Barth, 2016), and *psych* (Version 1.8.3.3; Revelle, 2017).

In the meta-analysis by Hofmann et al. (2010) no significant meta-analytic effect was found for studies using a subliminal US (k = 15; note that the studies by Rydell were not included in this meta-analysis). The effect size found by Hofmann and colleagues for an EC effect with subliminally presented CSs was $d = .49$ ($k = 8$), which is almost identical to the overall mean effect size found by the analysis above using all subliminal EC studies. Concluding, there appears to be an average effect that differs from zero, but the evidential value of the presented studies is weak (see p-curve analysis) and therefore further studies need to be conducted.

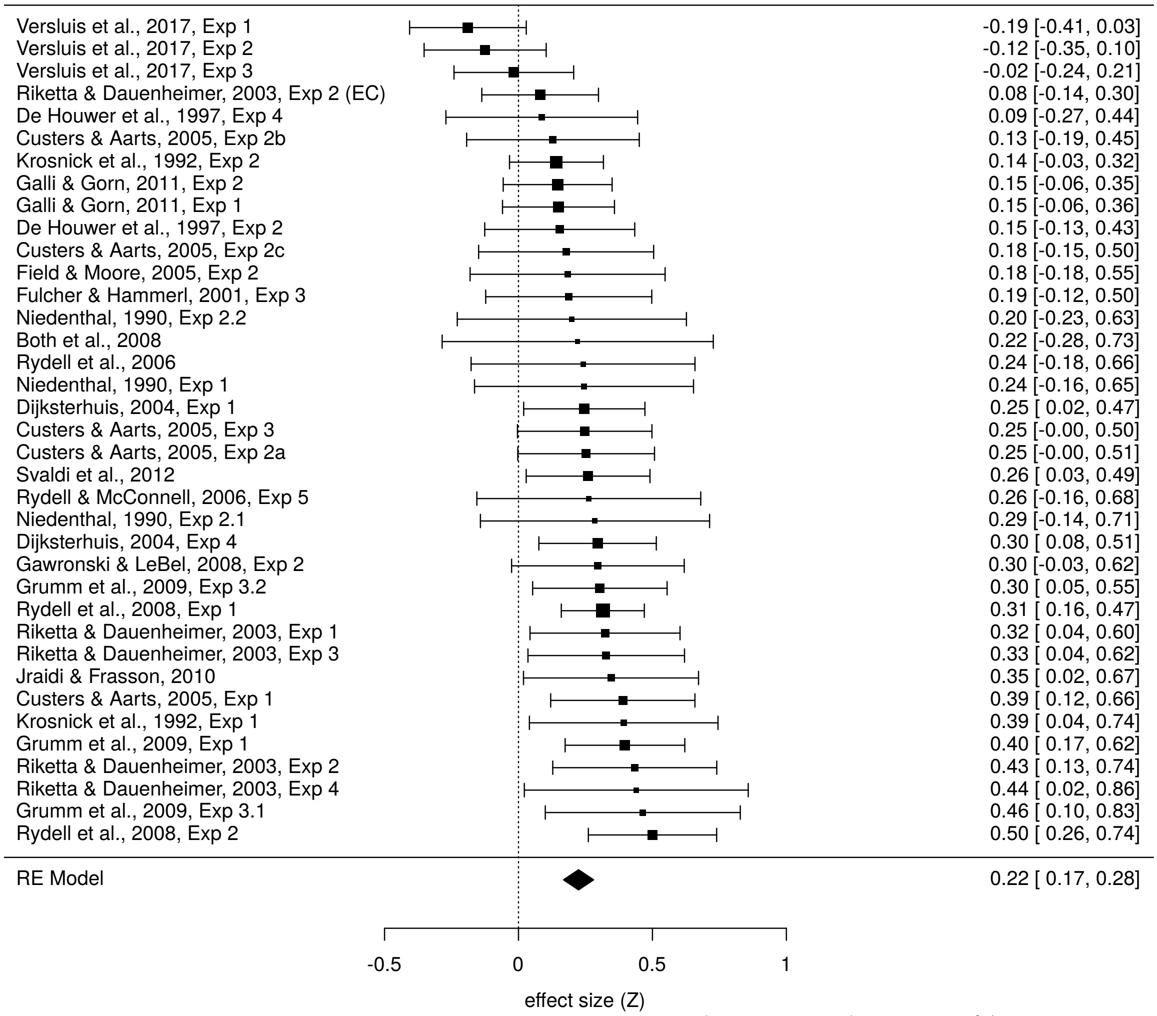

Figure 3. A forest plot showing the effect sizes (Fisher's Z) and 95% confidence intervals for each study included in the analysis based on a random effects model.

Studies showing no EC effect. Along with the two experiments reported by De Houwer et al. (1997) and the replications of Versluis et al. (2017), a series of six experiments by Stahl et al. (2016), testing for subliminal EC effects in a series of (potentially beneficial) conditions, are the only published subliminal EC studies showing a null effect for EC with briefly presented stimuli. In the studies by Stahl et al. (2016) the CSs were always presented briefly and the US was shown for a longer period, for the reasons discussed above. Importantly, in each study visibility of briefly presented CSs was checked during the learning phase by asking participants after every trial to select the presented CS from an array of stimuli. While the results from the visibility test showed that the CSs were recognized (slightly) above chance, no EC effect was found within each experiment for briefly presented CSs. A meta-analysis of the six experiments also revealed no EC effect for briefly presented stimuli.

In another study continuous flash suppression was used to experimentally manipulate CS-US contingency awareness: Specifically, Högden, Hütter, and Unkelbach (2018) used continuous flash suppression to suppress the conscious processing of CSs. Using a stereoscope, participants were presented with a CS to one eye and a high contrast dynamic pattern (e.g., flashing pixel masks) to the other eye. Using this technique the suppressed image (the CS in this case) can be presented for a prolonged period (e.g., 1800 ms) without being consciously perceived because a participant's attention is captured by the flashing pixels. The USs were shown in between the flashing pixel images and US valence was manipulated within participants. Therefore ideal circumstances for automatic EC effects should be met, as participants do not perceive the stimulus consciously, yet it is presented for an increased period compared to subliminal EC studies. An EC effect was found when CSs were not suppressed, but in all four studies no EC effect was found when the CSs were suppressed.

In another study, parafoveal stimulus presentation was used to experimentally manipulate the conscious awareness of the CS (Dedonder, Corneille, Bertinchamps, & Yzerbyt, 2014). Specifically, a US (valence manipulated within participants) was presented in the center of the screen and a CS was presented either 2.5° from the center (foveal condition) or 11.5° from the center (parafoveal condition). The parafoveal presentation allowed for a longer CS presentation than in the subliminal studies (i.e., 60 ms) and contingency awareness appeared to be absent for CS-US pairs in the parafoveal—but not the foveal—condition. EC effects on implicit and explicit measures were only found for CSs in the foveal, but not for CSs in the parafoveal condition.

These three studies contradict previous findings showing EC effects in the absence of awareness, by means of a subliminal presentation. Studies showing subliminal EC effects have been frequently used to demonstrate support for dual-process models such as the APE model (Gawronski & Bodenhausen, 2006, 2011). However, the evidence for EC effects in the absence of contingency awareness is mixed and additional studies did not find evidence for automatic EC effects under optimal conditions for automatic EC effects. It therefore appears necessary to further investigate EC effects with a subliminal presentation schedule to test the central prediction deviating between propositional single-process views and dual-process models: the necessity of contingency awareness in

EC. In the next section, I will discuss recent findings on beneficial factors of automatic EC effects, which will be considered in the empirical studies discussed in the next chapters.

Other beneficial factors for automatic EC effects

In recent publications on contingency awareness in EC, a set of factors has received some attention which will be considered in the series of studies presented in the next chapters. A first point concerns the presentation schedule of the CS-US pair. A second point is the relevance between CS and US in the learning phase and a third point is related to the optimal measurement of automatic EC effects. All three factors will be discussed in the following sections.

CS-US presentation schedule. A first characteristic that might play an important role for EC effects without contingency awareness acquired through associative processes is the presentation schedule of CS and US. One central characteristic in the implicit misattribution model (IMM, Jones et al., 2009, 2010) is that CS and US need to be presented simultaneously for a misattribution effect to occur. Recent studies have investigated the effect of sequential versus simultaneous CS-US presentation on (automatic) EC effects. A study by Sweldens et al. (2010) has shown that automatic EC effects seem to depend on a simultaneous presentation of the CS and US. Also, the previously discussed PD procedure showed that EC by way of an automatic process required simultaneous pairing of stimuli, whereas an EC effect by a propositional process was also found with sequential pairings (Hütter & Sweldens, 2013). Additionally, Stahl and Heycke (2016) showed that EC effects independent of US identity memory were only found with a simultaneous presentation schedule but not with a sequential presentation schedule (but in both conditions, EC effects were only found when participants had US valence memory). Taken together there are clear indications that a simultaneous CS-US presentation might be beneficial for EC produced by associative processes. In the studies presented in Chapter 6 and 7, we therefore paid careful attention to present the CS and US simultaneously.

Relevance. A second characteristic which might be beneficial for automatic EC effects is described in the research by Verwijmeren, Karremans, Stroebe, and Wigboldus (2012). In their study participants were presented with either water bottles or airline brands as CSs for a brief moment (17 ms). The CSs were followed by a US showing a face with a disgusted or fearful expression. As both USs have a similar negative valence the study should not be considered EC, however, the results are intriguing. Thirsty participants showed a preference for water bottles presented with fear compared to water bottles paired with disgust, suggesting that CS and US relevance enhances EC effects, but only when the pairing was (goal-) relevant to the participant, as the effect was only observed in thirsty participants. Another study also demonstrated that subliminal behavioral priming of drink brands was only effective for participants who were thirsty (Karremans, Stroebe, & Claus, 2006).

Studies using supraliminal CS-US presentations additionally demonstrated that EC effects were larger when a face showing disgust was paired with a taste of a beverage than when the face was paired with the color of a beverage indicating a potential importance of this basic human reaction (Baeyens et al., 1996). In Chapter 4, a study will therefore be presented investigating the influence of relevance on EC effects with briefly presented CSs by replicating and extending the study of Verwijmeren et al. (2012).

Choice as a more sensitive measure. A third characteristic worth investigating is the dependent measure assessing the change of preference after the learning phase. Even though EC effects appear to be captured in a more sensitive way by direct evaluative measures (e.g., Likert scales and slider scales) compared to indirect measures (e.g., reaction time tasks such as the IAT; Hofmann et al., 2010), the previously mentioned study by Verwijmeren found an effect for subliminally presented CSs only with a choice measure and not with an evaluative measure. Similar, an effect with choice as the dependent variable was also found in a paradigm similar to EC, where CSs were shown briefly and were predictive of a monetary reward/loss in a Go/No-Go task (Pine, Mendelsohn, & Dudai, 2014). The reward or loss could be considered a US in this study and after multiple learning trials, participants were given a 2-AFC task with all CS combinations. Results showed that CSs that were predictive of a reward were chosen over CSs predictive of a loss. Additionally, Jones et al. (2010) argue that in an explicit evaluation participants might overthink their response and the changed attitude—through implicit attribution—might not be acted upon. They therefore also suggest using a forced-choice dependent variable which might pick up the associative changes because participants simply have to choose one CS over another CS and not contemplate over one CS and possibly rely on other features of the CS (Jones et al., 2009). It therefore seems plausible that a small and subtle EC effect might not be reflected in evaluative ratings because they were not perceived as large enough to justify selection of a different point on, say, a 7-point scale—but may nevertheless tip the scale when being forced to choose between two CSs that were paired either with a positive or a negative US.

One additional argument comes from consumer sciences where one tool to measure spontaneous reactions is the swift-selection platform (Salcher, 1995). In this paradigm, several products are hidden behind a curtain on a platform and after the curtain opened for a few seconds, participants are asked to choose a product. The idea behind this technique is that consumers might not be able to verbalize the reasons for a product preference and an impulsive consumer choice might be based on different processes (Salcher, 1995). One could speculate that, if unconscious associations between a CS and US develop in a subliminal learning paradigm, participants might not be able to verbalize a preference for a given CS but might choose this CS in a forced choice task similar to the swift-selection platform. These speculations will be tested in the studies in Chapter 4 and 6. Regardless of the sensitivity of the measure of choice, it is interesting whether a choice measure can detect changes of preference after an EC learning phase as it constitutes an ecologically valid measure of preferences and might therefore link research on evaluative conditioning and consumer science.

The present research

So far I have outlined that evaluative conditioning appears to be a true effect and two different models are currently discussed that attempt to explain the mechanisms underlying EC. While a series of studies—using subliminal stimuli—showed support for dual-process models, recent studies have challenged these findings and methodological inaccuracies might explain why EC effects were found in previous studies. In a recent literature review on evidence for and against propositional single-process models, De Houwer states that "the idea that all associative learning is propositionally mediated should, however, not be dismissed too early" (De Houwer, 2014c, p. 22). As Lovibond and Shanks (2002) postulate that an EC effect without contingency awareness would be theoretically most interesting in differentiating between single- and dual-process models, I will report a series of studies testing the propositional nature of evaluative conditioning by attempting to show EC effects with a subliminal presentation schedule in the next chapters. The main question of this dissertation is therefore whether EC effects can be found in the absence of contingency awareness or whether contingency awareness is necessary for EC effects to form.

In the next chapter, a replication of the study by Rydell et al. (2006) will be presented. We chose a study reporting a subliminal EC effect that (1) shows strong statistical support for subliminal EC effects (see *p*-curve analysis), (2) appears to be replicable as two additional papers were published using this paradigm, and (3) is considered to be a methodologically sound investigation by the peer group (Dedonder et al., 2014; Sweldens et al., 2014).

Chapter 2: Of two minds or one? A registered replication of Rydell et al. (2006)

Additional materials are stored in an online repository: osf.io/c57sr

We constantly rely on preferences to make evaluations and choices in our daily life: When browsing through online shopping sites, watching political debates, or when forming impressions of other persons. One way of acquiring preferences is through evaluative conditioning (EC). An EC effect is defined as a change in liking that is due to the pairing of a negative or positive stimulus (unconditioned stimulus, US) with a target object (conditioned stimulus, CS; De Houwer, 2007).

Hofmann et al. (2010) conducted a meta-analysis on EC and concluded that, "beyond any doubt[,] EC is a genuine phenomenon" (*p.* 410). Yet, the mechanisms that enable evaluative conditioning are not fully understood. Considerable progress has been made in this area, but there is still disagreement about the mental processes underlying evaluative conditioning (Sweldens et al., 2014). The current debate is between single- and dual-process views. Single-process views claim that there is only one path that leads to preference acquisition (De Houwer, 2009; Lovibond & Shanks, 2002; e.g., Mitchell et al., 2009). According to single-process propositional views, evaluative conditioning results from the conscious formation of a proposition about the relation between a CS-US pair. Consequently, propositional views predict that awareness of the contingency between the CS and US would be necessary to establish an EC effect. Dual-process theories propose two distinct systems that may lead to preference acquisition: In addition to a conscious, propositional system, a second associative system is proposed in which a link between the stimuli can be formed automatically without contingency awareness. Thus, according to dual-process models, an EC effect could result from an associative system even when participants are unaware of the CS-US link. Influential examples for such accounts are the associative-propositional evaluation model (APE; Gawronski & Bodenhausen, 2006) and the systems of reasoning account (Rydell & McConnell, 2006). In addition to the distinction between propositional and associative systems, especially the latter model proposes a distinction between implicit and explicit attitudes, such that associations due to CS-US pairings are typically reflected in implicit evaluations, whereas propositions about CS-US relations typically form the basis of explicit attitudes (Gawronski & Bodenhausen, 2011; Rydell & McConnell, 2006).

One testable difference between dual- and single-process theories is therefore the question whether contingency awareness is a necessary condition for an EC effect. One way to experimentally assess the effect of contingency awareness on EC is via *subliminal* or suboptimal (i.e., brief and masked) presentation (Dehaene et al., 2006): If USs (or CSs) are presented suboptimally, they often do not enter conscious awareness, and the conscious formation of a proposition about the CS-US relationship is not possible (Dehaene et al., 2006). Following dual-process models, it should nevertheless be possible to form an unconscious association between CS and US through repeated pairing, because awareness of the relationship between stimuli is not needed for the formation of associations. These pairings would ultimately result in EC, which one should be able to detect with implicit evaluation measures (Gawronski & Bodenhausen, 2011). Finding EC with a subliminal conditioning procedure would be strong evidence for dual-process accounts.

Research on subliminal EC has so far yielded mixed results: On the one hand, there are several studies that found subliminal EC (e.g. De Houwer et al., 1994, 1997; Field & Moore, 2005; Fulcher & Hammerl, 2001). Most of these studies, however had methodological issues (see Sweldens et al., 2014 for an overview). On the other hand, numerous studies support the claim that awareness is necessary for EC (Dedonder et al., 2014; Pleyers et al., 2007; Stahl et al., 2016, 2009). Most of these studies, regardless of the results, have been criticized or disregarded by researchers in the field.

One study, however, is considered particularly convincing evidence for dual-process accounts (Sweldens et al., 2014). In this study, Rydell et al. (2006) showed a selective influence of briefly presented USs on implicit attitudes. In this experiment, the CS was a picture of a character named *Bob*, which was preceded by a briefly presented prime and followed by explicit, behavioral information about *Bob*. Prime and explicit information had opposite valence. The key finding of the study is that explicit attitudes were affected by behavioral information, whereas implicit attitudes reflected the valence of the briefly presented prime. Rydell et al. (2006) conclude that their participants held opposite implicit and explicit attitudes at the same time reflecting on the two types of information received. A prime-recognition check at the end of the study confirmed that the primes were not recognized above chance and Rydell et al. (2006) therefore concluded that the primes were indeed presented subliminally.

The finding is clearly interesting and theoretically relevant. The original study was cited more than 200 times and is generally viewed as methodologically strong (i.e., has fewer methodical issues than other studies claiming to have found subliminal EC; Dedonder et al., 2014). On the other hand, it is also based on a complex interaction obtained from a relatively small sample, inviting questions about its reliability. Furthermore, the study has not been independently replicated, as Sweldens et al. (2014) point out in their review (but see two similar studies by the original authors; Rydell & McConnell, 2006; Rydell et al., 2008). Considering the important contribution to the discussion about EC without contingency awareness and dual-process theories in general (Gawronski & Bodenhausen, 2011), an independent replication would increase the confidence in the results, and thereby, the evidence for EC without contingency awareness.

In the present study, we conducted two replication attempts to investigate whether implicit attitudes can indeed form on the basis of briefly presented information (independently to contradicting explicit attitudes). Experiment 1 is a pre-registered replication using a German sample (see osf.io/cm2sj for the registration of Experiment 1 and osf.io/c57sr for the analysis scripts and material). Experiment 2 is a second pre-registered replication that aims at overcoming some of the potential limitations of Experiment 1: Specifically, to avoid potential cultural or language differences, we used the original material and sampled participants from a similar population as the original study. In addition, we pretested the prime presentation duration to ensure that the primes were indeed presented briefly enough so that they were not remembered at above-chance levels in the awareness check after the learning phase. The goal of both studies was to replicate the original study as closely as possible, while making sure that the experimental procedure would have the same psychological effects as the original procedure, and thus, would yield theoretically relevant results that allow for a valid test of the underlying theoretical claims by Rydell et al. (2006). On a continuum between exact/direct versus conceptual replications, therefore, Experiment 1 may be considered a conceptual replication (only) in the sense that the original material was translated into German (to make sure our German participants understood its meaning); Experiment 2 may be considered a conceptual replication (only) in the sense that the original prime presentation duration was adapted to ensure chance-level prime awareness as reported in the original study.

Experiment 1

The goal of the first experiment was to replicate the study of Rydell et al. (2006) as closely as possible. We contacted the first author and received the original material of the study. All instructions and the material was translated into German to conduct the study at the University of Cologne in Germany.

Method.

Design.

As in the original study, we realized a 2 (valence condition: half of the participants received negative primes and positive behavior in the first block and positive primes and negative behavior in the second block, while for the other half the blocks were the other way around) × 2 (times of measurement: after block 1 and after block 2) × 2 (dependent variable: implicit or explicit attitude) design. The last two factors were manipulated within participants, the first factor between participants.

Procedure.

For efficiency reasons, we decided to conduct the experiment solely at the computer and therefore computerized the measures that had been administered as paper-pencil tasks in the original study. We used PsychoPy as the experimental software (Peirce, 2007). As in Rydell et al. (2006), we conducted a modified version of an attitude-learning paradigm developed by Kerpelman and Himmelfarb (1971).

Participants were seated at a computer and instructed that they would be learning something about a person (called either Bob, Jan, Tim, Tom or Ben). Only the name Bob was used in the original study. While Bob may be a common American name, it is not very common in German, so we included the above mentioned additional names that are more common in Germany. One of those five names was randomly assigned to the target character for each participant anew.

Each trial started with a blank screen for 1000 ms. Afterwards a fixation cross was displayed in the middle of the screen for 200 ms, which was replaced by either a positive or negative prime word in white letters on a black background, which was presented for 23.5 ms. The prime was presented for 23.5 ms instead of 25 ms (as in the original study) due to a slight difference in CRT monitor parameters (i.e., 85 Hz instead of presumably 80 Hz in the original study). We tried to mimic the exact prime presentation conditions as in the original study, based on the information given by the first author of the original study.

All in all we used 10 positive and 10 negative primes. Each prime was presented 10 times, resulting in 200 trials altogether. Half of the participants received 10×10 positive primes in the first block and 10×10 negative primes in the second block. The other half received negative primes first and positives primes in the second block. The order of the 100 primes within each block was randomized for each participant anew. The prime was immediately followed by an image of the target person, which was shown without any additional information for 247 ms.

After the target person was shown alone, negative or positive behavioral descriptions -written in white letters- appeared underneath the image, until the participant pressed either the “c” key or the “u” key to indicate whether he/she thought that the behavior was characteristic or uncharacteristic for the target person. After the judgment, the participant received feedback for 5000 ms about whether their judgment was correct or incorrect (i.e., the corresponding German word was shown on the screen). In the first block of the learning phase, 50 positive behaviors and 50 negative behaviors were randomly drawn out of a list of 100 positive and 100 negative behaviors respectively, and presented in random order (determined for each participant anew). In the second block, the remaining 50 positive and 50 negative behaviors were presented in random order. The valence of the primes was always opposite to the valence of the characteristic behaviors (i.e., when a negative prime was shown, the negative behavior was uncharacteristic and the positive behavior characteristic of the target character).

After the first block, participants performed the first set of attitude measures, consisting of explicit evaluations of the target person and an IAT. We used the same measures as in the original study (but in a computerized version). The order of implicit and explicit measures was counterbalanced across participants, with the same order of measures after the first and second block. In the explicit attitude measure, participants were asked to judge how likable they found the target person, on a 9-point scale, ranging from 1 (very unlikable) to 9 (very likable). After the participant indicated his/her response, five 9-point semantic differential scales (Bad-Good, Mean-Pleasant, Disagreeable-Agreeable, Uncaring-Caring, Cruel-Kind) were presented together on a computer screen and participants were asked to judge the target person. Subsequently the participants were requested to evaluate the target person on a feeling thermometer ranging from 0° to 100°. All scores were z-standardized and a mean explicit rating score was calculated.

To assess implicit attitudes the participants may hold toward the target person, an implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998) was administered. As in Rydell et al. (2006), we used a modified IAT version reported by McConnell and Leibold (2001) consisting of seven 20-trial blocks. Participants had to press the left key ("d") or right key ("k") as part of a classification task. In the first block, images of the target character as well as images of other persons had to be categorized as depicting the target character or other persons (counterbalanced over participants whether the target character was on the left or the right key). The target character name and "not" plus the target character name were displayed on the left or right bottom of the screen as category labels. As in the original study we selected one image—out of the six images in the original material at random, for each participant anew, to represent the target character (in the learning phase and in the IAT). The remaining five other images were then used in the IAT to represent other individuals that were not the target character. In the second block, positive and negative words had to be categorized as positive or negative (with the position of the corresponding key for positive/negative counterbalanced across participants). Now the labels "positive" and "negative" were displayed on the left and right bottom of the screen. In the third and fourth block, the first and second block were combined: Half of the participants had the target character together with negative words on one key and all other images and positive words on the other key (Combination 1), while the other half of participants classified the target character together with positive words on one key and all other images and negative words on the other key (Combination 2). The category labels from block 1 and block 2 were combined in this block (i.e., Bob + positive). In the fifth block, a classification of the target character vs. other persons was again performed, but the location of the key for classifying the target character (vs. other persons) was switched compared to block 1 and block 3/4. In the final two blocks (i.e., blocks six and seven), participants again worked on a combined classification task (i.e., a combination of blocks 2 and 5): Participants who worked on Combination 1 in blocks 3 and 4 now worked on Combination 2 (and vice versa). The order of trials in the critical blocks 3/4 and 6/7 was randomized for each participant anew, controlling that there were no more than 5 repetition trials (same key pressed as in previous trial) and controlling that there were approximately the same amount of response-switch (switching between the d and k key) and response-repetition trials.

We processed the reaction times (RT) of the critical blocks 3/4 and 6/7 in the IAT as described in the original study (Rydell et al., 2006). All trials were used and responses faster than 300 ms recoded to 300 ms and responses slower than 3000 ms recoded to 3000 ms. Reaction times were transformed to reduce skewness (natural log) and a mean log reaction time was calculated for each participant in each critical block. An IAT effect was calculated subtracting the mean log RT of the block in which the target character and positive words (Combination 2) were presented from the mean log RT of the block in which the target character and negative words (Combination 1) were presented. The IAT effect was then z-standardized (across both IAT measures and all participants). Therefore a larger IAT effect indicated a more positive evaluation of the target character.

After the second experimental block and the second set of measures, a recognition test of the briefly presented primes was administered: 40 words were presented in a random order on the computer screen simultaneously (20 of them were the previously presented primes, the other 20 were words that had not been shown in any task previously). Participants were asked to select those 20 words which they thought had been shown as primes during the learning phase. At the end of the study, we asked demographic questions (Age, Study/Job, Gender, Suspected goal of study, Comments).

Materials.

Two individuals translated the instructions and materials independently and a third person adjusted the translations, if necessary. The original material we received contained eight positively and eight negatively valenced words used in the IAT, but the article referred to ten positive and ten negative words.[3] We therefore added the four words “successful” (erfolgreich), “talented” (begabt), “hectic” (hektisch) and “disgusting” (widerlich) to complete the required ten items for each category.[4] We also noticed that some valenced IAT targets were quite similar to the briefly presented primes during the learning phase. The word „laughter“ (Lachen), which translated into German appears very similar to the prime word „smile“ (Lächeln) and the word „friend“ (Freund), which was used as a prime, were replaced in the IAT with „trust“

[3]After the data of Experiment 1 was collected we were informed that a different set of target words was used in the original study.

[4]The first three words were taken from a study by Bluemke and Friese (2006), while the last word was explicitly mentioned in the original article by Rydell et al. (2006).

(Vertrauen) and „humor“ (Humor).[5]

In our experiment we addressed a potential issue with the original prime-recognition check: The words “help“, “candy“, “crash[ed]“ and “gun“ occurred as new words in the recognition task but also as behaviors that describe the character “Bob.”[6] Participants might have selected those words because they remembered that they saw those words before, without remembering their source (i.e., whether they were primes or words in a different task, such as the IAT, or presented as part of the sentences describing Bob). If those words were selected frequently it might have resulted in an underestimation of participants‘ knowledge of the primes. The conclusion that participants were not able to recognize the primes during learning must therefore be interpreted with caution.

To eliminate potential issues regarding these words, we substituted the word “help” (Hilfe) by the word “beautiful” (schön), “crash” was replaced by “fear” (Angst) and “gun” was replaced by “pistol” (Pistole) in the recognition task. “Candy” was replaced in the behaviors with “magazines” (Zeitschriften).

In a small pilot study (N = 16) each valent word was rated on a 101 point likert scale, ranging from -50 to 50. The results confirmed that all translated primes (M_{pos} = 78.17; M_{neg} = 21.34), new words in the prime-recognition task (M_{pos} = 74.04; M_{neg} = 22.14), and words used in the IAT (M_{pos} = 83.54; M_{neg} = 21.31) were appropriately positive (or negative) in the German language.

As Rydell et al. (2006), we also made sure that primes and new words in the recognition task were on average of approximately equal length and equal frequency in the German language (Primes: M_{length} = 6.1, $M_{frequency}$ = 12460.35; New words: M_{length} = 7.2, $M_{frequency}$ = 11778.9).

Changes compared to original Experiment.

In addition to computerizing all parts of the study and translating the material (thereby adding multiple different German names for the target character) we resolved minor inconsistencies between the paper and the material we received: (1) We received

[5] Neither “laughter” nor “friend” were used as targets in the original Experiment, but were included in the material we initially received.

[6] Additionally the words “Love” and “Peace“ were used as new words in the recognition test, but were also previously used as valent words in the IAT after block 1 and block 2. To avoid potential problems in the prime recognition check, we substituted the words”love" (Liebe) and “peace” (Frieden) with “nature” (Natur) and “cash” (Bargeld) in the prime-recognition task.

16 IAT attribute words (8 positive, 8 negative) and added 4 words (2 positive, 2 negative) for a total of 20 as mentioned in the method section of the original study. Of the positive IAT attribute words, two were substituted by new words because one was very similar to a prime from the learning phase after the translation to German and another was used as a prime in the learning phase. (2) Beyond the translation and resolving the inconsistency between the paper and the material received, the only adjustment was that we replaced distracter words from the recognition test that had in fact been part of the IAT and/or behavioral descriptions (and were therefore not new to participants) with truly new words, so as to obtain a valid estimation of prime recognition. As the memory measure was administered in the very end of the experiment, we can be confident that this change cannot influence the main pattern of results obtained in the IAT and evaluative ratings.

Hypotheses.

Explicit evaluation.

Under both propositional single- as well as dual-process theories we would expect a more favorable explicit rating of the target stimulus (i.e., Bob) when positive behavior was characteristic of him than when negative behavior was characteristic of him (Rydell et al., 2006). We thus expect a significant difference between the results of the explicit attitude measure at time 1 compared to time 2, in the direction described above. Because half of the participants received positive behavioral information in the first block and negative information in the second block, whereas the order was reversed for the other half, we expect a 2 (time of measurement) $\times$ 2 (valence condition: negative primes and positive behavior first vs. positive primes and negative behavior first) interaction.

IAT effect.

Based on the findings by Rydell et al. (2006), and as predicted by a dual-process account, we would expect that the implicit measure contradicts the explicit measure in each experimental block. Specifically, we would expect an IAT effect to reflect the valence of primes (and opposite to the behavior information), resulting in a significant difference between the results of the implicit attitude measure at time 1 compared to time 2.

A single-process view, in contrast, would predict that a subliminal prime has no influence on the implicit measure. We would therefore expect that the IAT reflects the supraliminal behavioral information about the target character as does the explicit rating.

Both of these opposing patterns would be reflected – similar to the explicit evaluation – in a 2 × 2 interaction of time of measurement and the valence condition (negative primes and positive behavior first vs. positive primes and negative behavior first).

Analysis plan. The data were analyzed in two ways. First, the same frequentist analyses were conducted as in the original article. They were based on the initial fixed sample size for which the underlying sampling assumptions hold. Second, we planned to continue sampling, while sequentially computing Bayes factors, until the relative evidence was strong enough for either the null or alternative hypotheses. For the frequentist analysis we used $\alpha = .05$ as inference criterion. In all Bayes Factor ANOVAs we used the following method to estimate whether a given factor or interaction should be interpreted as having an influence on the dependent variable: The full model with all effects (i.e., main effects and interactions) was compared to a model without the effect of interest (i.e., a main effect or an interaction). A Bayes factor for the alternative hypothesis over the null hypothesis (i.e., $BF_{10} > 1$) for a given effect indicates that the full model (that includes the predictor term corresponding to the effect) is favored over the model without the effect given the data (i.e., the predictor term improved the model, suggesting that the effect had an influence on the dependent variable). We follow Rouder and Morey (2012) and interpret the Bayes factor as the evidence for one model (in this case the full model) *relative* to a competing model (in this case the restricted model).

In model comparison using Bayes factors, prior distributions on target parameters represent the hypotheses to be tested. We chose weakly informative priors for the analysis here, as recommended by Rouder, Morey, Speckman, and Province (2012). Previous EC studies on contingency awareness using the IAT found small to medium effect sizes (Rydell et al., 2006: $\eta^2 = .1$; Olson & Fazio, 2001: $d = 0.37$). We used these previous findings as guideline for our expectation for this study and chose our priors accordingly. For our ANOVA analyses we therefore used default multivariate Cauchy priors on the effect size parameters (Rouder et al., 2012), with a scaling parameter of r = 0.5. For all Bayesian t tests Cauchy priors with a scaling parameter of $r = \sqrt{2}/2$ were used (Rouder et al., 2012; Rouder, Speckman, Sun, Morey, & Iverson, 2009), which expresses our expectation to find small to medium effects. When informative prior distributions are used, the question can be asked how the chosen priors affect the results. To test the robustness of the results to changes of the prior we analyzed the data with other priors around the chosen scaling factor. We used smaller and larger prior scaling factors of 0.3 and $\sqrt{2}/2$, respectively, for the ANOVA. The pattern of Bayes factors did not change substantially (all BF that were >1000 were still >1000 and all BF > 10 were still larger 10). We additionally calculated all t tests with prior scales of 0.5 and 1: all Bayes Factors originally reported to be larger 1000 were still larger than 1000, all Bayes Factors that were initially larger than 10 were still larger than 10 when changing the prior width; and all Bayes Factors that were smaller than 10 remained smaller than 10.[7]

We additionally report the median and the 95% highest density interval (HDI) of the posterior distribution of the estimated effect size d. The 95% HDI contains the true population effect size with a probability of 95%. We considered a Bayes factor > 10 or $< 1/10$ to be convincing evidence for the alternative hypothesis relative to the null, or for the null hypothesis relative to the alternative, respectively. We use BF_{10} to indicate a Bayes factor for the alternative Hypothesis and BF_{01} as a Bayes factor for the null hypothesis (dividing 1 by the calculated Bayes factor). BF_{10} or BF_{01} smaller than 3 will only be considered to be anecdotal relative evidence (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2015). We performed all analyses in R (Version 3.4.4; R Core Team, 2016) and the R-packages *afex* (Version 0.20.2; Singmann, Bolker, Westfall, & Aust, 2016), *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2015), *dplyr* (Version 0.7.5; Wickham & Francois, 2016), and *papaja* (Version 0.1.0.9735; Aust & Barth, 2016).

Sample size rationale.

For the initial frequentist analyses, we planned to collect the same number of participants as in the original study ($N = 50$), even though an a priori power analysis with the smallest interaction effect reported in the original study ($f = 0.32$) and $\alpha = \beta = .05$ yielded a required sample of 34 participants. Since we did not want to sample fewer participants than in the original study, an experiment with $N = 50$, would allow us to detect an interaction effect of $f = 0.26$ or larger (with $\alpha = \beta = .05$). Starting after reaching this initial sample size (thereby lowering the possibility of a false positive finding in the Bayesian analyses; Schönbrodt et al., 2015), we planned

[7]See table robustnessBF.pdf on osf.io/c57sr for a detailed overview of all Bayes Factors with different prior scaling factors.

to continue the data collection and computed Bayes factors. We planned to stop data collection once we had gathered sufficient information (BF_{10} or BF_{01} larger 10) regarding our hypotheses of main interest (namely the two-way interactions, for both dependent measures, of time of measurement × order of information received) or once we had reached a maximum of 125 participants (which is 2.5 times the original sample size as recommended by Simonsohn, 2015 for replication studies).

Particicpants.

We initially collected the data of 53 participants and - based on our Bayesian analyses - did not collect any additional data. One participant aborted the experiment and one data file was corrupted; both of these events were known during data collection, and these missing participants were replaced by 2 additional participants. We used the data of 51 participants in the final analysis.[8]

Confirmatory analysis. We first report the set of planned and pre-registered analyses to test our hypotheses, before we turn to additional exploratory analyses.

Recognition test.

A one-tailed t test indicated that the primes were recognized above chance in the recognition task at the end of the study, $M = 12.53$, with chance level at 10, $t(50) = 5.22$, $p < .001$; $BF_{10} = 10,227.03$, $d = 0.70$, 95% HDI [0.40, 1.01].

Explicit and implicit evaluation.

As in the analysis by Rydell et al. (2006) we first ran a 2 × 2 × 2 ANOVA with the type of the dependent variable (IAT, Explicit), the time of measurement (time 1, time 2) and the valence condition (negative primes and positive behavior first vs. positive primes and negative behavior first) with the first two factors manipulated within participants. As in the original study, we found statistical evidence for a three way interaction of the variables mentioned above, $F(1, 49) = 49.16$, $MSE = 0.36$, $p < .001$, $\hat{\eta}^2_G = .149$, $BF_{10} = \mathit{1.14} \times 10^7$.

Explicit evaluation.

[8] In addition, during the final day of data collection, two participants instead of only one were collected accidentally, resulting in 51 (instead of the planned 50) complete data sets after testing a total of 53 participants. Results do not change when we excluded the 51st participant (largest deviation of $d = 0.05$; largest deviation of $\eta^2_G = .005$). Two participants had to rate the target character also before the first learning phase, due to a programming error. Excluding these participants did not substantially alter the results (largest deviation of $d = 0.07$; largest deviation of $\eta^2_G = .01$).

We subsequently tested for a 2 (time of measurement) $\times$ 2 (valence condition) ANOVA in the explicit rating data only. As in the original study, we indeed found a significant interaction of the time of measurement and whether positive behavior was presented first and negative second or vice versa, $F(1, 49) = 274.76$, $MSE = 0.26$, $p < .001$, $\hat{\eta}^2_G = .757$, $BF_{10} = \mathit{4.07} \times 10^{30}$.

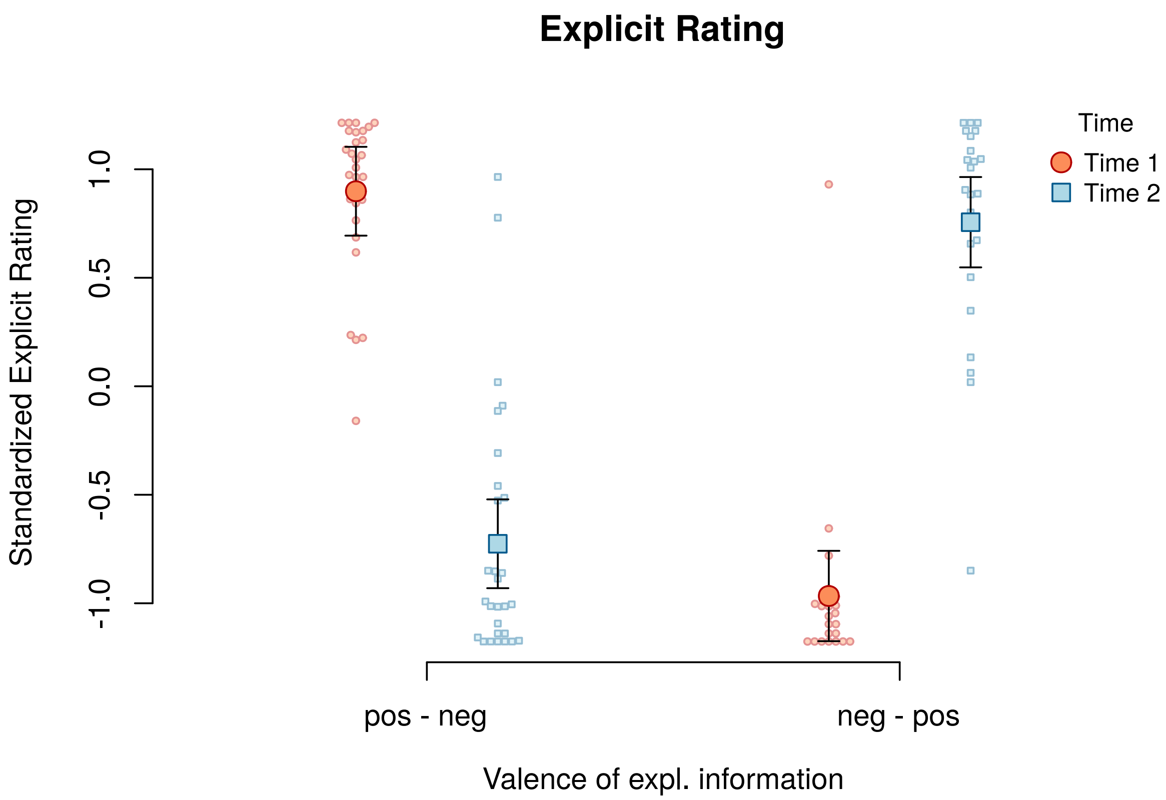

Figure 4. Standardized explicit evaluation effects of Experiment 1, split by the valence of the explicit information and the time of the explicit measurement. Error bars represent 95% within-subject confidence intervals. Pos: positive information, neg: negative information.

As can be seen in Figure 4 and Table C1, the rating of the character was according to the explicit information given to the participants. When given positive information first, the rating of the target character was more positive than when given negative information, $t(27) = 11.52$, $p < .001$; $BF_{10} = \mathit{1.37} \times 10^9$, $d = 2.09$, 95% HDI [1.41, 2.79]. Similar, when negative behavior was learned in the first block, the rating of the target was more negative than when positive information was learned in the second block, $t(22) = 12.14$, $p < .001$; $BF_{10} = \mathit{2.8} \times 10^8$, $d = 2.41$, 95% HDI [1.57, 3.26].

Implicit evaluation.

The analysis of the IAT data also revealed a significant 2-way interaction of time of measurement and condition, $F(1,49) = 14.24$, $MSE = 0.44$, $p < .001$, $\hat{\eta}^2_G = .075$, $BF_{10} = 66.31$.

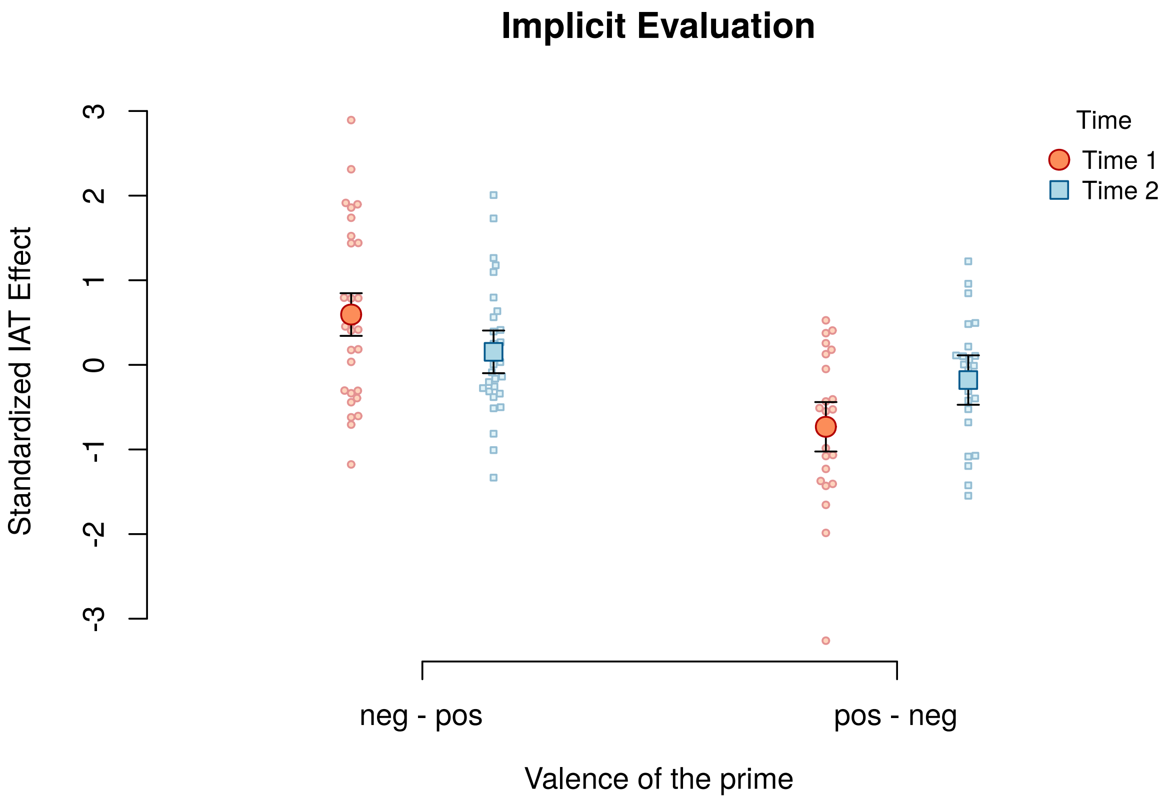

Figure 5. Standardized IAT effects of Experiment 1, split by the valence of the primes and the time of IAT measurement. Error bars represent 95% within-subject confidence intervals. Pos: positive prime, neg: negative prime.

As can be seen in Figure 5 and Table C2, when negative primes were presented first and positive primes afterwards, we can observe some indications that the IAT effect became more negative $t(27) = -2.54$, $p = .017$; $BF_{10} = 2.92$, $d = -0.44$, 95% HDI [-0.82, -0.06]. This pattern is inconsistent with the prime valence but consistent both with the verbal behaviors presented to the participants, as well as the explicit ratings. Similarly, when positive primes were presented first and negative primes afterwards, the IAT effect became more positive $t(22) = -2.77$, $p = .011$; $BF_{10} = 4.50$, $d = -0.52$, 95% HDI [-0.95, -0.10]. This pattern is again consistent with the explicit evaluation and inconsistent with the notion that the IAT effect reflects prime valence.

Exploratory analysis.

Our counterbalancing factors had no influence on the pattern of results in the implicit nor explicit evaluations (but we cannot confidently conclude an absence of such effects because the Bayes factors were mostly not conclusive). Additionally, no correlation between the number of remembered primes and the IAT effect score was found, $r = .15$, 95% CI $[-.13, .41]$, $t(49) = 1.05$, $p = .301$, $BF_{01} = 2.28$.

Composite effect.

We additionally computed a composite IAT effect score for both groups together (i.e., negative primes first and positive primes first) and compared it to zero. We subtracted the mean log IAT effect when the target character was shown with a negative prime from the mean log IAT effect when the character was shown with a positive prime.

Values larger 0 would therefore indicate that the target character's IAT score was in line with the valence of the briefly presented primes; values smaller than 0 would indicate that it was in line with the behavioral information received. We observed a negative value for the composite score (M = -0.08), and a one-sided t test of the hypothesis derived from the results of Rydell and colleagues (2006) that the effect is *larger* than zero was consequently not significant, $t(50) = -3.78$, $p > .999$. The Bayes factor yielded relative evidence for the null hypothesis over the alternative hypothesis, $BF_{01} = 28.58$, $d = -0.50$, 95% HDI [-0.79, -0.21], which clearly indicates that the target character was not implicitly evaluated in line with the valence of the briefly presented primes.

Similarly, we computed a difference score for the explicit rating. We subtracted the z-standardized mean explicit rating when negative behavior was learned from the mean rating when positive behavior was learned. A positive value therefore indicated an explicit rating in line with the behavioral information, while a negative value would indicate an explicit rating in line with the briefly presented primes. Unsurprisingly, a one-sided t test revealed that the target character was rated in line with the behavioral information received, $t(50) = 16.74$, $p < .001$; $BF_{10} = 2.16 \times 10^{19}$, $d = 2.29$, 95% HDI [1.77, 2.83].

Discussion Experiment 1. We replicated the pattern of results of the original study for the explicit evaluation. When participants learned that the target character was a good person, he was evaluated more positively than when participants learned that he was a bad person. Considering that participants had 100 trials to learn about the target character in each block, we would consider this is a trivial learning effect consistent with both single- and dual-process accounts.

Contrary to the original study, we did not replicate the pattern of effects for the implicit measures. While the IAT effects in the original study reflected a change in line with the briefly presented primes, the pattern observed in this experiment was opposite to it (i.e., in line with the explicit information about the target character). We therefore conclude that we did not find evidence that the briefly presented primes had an effect on implicit evaluations.

Limitations of Experiment 1.

There are two aspects that might account for the difference in result patterns between the original study and the present replication: First, the translation of the material into German may have affected the results. For instance, the German prime words were slightly longer (6.05 letters) than the original primes (4.9 letters), so one might speculate that the primes could not be fully processed in the brief presentation time. However, this interpretation does not fit with the prime-recognition result, which showed a higher rate of prime recognition than in the original study. We would therefore conclude that prime processing was not reduced in this replication. However, other language or cultural factors may have moderated the influence of the briefly presented primes.

Second, perhaps more importantly, our results suggest that prime processing was not fully subliminal. As in the study by Rydell et al. (2006), we measured the recognition of the briefly presented primes with a recognition test in the end of the study. Recall that one half of the primes were shown in the first block only, the other half in the second block only. Since participants additionally received no information about the primes, the prime-recognition test in the end of the study is a weak proxy for the actual visibility of the primes during the learning phases (i.e., decay over time and interference from subsequent stimuli may have weakened prime memory). We additionally substituted multiple words from the list of new words used in the original study with words that were not previously used (see above). We found – contrary to the original study – that the primes were recognized above chance. We do not know whether this difference is purely due to a more strict measure of memory (i.e., due to our using truly new words as distracters in the recognition list given to the participants, instead of including words that were also used in the behavioral statements), or due to a higher visibility of the primes. Note that we aimed at closely replicating the presentation conditions of the original study (black background, white letters, font type and size), used the same images as backward masks, and even realized a slightly shorter prime presentation time compared to the original study (i.e., 23.5 ms instead of 25 ms). Together with the slightly increased word length, it could be argued that, if anything, it should have been more difficult to perceive the briefly presented primes. However, in contrast to the original study, the result of the prime-recognition check suggested that the primes were not subliminal, but were supraliminally processed (at least in part).

Note that prime presentation may also have been supraliminal in the original study: The differences in prime-recognition performance may be explained by the improvement made to the prime recognition test (i.e., by using truly new words). However, we cannot rule out the possibility that the primes were too easy to read in the present Experiment 1, and were therefore processed in a propositional fashion instead of the automatic associative process. Following that logic, it would not be surprising that we did not observe an effect of the primes on the IAT, considering both primes and verbal information were supraliminal: The clearly visible behavioral information participants learned about the target person, should have a much larger effect on preferences than a briefly presented word with no clear relation to the target person. In order to allow for automatic associative effects on implicit attitudes, it might be necessary to ensure that prime processing is truly subliminal in the sense that prime recognition is at chance as reported in the original study. Keeping the original presentation duration would therefore lead to functionally different presentation conditions (i.e., above-chance awareness as opposed to chance-level awareness in the original study), and this would render the results of the replication study all but uninterpretable. In order to allow for a valid interpretation of the results of our replication attempt, we therefore aimed to realize presentation conditions that ensured chance-level awareness of the primes (see Stroebe & Strack, 2014).

Experiment 2

The goal of the second experiment was to conduct a second replication of the original study, addressing the potential limitations outlined above. We used the same procedure as in Experiment 1, but used the original materials and a sample of participants from the same population of US undergraduate students. Primes in

Experiment 2 were presented for the longest possible duration for which prime memory (measured as in the original study) did not exceed chance level in a pretest.

Pretest. Participants completed a single learning block with 10 $\times$ 10 positive (negative) primes, receiving the same instructions as in Experiment 2. They were then asked about their explicit evaluations of the target person (this was done to erase any prime knowledge from working memory) before completing an adapted version of the prime-recognition task that contained only the positive (negative) words. We successively tested prime presentation durations starting with 23.5 ms (continuing with 20 ms, 17 ms, 13 ms etc.), until the primes could no longer be identified above chance. We used sequential Bayesian analysis (with the same priors and specifications as in Experiment 1, starting each analysis after 10 participants) and stopped data collection for any given presentation time as soon as the Bayes Factor for the test of prime memory performance against chance level has reached $BF_{10} > 5$ or $BF_{01} > 5$ (with a maximum N of 25 participants per condition). We decided a priory to use the longest presentation time for which prime identification did not exceed chance level (i.e., until we obtain a Bayes factor in support of the null hypothesis).[9]

With a prime presentation time of 23.5 ms, the primes were visible above chance, $BF_{10} = 5.41$, $d = 0.48$, 95% HDI [0.05, 0.91], $N = 22$. This also tended to be the case with a presentation time of 20 ms, $BF_{10} = 2.16$, $d = 0.36$, 95% HDI [-0.04, 0.76], $N =$ 24. A test of a prime presentation duration of 17 ms yielded convincing evidence that prime stimuli were remembered at chance level, $BF_{01} = 5.78$, $d = -0.08$, 95% HDI [-0.48, 0.32], $N = 21$.

Procedure. We used the same experimental scripts as in Experiment 1, but with the original English instructions and stimulus material (including only the name Bob for the target character). In contrast to Experiment 1, we used the same IAT target words as in the original study. As in Experiment 1, we used new words (i.e., words that were not used during the learning phase) for the recognition task, to ensure

[9] The data of the pretest and the main study were collected under a Born Open Data protocol (Rouder, 2016) in which they were automatically logged, uploaded, and made freely available after every day of data collection (github.com/methexp/rawdata/tree/master/croco4, github.com/methexp/rawdata/tree/master/croco4b, github.com/methexp/rawdata/tree/master/croco5).

a more exact approximation of the visibility of the primes.[10] Most importantly, we changed the presentation time of the prime to 17 ms.

Analysis plan. As in Experiment 1, we conducted frequentist and Bayesian analyses after a fixed initial sample size and used sequential Bayesian testing for all further analyses. We used the same annotation, inference criteria, and priors as in Experiment 1.

Sample size rationale.

As in Experiment 1, we initially collected the data of 50 participants. We additionally decided to continue recruiting participants using sequential Bayesian analysis (analyzing the data after every day of data collection) until the two Bayes factors of interest (see below) were larger than 10 (or smaller than 1/10), or until a maximum of 125 participants was reached. No new participants would be recruited after reaching this criterion, but those already signed up at this point were allowed to take part in the study. We decided to exclude participants who (a) aborted the experiment or (b) experienced or reported unforeseeable circumstances (e.g., noises, difficulty with the instructions; based on our experience we did not expect any of those cases to occur).

Data collection stopping rule and confirmatory Analysis.

As in the exploratory analyses of Experiment 1 we planned to calculate a composite effect both for the IAT and the explicit evaluation. We planned to run one-sided t tests and compare the score to 0 (see above) to determine whether (a) the explicit evaluation is in line with the behavioral information and (b) whether the IAT effects are in line with the prime valence. We planned to stop data collection once we have gathered sufficient information on these two contrasts (i.e., $BF_{10} > 10$ or $BF_{01} > 10$).

Additional analyses.

We planned to additionally conduct the same confirmatory analyses as in Experiment 1, but without using these as stopping rules for the data collection.

[10] In the original study, some of the prime words, as well as some of the new words in the prime recognition task, also appeared as part of the behavioral statements during the learning phase. We slightly adjusted the behavioral statements (e.g., substituted single words in the behavioral description with similar but different words or rearranged the sentence and omitted the word in question) to remove this possible confound. Two independent raters confirmed that the behavioral statements were still clearly identifiable as positive or negative.

Participants.

We collected the data of 57 participants and—based on our Bayesian analyses—did not collect any additional participants. We collected more than the planned 50 participants, as we allowed participants who had already signed up to participate as well (see above).

Confirmatory analysis. We will first report on the performance of the participants in the prime recognition test and compare it to the chance level, then analyse the composite effect score testing the two crucial hypotheses (i.e., explicit rating in line with the behavioral information and IAT effect in line with the valence of the primes), and we will additionally report the same confirmatory analyses as in Experiment 1 and the original study.

Recognition test.

A one-tailed t test indicated that the primes were not recognized above chance in the recognition task at the end of the study, M = 10.07, with chance level at 10, $t(56) = 0.46$, $p = .325$; $BF_{01} = 4.66$, $d = 0.06$, 95% HDI [-0.19, 0.31]. This was contrary to the results of the first experiment but in line with the results of the original study.

Stopping rule analysis: Composite effect.

We again computed a composite effect score for the explicit ratings and the IAT data. Values larger than zero in the explicit rating indicated ratings in line with the valence learned from the behavioral statements, values larger than zero from the IAT data indicated IAT effects in line with the valence of the briefly presented primes.

Unsurprisingly, a one-sided t test of the explicit ratings revealed that the target character was rated in line with the behavioral information received, $t(56) = 20.31$, $p < .001$; $BF_{10} = \mathit{3.83} \times 10^{24}$, $d = 2.64$, 95% HDI [2.09, 3.21].

Importantly, as in Experiment 1 we found a negative value for the composite score of the IAT (M = -0.11), indicating that the IAT effect was not in line with the valence of the primes. The pre-registered one-sided t test of the composite score of the IAT data, testing whether the effect is *larger* than zero, was consequently not significant, $t(56) = -4.41$, $p > .999$. The Bayesian analysis also yielded relative evidence for the absence of the predicted effect, $BF_{01} = 33.85$, $d = -0.56$, 95% HDI [-0.83, -0.28].

Overall confirmatory analysis.

As in Experiment 1 and in the original analysis we additionally ran a $2 \times 2 \times 2$ ANOVA with the type of the dependent variable (IAT, Explicit), the time of measurement (after block 1, after block 2) and the valence condition (negative primes and positive behavior first vs. positive primes and negative behavior first) with the first two factors manipulated within participants. We again found statistical evidence for a three way interaction of the variables mentioned above, $F(1, 55) = 54.69$, $MSE = 0.35$, $p < .001$, $\hat{\eta}^2_G = .149$, $BF_{10} = 3.74 \times 10^7$.

Explicit evaluation.

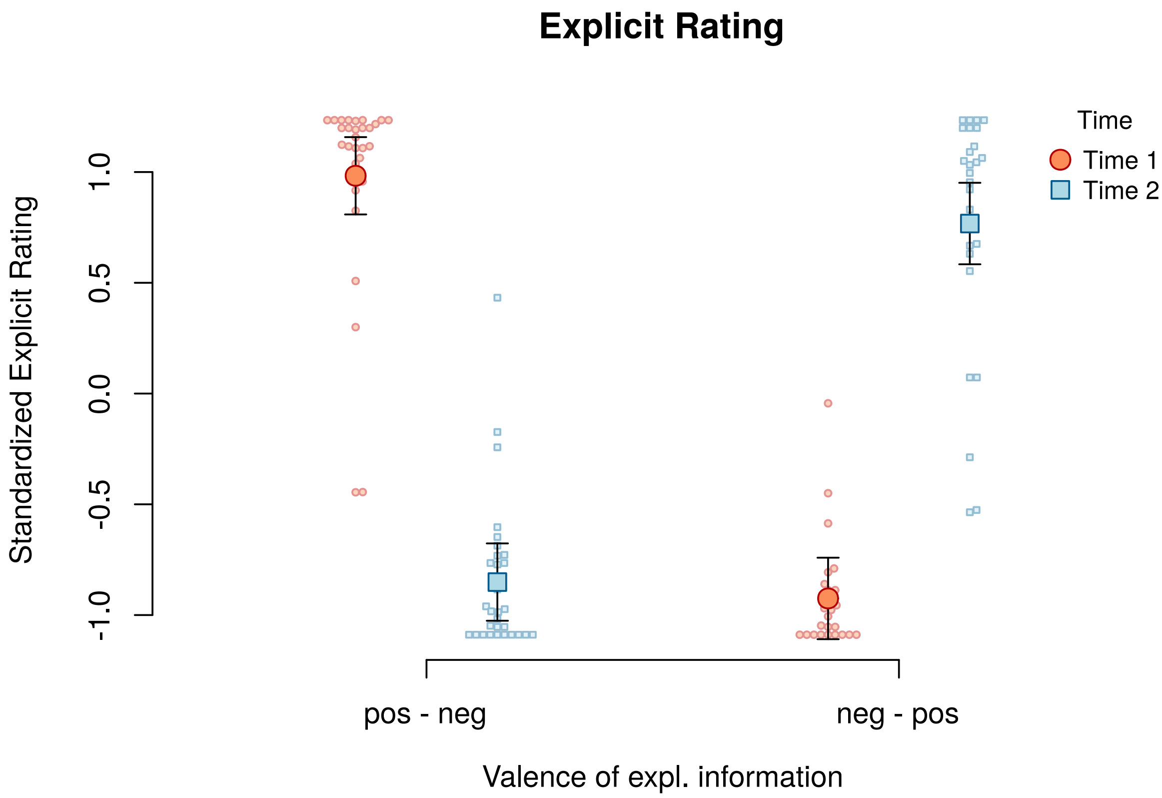

Figure 6. Standardized explicit evaluation effects of Experiment 2, split by the valence of the explicit information and the time of the explicit measurement. Error bars represent 95% within-subject confidence intervals. Pos: positive information, neg: negative information.

We therefore computed a 2 (time of measurement) $\times$ 2 (valence condition) ANOVA of the explicit rating data only. A significant interaction of the time of measurement and whether positive behavior was presented first and negative second or vice versa was found, $F(1, 55) = 403.79$, $MSE = 0.22$, $p < .001$, $\hat{\eta}^2_G = .828$,

$BF_{10} = \mathit{1.83} \times 10^{42}$.

As in the previous studies, the rating of Bob was according to the explicit information given to the participants (see Figure 6 and Table D1). When learning that Bob was a good person in the first block, the rating of the target character was more positive than when learning Bob was a bad person, $t(30) = 15.18$, $p < .001$; $BF_{10} = \mathit{4.66} \times 10^{12}$, $d = 2.63$, 95% HDI [1.87, 3.42]. Similar, when negative behavior was learned about Bob in the first block, the explicit rating of Bob was more negative than when positive information was learned, $t(25) = 13.40$, $p < .001$; $BF_{10} = \mathit{1.16} \times 10^{10}$, $d = 2.52$, 95% HDI [1.70, 3.35].

These results again show a simple learning effect, but also demonstrate that the participants paid attention to the instructions and the information presented.

Implicit evaluation.

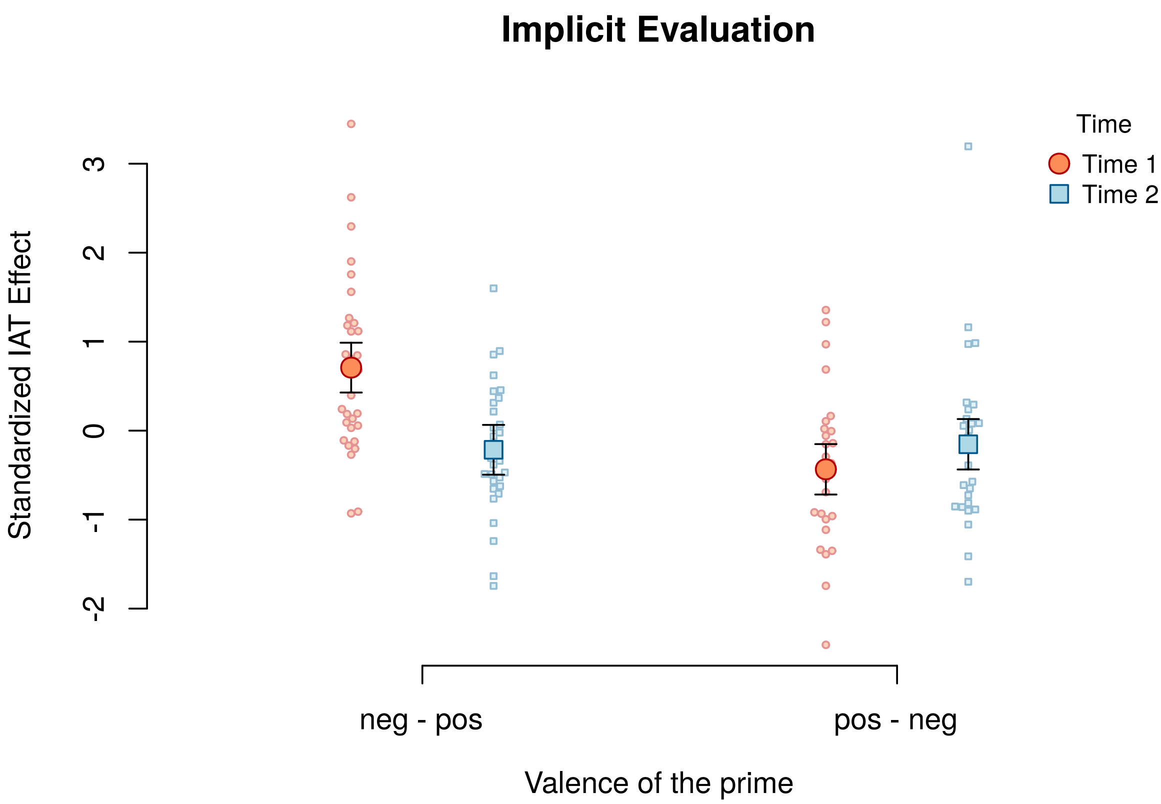

Figure 7. Standardized IAT effects of Experiment 2, split by the valence of the primes and the time of IAT measurement. Error bars represent 95% within-subject confidence intervals. Pos: positive prime, neg: negative prime.

An ANOVA of the IAT data also revealed a significant 2-way interaction of time of measurement and valence condition, $F(1, 55) = 18.96$, $MSE = 0.54$, $p < .001$, $\hat{\eta}^2_G = .102$, $BF_{10} = 535.23$.[11]

As shown in Figure 7 and Table D1, when negative primes were presented first and positive primes afterwards, the IAT effect became more negative $t(30) = -4.77$, $p < .001$; $BF_{10} = 529.65$, $d = -0.80$, 95% HDI [-1.21, -0.39]. This pattern is inconsistent with the prime valence but consistent with the behavior descriptions presented to the participants, with the explicit ratings, as well as with the results of Experiment 1. Even though descriptively true, and contrary to Experiment 1, there was no statistical evidence that the IAT effect became more positive when negative primes were shown after positive primes, $t(25) = -1.45$, $p = .161$; $BF_{01} = 1.91$, $d = -0.25$, 95% HDI [-0.63, 0.12]. One should note, however, that we ran two tailed t tests in these analyses; the direction of the effect was, for both valence order conditions, not in line with the prime valence but instead in line with the valence of the explicitly learned information. Taken together, these analyses—similar to the composite effect analyses— showed that there was no evidence for the pattern described in the original study (i.e., no evidence for IAT effects in line with the valence of the briefly presented primes).

Discussion Experiment 2. In Experiment 2, we again set out to test whether implicit attitudes could be selectively influenced by briefly presented primes, while explicit attitudes were influenced by explicitly learned behavioral information. In contrast to Experiment 1, we used the original instructions and prime words, and collected data from a similar population as in the original study (i.e., US undergraduate students). We also used a shorter presentation time of the briefly presented primes (i.e., 17 ms compared to 25 ms). We speculated that we did not find an effect of the primes in Experiment 1 because they were sometimes visible and therefore may have been processed consciously, thus competing with the explicit statements instead of affecting implicit attitudes. In Experiment 2, we found that primes were not recognized above chance in the prime recognition task. This result is contrary to the above-chance recognition in Experiment 1, but in line with the original results. If the primes in Experiment 1 indeed did not affect the IAT effect because they competed with explicit information, we should be able to detect an effect of the primes on the IAT in Experiment 2, where primes were not processed consciously. In a composite

[11] Additionally, we found no indications for an influence of the order of dependent measures or the key assignment during the IAT on the IAT results (as in Experiment 1 the Bayes factors were mostly not conclusive).

effects analysis, however, we found compelling evidence that the IAT effect was not in line with the valence of the briefly presented primes. All analyses of the explicit rating support the previous findings that the explicit ratings were in line with the valence learned about Bob throughout 100 trials. These findings indicate that participants paid attention to the instructions and stimuli during the learning phase.

One noteworthy observation from the unstandardized and untransformed IAT effects from both Experiment 1 and Experiment 2 (see table C2 and table D2) is that IAT effects appear to generally reflect a positive evaluation of the target character. As one might expect, IAT effects were positive (i.e., greater than 0) when a positive explicit behavior was learned. Interestingly, IAT effects were also greater than zero in the second IAT after negative explicit behavior had been learned in the second block (Valence Condition 1). Importantly, when negative explicit information was learned first, IAT effect scores in the first IAT do not differ from zero (and are greater than zero in the second IAT). Thus, there appears to be an asymmetrical effect of negative versus positive learned behavioral information on the IAT, which should be investigated by future studies and replication attempts.

Given the results of Experiment 2, we were not able to replicate the dissociating between implicit and explicit measures described in the original study, despite close adherence to the original methods and procedures. Compared to the original study, we used the same instructions, IAT target words, and primes. We slightly modified the behavioral descriptions to eliminate overlap between the words used in the descriptions and those used as primes or new words in the recognition task. The most relevant modification to the original study protocol was the reduction of the presentation duration of the briefly presented primes: We presented primes for only 17 ms in order to achieve a functional equivalence in the perception of primes compared to the original experiment. Nevertheless, we were unable to reproduce the implicit attitude findings of the original study and instead found *evidence for the absence* of the effect.

General Discussion

The main goal of this set of experiments was to investigate the potential influence of subliminal stimuli on (implicit) attitudes. In two experiments, we did not observe such an influence. We decided to reduce the visibility of the briefly presented stimuli in Experiment 2 compared to Experiment 1, in order to achieve at-chance prime recognition as described in the original study. However, even with prime recognition at chance level, we were unable to find an effect of the primes on the IAT effect.

A key finding of the original work was that participants held two distinct attitudes at the same time. Replicating this finding—regardless of the actual visibility of the primes—would be an important contribution to the literature. However, we were unable to reproduce this "dual attitude" effect using the same presentation time of prime words as in the original study in Experiment 1, or using a shorter presentation time in Experiment 2.

One noteworthy finding of Experiment 1 and the pilots prior to Experiment 2 was that prime stimuli seemed to be visible when presented for the same amount of time as in the original study. We can only speculate whether the prime recognition tests in our experiments were more sensitive (as we used truly new words versus words that had previously appeared in other parts of the study), or whether differences in the presentation method of the words (e.g., font size, font color) caused prime words to be recognized above chance in our studies with a presentation time of 23.5 ms. Future studies should take into account that merely using the same presentation time as reported in another study does not guarantee that stimuli are perceived in a similar way; ideally, trial-by-trial visibility checks should be used to determine actual visibility (Stahl et al., 2016).

Limitations

In Experiment 2, we reduced the presentation time of the primes to ensure a prime presentation below the visibility threshold (i.e., to present the prime words subliminally as described in the original study). It seems possible that the reduced presentation time—while successfully reducing visibility—was insufficient for a subliminal processing of the prime words. Future studies could investigate the possibility of a longer (i.e., 25 ms) presentation time, while ensuring that the prime words are indeed not visible to the participants. This could be achieved by a parafoveal and/or masked presentation of the prime word, as implemented by Rydell and colleagues in a subsequent replication study of this effect (Rydell et al., 2008). Including a forward mask and a parafoveal presentation could have the effect that prime words are not visible to the participants, but a longer presentation time could result in an unconscious processing of the primes and possibly therefore in an effect of prime words on implicit attitude measures. It should be noted however, that another study on evaluative conditioning, using a masked and parafoveal presentation method, was not able to detect EC effects on implicit or explicit measures (Dedonder et al., 2014).

One additional noteworthy characteristic of the method used in the original experiment—which has not been discussed previously—is the length of feedback during the learning phase. At 5000 ms, this feedback is relatively long; this was sufficiently noteworthy that some participants mentioned it after the study. While we did not change the length of feedback—in order to keep the method as close as possible to the original study—it seems possible that participants' attention may not remain focused on the screen for the entire duration of the feedback. Importantly, prime processing may suffer from this reduction of attention because the brief prime words were presented just after the feedback, with only a short fixation cross preceding the prime. As attention towards the subliminally presented stimulus might be a necessary precondition for their cognitive processing (Dehaene et al., 2006), this characteristic of the learning phase could be further investigated—and perhaps improved upon—in future studies. Such improvements (i.e., ways to ensure full attention on the briefly presented primes) might help strengthen the effects of primes on attitudes, but they would simultaneously induce greater prime visibility, thereby rendering it less likely to obtain the original finding.

Implications & Outlook

While the original study is cited as an important finding demonstrating the influence of briefly presented primes on participants' implicit attitudes (Sweldens et al., 2014), our results show that the finding is not easy to replicate. However, at this point we would refrain from drawing strong conclusions from the present two replication studies, for two reasons: First, while we ran two replication in two different labs, we used the same experimental script (i.e., same font, same colors, etc.). Specific—unknown—settings might have interfered with the original effect; and replications by other labs, perhaps using a variety of sensible presentation parameters, would be highly desirable. Second, there are two replications of the original finding (albeit with slightly altered methods; Rydell & McConnell, 2006; Rydell et al., 2008). Those studies—while not independent—are important replications of the effect and should be taken into account when running additional replication attempts. Nevertheless, should future studies confirm the interpretation that the original findings are not robust, this will have important theoretical implications as discussed next.

One theoretical aspect of the original study worth discussing is the duality of attitudes and their measurement, as proposed in the systems of reasoning account (Rydell & McConnell, 2006). The original finding suggests that explicitly learned attitudes map onto explicit rating measures, whereas implicitly acquired attitudes map onto implicit measures. In other studies in support of the systems of reasoning account, it was shown that explicit attitudes adapted quickly to new (explicit) information, while implicit attitudes did not adapt quickly, when new (explicit) information was encountered (Rydell & McConnell, 2006). This effect was taken as an indication for the duality of fast-changing explicit attitudes versus slow-changing implicit attitudes. However, as the authors themselves suggest, the delay in adaptation to new information exhibited by implicit measures might be due to their reduced sensitivity (i.e., the IAT may simply be less sensitive to the information), instead of supporting the existence of dual attitudes. According to the systems of reasoning account (Rydell & McConnell, 2006), demonstrating an influence of subliminal primes on implicit attitudes (while influencing the explicit attitudes using supraliminal behavior descriptions) was therefore important to lend credibility to the proposed duality of attitudes. We were not able to demonstrate this important effect in our two replication studies. If additional studies support the present finding, the differences in implicit and explicit attitudes found previously might indeed turn out to be best explained by a difference in the measures, not in mental constructs.

Interestingly, in the above-mentioned studies, it was shown that 100 trials of explicit counter-attitudinal information had an effect on implicit attitudes (Rydell & McConnell, 2006). With these results in mind, it would be possible that the explicit information presented to participants in our replication studies might similarly have interfered with the opposite-valence information provided by the briefly presented primes. Additional findings also contradict the proposition of an unequivocal mapping of explicit and implicit measures and attitudes. First, several EC studies reported unconscious EC effects on explicit ratings (Hütter et al., 2012; Olson & Fazio, 2001). Second, recent work also showed that implicit attitudes can be accessed and reported in a conscious manner (Hahn, Judd, Hirsh, & Blair, 2014). Third, additional work showed that even indirect measures of EC effects relied on contingency awareness between CS and US during the learning phase (Pleyers et al., 2007; Stahl et al., 2009). Taken together, the assumption of dual attitude models, that "people can hold different implicit attitudes and explicit attitudes about an attitude object at the same time" (Rydell & McConnell, 2006, p. 1007) is challenged both by recent work and the present replication attempts.

Another theoretical aspect of this replication attempt is the role of contingency awareness in EC. The results of the present replication studies are in line with other studies claiming that contingency awareness is necessary for evaluative conditioning (Dedonder et al., 2014; Pleyers et al., 2007; Stahl et al., 2016, 2009). The present studies did not provide evidence for EC in the absence of contingency awareness; yet the existence of such an effect has been proposed by dual-process models of attitude acquisition (e.g., Gawronski & Bodenhausen, 2006). As the original study has previously been cited as (methodologically) strong evidence for EC with subliminal stimuli–and therefore for EC without contingency awareness—the results of these replication attempts question the robustness of this effect and therefore lend support for the notion that EC requires contingency awareness. Clearly, the present work realizes only one of many possible ways to investigate contingency awareness in EC; other possible realizations should be taken into account by future studies addressing the necessity of contingency awareness in EC.

To conclude, we set out to replicate the finding that briefly presented primes had a selective influence on participants' implicit attitudes. In two experiments, we were not able to replicate the original finding but instead found evidence for its absence. If confirmed by future work, this suggests a lack of robustness of the original finding that would considerably weaken the empirical support for dual-attitude theories.

Chapter 3: Introducing potentially beneficial factors for subliminal EC effects

The evidential value of the studies by Rydell and colleagues—according to the *p*-curve analyses—appeared to be very promising. The goal of the previous investigation was therefore to establish an EC effect of subliminally presented stimuli using the paradigm of Rydell et al. (2006). Once an EC effect with briefly presented stimuli was established, we would be able to test specific problems in the measurement of subliminality (i.e., studies measuring memory after the learning phase as a proxy for stimulus visibility; Lovibond & Shanks, 2002) and might be able to adapt the procedure. However, in two studies we were not able to detect this EC effect and replicate the finding by Rydell and colleagues (2006).

An additional important finding of the replication studies was that recognition of the briefly presented words was above chance in the study using similar presentation conditions as in the original experiment (i.e., Experiment 1). This is particularly surprising as the recognition measure was administered long after the learning phase (i.e., half of the stimuli were presented only in the first block and the measure was therefore taken approximately 30 minutes after the end of this learning block). As memory might be influenced by multiple factors between the learning phase and the recognition test (Lovibond & Shanks, 2002), this result strongly indicates that the stimuli were (partially) visible to the participants during the learning phase. These results demonstrate that a better test of stimulus visibility during the learning phase needs to be employed. A study reporting EC effects with briefly presented stimuli is only of theoretical relevance if the stimulus presentation was truly subliminal, as only then we can rule out contingency awareness between CS and US. In the experiments discussed in the next chapters, stronger tests for stimulus visibility will therefore be deployed.

Importantly, we were not able to replicate the finding by Rydell and colleagues (2006) that a briefly presented word had an influence on (implicit) evaluations. One could think of multiple reasons why we did not find an EC effect of the briefly presented stimuli. The experimental setup was only suited to test for a strong dual-process claim (i.e., independent influences on implicit and explicit processes that cannot communicate). While this pattern of effects is possible in the APE model and the studies by Rydell are explicitly discussed (Gawronski & Bodenhausen, 2011), other circumstances in the processing of stimuli might cause associative information to be "overwritten" by propositional information. In the next chapters, the focus will be on a "simple" effect of subliminally presented stimuli on evaluations instead of the double dissociation presented by Rydell and colleagues (2006). We will consider a series of factors that might have hindered an EC effect in the previous studies or could be considered beneficial for automatic EC effects.

First, one problem might be that the US was presented subliminally. As previously discussed, an affective reaction towards the US might require conscious processing (Lähteenmäki et al., 2015). The absence of an EC effect might therefore be due to the lack of an affective reaction even though the stimulus might have been processed unconsciously. Put differently, when the US is presented subliminally it has to (1) be processed unconsciously and (2) produce an affective reaction, while when the CS is presented subliminally it only has to be processed (and associated with a supraliminally presented US). One could therefore argue that EC effects with subliminal USs pose a higher burden on the subconscious processes than an EC effect with subliminally presented CSs. In the following chapters, therefore, CSs will be presented subliminally and USs supraliminally.

Second, as associations in the APE model are described as "co-occurrences between objects and events result[ing] in a co-activation of their corresponding mental concepts, which in turn creates an associative link between the two" (Gawronski & Bodenhausen, 2014b, p. 191), it might be beneficial for associative EC effects to make use of CSs and USs that might already have some associations. Specifically, Verwijmeren et al. (2012) showed in a study using subliminal CSs that a stronger conditioning effect was found when the US carried relevance to the CS. We therefore made use of the procedure by Verwijmeren et al. (2012) as this study explicitly addresses the factor of relevance as possibly beneficial for a conditioning effect with a subliminally presented CS.

Third, an evaluation as the dependent variable might not capture small changes in preferences after an EC learning phase with subliminally presented stimuli as the effect size might be considerably smaller than with supraliminally presented stimuli (however, note that Hofmann et al., 2010 did not find any indications for a smaller effect). In the study by Verwijmeren et al. (2012) a two-alternative forced choice measure captured an EC effect after the learning phase while an evaluation did not show a difference. In the next chapter, a choice measure will therefore be used alongside an evaluation to investigate the question whether the choice measure might be more sensitive to detect a subliminal EC effect than an evaluation.

The aim of the study presented in the next chapter was therefore to find an EC effect with briefly presented stimuli by following the procedure of a study successfully demonstrating subliminal conditioning effects (Verwijmeren et al., 2012). We speculated that (1) a subliminal CS instead of a subliminal US, (2) a relevance between CS and US with an additional personal relevance to attend the stimuli, (3) and a choice measure might constitute as beneficial methodological circumstances to find an EC effect with briefly presented stimuli. We also made sure that CSs were indeed presented subliminally by using a strict visibility test.

Chapter 4: Relevance effects in evaluative conditioning: A pre-registered replication and extension of Verwijmeren, Karremans, Stroebe, & Wigboldus (2012)

An important and old research question is whether our preferences and actions are driven solely by our conscious and deliberate decisions, and to which degree unconsciously processed stimuli and spontaneous impulses also influence our behavior. The issue of whether preference acquisition can occur automatically is central to current dual-process theories of attitudes (e.g., Gawronski & Bodenhausen, 2006). Preference acquisition in the absence of conscious awareness is one of two examples for automatic or unconscious learning phenomena (Mitchell et al., 2009). One way in which stimuli may be unconsciously processed is when they are presented subliminally (e.g., very briefly), resulting in a stimulation too weak to enter consciousness (Dehaene et al., 2006).

Preferences can be acquired when an initially neutral stimulus (the conditioned stimulus, or CS), after being paired with a valent (positive or negative) stimulus (the unconditioned stimulus, or US), is evaluated in accordance with the valence of the US; this phenomenon is called evaluative conditioning (EC). It is of theoretical importance for theories of learning as well as theories of attitude acquisition and expression because it is one of very few phenomena that have sometimes been found to demonstrate automatic learning, or, more specifically, learning without awareness of what is learned. For instance, EC effects have been found in the absence of participants' awareness of the CS-US contingencies (e.g., Fulcher & Hammerl, 2001; Olson & Fazio, 2001). Yet, these findings have been criticized on methodological grounds (e.g., Pleyers et al., 2007), and there is ongoing debate about the adequate method of studying learning without awareness (Gawronski & Walther, 2012; Hütter et al., 2012; Stahl et al., 2009) and of establishing the subliminality of stimulus presentations (e.g., Pratte & Rouder, 2009).

A central criticism is that in most studies of awareness in EC, memory for CS-US pairings assessed *after* the learning phase is used as a proxy for awareness of these pairings *during* learning; this is problematic because participants may be aware of a CS-US pairing during learning but forget about this pairing before the memory test. This may lead to underestimation of awareness, and thereby, to erroneous conclusions about automatic EC effects (e.g., Gawronski & Walther, 2012). As one possible solution to this problem, awareness can be assessed during learning. Another possible solution is to manipulate awareness experimentally, for instance by presenting the CSs subliminally (i.e., preventing the CS from entering awareness implies that the CS-US relation remains unaware).

A recent paper (Verwijmeren et al., 2012) using subliminally presented CSs, provides a possible condition for automatic learning effects: goal-relevance of the learning procedure. In this study, participants attended a stream of stimuli presented on the screen and were instructed to performed a simple identification task (i.e., they had to indicate whether an X or an O was presented on the screen). As part of this stream, they were presented with bottled-water brands for 17 ms, followed by faces with expressions of either fear or disgust for 2s. The brand paired with a facial expression of fear was preferred over the brand paired with an expression of disgust. Importantly, this was only the case for participants who reported higher levels of thirst, not for those with lower-than-average thirst levels. Similarly, evaluations of goal-irrelevant airline brands were not affected by the pairings. In another study, the same finding was reported for supraliminal CS presentations. Thus, an evaluative learning effect was found for subliminally (and supraliminally) presented CSs that was modulated by participants' motivation or goal.

The study by Verwijmeren et al. (2012) is one of very few reports of conditioning with subliminal CSs. It is consistent with a recent meta-analysis of EC that, among many other findings, found comparable effect sizes for sub- and supraliminal CS presentations (Hofmann et al., 2010). However, this conclusion must be interpreted with caution, for several reasons. First, only eight studies with subliminally presented CSs were considered. Second, most of the studies that were analyzed are not easily interpretable in terms of subliminal EC. In some studies, it was unclear whether the results reflect an EC effect (e.g., Dijksterhuis, 2004) or whether presentations were in fact subliminal (e.g., Gawronski & LeBel, 2008, see Chapter 1 for a discussion). Furthermore, recent studies from our lab consistently showed no EC effects with subliminal CSs (Stahl et al., 2016). In sum, previous evidence for subliminal EC effects appears to be weak. It is therefore important to investigate whether the finding by Verwijmeren et al. (2012) can be replicated.

The finding by Verwijmeren et al. (2012) is not only a demonstration of successful subliminal conditioning, but it also addresses possible beneficial conditions for subliminal conditioning. An important role of motivational relevance has been suggested by the experimental manipulation of motivational relevance by Verwijmeren et al. (2012) – see also the role of motivation on the effect of subliminal presentations on brand choice (Karremans et al., 2006). If we can reproduce this pattern of subliminal EC under conditions of motivational relevance and its absence otherwise, we will be one important step closer to reconciling the inconsistent findings in the EC literature.

The present study. We aimed at determining whether the results by Verwijmeren et al. (2012) could be independently replicated. In close collaboration with the first author of the original paper, we planned to conduct such an independent replication in a strictly confirmatory manner by conducting a pre-registered study. We also extended the original design to clarify three important questions that were not answered in the original work, in order to investigate whether the finding's interpretation as a subliminal EC effect modulated by participants' motivation could be corroborated.

First, it is unclear whether the preference of a brand paired with a fearful face over a brand paired with a disgusted face constitutes an instance of evaluative conditioning. In EC, negative USs are typically contrasted with neutral or positive USs, not with other negative USs. In the Verwijmeren et al. (2012) study, both CSs could have been evaluated more negatively than a baseline (e.g., a neutrally paired CS), reflecting an overall EC effect whose magnitude was modulated by goal-relevance; such an interpretation would be more easily reconciled with the host of EC findings in the absence of goal-relevance. Alternatively, an EC effect could have been restricted to the goal-relevant condition (disgusted faces paired with water brands), with EC effects absent for (negative) fearful faces paired with water brands (as well as airplane brands); the absence of EC for these stimuli would be more difficult to reconcile with the literature. To clarify whether the observed effects reflect negative evaluations of the CSs, and therefore, an EC effect, we introduced a baseline condition of CSs that were paired with neutral USs. If the findings are in fact EC effects (i.e., due to stimulus valence), we expect both fear-CSs as well as disgust-CSs to be evaluated more negatively than the baseline.

Second, it is unclear whether the effect is due to a simple association between CS and US (as predicted by dual-process models of attitude acquisition; Gawronski & Bodenhausen, 2006), or whether it reflects more complex propositional or relational learning. For instance, a recent study paired neutral objects (CSs) with facial expressions of happiness or disgust (Bayliss, Frischen, Fenske, & Tipper, 2007). A main effect of emotion was found that could be interpreted as an EC effect: CSs paired with happy faces were evaluated more positively than CSs paired with disgusted-looking faces. However, this main effect was qualified by an interaction with the gaze direction of the face (towards vs. away from the CS): The effect of emotion was restricted to those CSs that were “looked at” by the face, whereas there was no effect of facial expression on CS evaluation when the face looked away from the CS. This gaze-cueing effect cannot be explained by simple (unqualified) CS-US associations but reflects a relational effect that requires more complex representations such as propositions (see, e.g., Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2012; Moran & Bar-Anan, 2013). In other words, it does not rely on simple association by spatiotemporal co-occurrence but depends on participants’ interpretation of the facial expression (e.g., of disgust) as a reaction to the CS (e.g., a water bottle). A similar gaze-cueing effect may have been operating in the Verwijmeren et al. (2012) study (i.e., a picture of a water brand CS was followed by the disgusted-looking face, creating the impression that the depicted person gazed at the CS and responded by expressing disgust). To exclude a potential gaze-cueing explanation of the finding, we introduced an additional condition in which we used non-face USs (i.e., IAPS pictures) to elicit the emotion of disgust that did not convey any relational information. If the finding depended on gaze-cueing (i.e., if it is based on a qualification of the CS-US relation), we would expect to replicate the effect only for face USs. If the finding was due to unqualified associations, we expected an EC effect for both US materials.

Third, it is unclear whether the CSs were in fact presented subliminally. As argued by Pleyers et al. (2007), it is necessary to establish awareness not on the level of the participant but at the level of each of the several CSs. In subliminal-EC studies, an analysis at the trial level is required: In order to conclude that an EC effect is based on unconscious processes, it must be established that, on every single presentation trial, the stimulus was not identified. Otherwise, if, for example, one of the water brands was consciously visible only once, and perceived by the participant as paired with a facial expression of disgust, this might be sufficient to account for the effect. Setting the CS presentation duration to 17 ms is not sufficient to establish subliminality; visibility also depends on the features of stimulus, mask, context, and even participants’ motivation or engagement in the task (Pratte & Rouder, 2009). To establish subliminality more strictly, we assessed awareness during learning on a trial-by-trial basis.

Integrating the assessment of awareness into the learning phase also solves two methodological problems of standard subliminal presentation methodology: First, it allows an assessment of awareness on a trial-by-trial basis, instead of an overall assessment at the participant or group level. Second, it replaces an indirect assessment of awareness from a separate subliminality check phase, typically administered after the main experimental task, by a direct assessment. Typically, the main experimental task—for instance, a priming task with masked primes and clearly visible targets in which participants are instructed to classify the targets—is followed by a separate subliminality check task (in which participants are instructed to classify the masked primes) to assess whether the masked primes were in fact subliminal. If participants are not able to do so above chance level in the subliminality check task, it is concluded that the masked primes were also not visible in the main experimental task. This conclusion relies on the questionable assumption that prime visibility is identical in both tasks. In fact, Pratte and Rouder (2009) showed that this is not necessarily the case because task difficulty is often not held constant (i.e., identifying targets in the priming task is considerably easier than identifying masked primes in the visibility check): In their study, visibility of masked primes was improved when clearly visible primes were interspersed, thereby maintaining participants' motivation to perform the task.

In addition to assessing awareness on a trial-by-trial basis during the learning phase, we combined subliminal and supraliminal presentations to avoid the task-difficulty artefact. In sum, we asked participants, after each CS presentation, to identify the just-presented CS (i.e., select it out of a set of options). If participants were not able to select the correct CS, we concluded that it was not consciously visible on that trial. Only those CSs that were never identified correctly were counted as truly subliminal; we compared the pattern of results for those subliminal CSs with those CSs that were identified on one or more trials, and with CSs that were presented clearly supraliminally.

To summarize, the present study aimed at replicating and extending the original finding by Verwijmeren et al. (2012), with the goal to test whether it (a) could be replicated, (b) reflects a change in liking (i.e., an EC effect) of (fear- and) disgust-CSs when compared with a neutral baseline, (c) reflects an association between the CS and an experience of disgust, or instead a relational gaze-cueing effect, and (d) operates on truly subliminal CSs that were never consciously visible to participants.

Method

The original design was used, with the following modifications to address the above-mentioned goals: (a) In order to avoid the task-difficulty artefact in subliminal presentation, we combined the designs of both studies by manipulating CS presentation duration (sub- vs. supraliminal) as a within-participants factor. This allowed us to replicate both the sub- and the supraliminal effects reported by Verwijmeren et al. (2012) in a single study. (b) We added a neutral condition in which CSs were not paired with a valent US but with a neutral filler stimulus. (c) We added a second disgust condition in which IAPS pictures were used as disgust-eliciting USs. (d) We replaced the X-O identification task, which was used in the original study to ensure participants' attention on the CS and US stimuli, by a CS identification task that served the same purpose and allowed us to assess CS visibility on each trial.

Modifications to materials and procedure included: the generation and pre-testing of new CS materials (which was necessary because we used 8 instead of only 2 brands of each type), the choice of a stronger pre-mask (i.e., because the CSs appeared to be quite readily visible when using the original mask together with the original presentation parameters, we replaced the original mask, which consisted of a blurred version of the CSs, by a random-checkboard-pattern mask), the choice of rating scales for the evaluation and thirst measures (i.e., 200-point instead of 7-point scales), and the realization of a four-alternative forced choice instead of a two-alternative forced choice (AFC) task to accommodate the two additional conditions (i.e., neutral baseline and disgust-eliciting pictures).

Design. A 4 (US type: disgusted face, fearful face, disgust-eliciting picture, neutral) $\times$ 2 (CS presentation duration: 17 ms vs. 1000 ms) $\times$ 2 (CS type: water vs. airline brand) within-subjects design was realized.

Sample size. To be able to replicate the original finding of a subliminal EC effect, we aimed at a sufficiently large sample size and followed suggestions to multiply the original sample size by 2.5 (Simonsohn, 2015). With 46 participants in the central analysis of the original paper (Verwijmeren et al., 2012, Exp.2), this requires at least 115 participants. To err on the safe side, we decided to recruit $N = 135$ participants from the student pool. Due to our participant exclusion criteria, we expected to exclude approximately 7% of the participants resulting in a total of 125 participants included in the analysis. At $\alpha = .05$, this allowed us to detect even small effects

(d = 0.22) with an acceptable power (1-$\beta > .79$), and small-to-medium effects (d = 0.3) with high statistical power (i.e., 1-$\beta > .95$). EC effects of the magnitude of approximately $d = 0.4$, which were obtained by the meta-analysis (Hofmann et al., 2010) for both subliminal as well as supraliminal CSs, could be detected with a very high power of 1-$\beta > .997$.

Materials. As CSs for the thirst-relevant condition, we used 8 water bottles, all with a different logo (but the same bottle), found in a web search. For the thirst-irrelevant condition we used 8 pictures of airline logos. The two sets were taken from a pilot study ($N = 31$) in which a larger set of bottles and airplane company logos were rated on an evaluative rating scale, and neutrally evaluated pictures were selected. No selected picture was reported as known by more than 6.6 % of the participants in this pilot study.

As USs, we used 5 pictures of faces looking disgusted and 5 pictures of faces looking fearful (Langner et al., 2010), which were the same USs as in the original study. In addition, we used 5 disgust-eliciting IAPS pictures as USs (P. J. Lang, Bradley, & Cuthbert, 2008) that were judged in another pilot study ($N = 30$) to reliably elicit disgust and that were not associated with food/drink or transportation. Six colored pixel masks were created, one to be used as a forward-mask in every trial, including the supraliminal trials, and five to serve as neutral USs.

Each CS of a given type was assigned randomly, for each participant anew, to one of the cells of the 4 (US) × 2 (presentation duration) design. Each CS was presented 5 times (once with each US from a given category). This procedure was repeated once, resulting in 160 critical trials.

Procedure. Participants were told that they would see a lot of pictures, some very briefly. It was their task to attend to all pictures in order to identify them and select the appropriate one from the set. They were told that the task was difficult and that they had to guess when not sure.

A trial was constructed as follows: A colored, squared sized, forward-mask in the middle of the screen announced every trial. Immediately afterwards, a CS was shown with the same size as the forward-mask. Depending on the condition, the CS was shown for either 17 ms[12] or for 1000 ms. The CS was replaced by the US which also functioned as a post-mask. Afterwards the screen was cleared. After every trial, a screen appeared showing all 8 CS objects (either all 8 water-bottles or 8 airline logos, depending on the condition in the trial). To report which object they have identified, participants had to select an object by pressing a corresponding number. The arrangement of objects in the identification task was randomized for each trial.

In addition to the 160 critical trials, we also presented 10 trials in which a new bottle (and 10 trials with a new airplane logo) was presented for 83 ms, followed by a neutral US. This was an additional measure to keep up participants' motivation to detect a briefly presented CS (Pratte & Rouder, 2009). The data from these trials were not relevant for our hypotheses. This setup resulted in a total of 180 trials that were part of the learning phase. Order of trials was randomized for each participant anew.

After the learning phase, evaluative ratings and choices were collected. The order in which the two depended variables were collected was counterbalanced over participants. In the evaluative ratings, each critical water bottle and each airplane logo were shown separately and rated on a 200-point slider scale (anchored "very negative" … "very positive"). Water bottle ratings were always asked first, followed by the ratings of the airline logos.

In the choice measure, participants were instructed to select the water bottle they would like to drink. First, the 4 bottles from the subliminal condition were shown on a screen and participants selected a bottle by typing the corresponding number. After one bottle was selected, the remaining 3 bottles were shown again and a second choice was asked for. This process was repeated a third time with the remaining two bottles. The same procedure was then repeated for the supraliminally presented CSs, and accordingly with the airline logos.

Thirst was measured in the end by two questions (e.g., "How thirsty do you feel at this moment"), using a slider with two anchors (i.e., "not at all thirsty" and "very thirsty"). We also assessed familiarity with the brands. In the end the following demographic items and debriefing data were collected: age, study, gender, as well as

[12]The brief presentation duration of one frame on a 60Hz refresh rate monitor was based on two pilot studies (total $N = 8$) which showed a correct identification of the water bottles in 16.4% of the cases, with a chance level of 12.5%. Airplane logos were correctly identified 32.8% of the times. Based on this the airplane logos were shrunk to ensure a lower identification rate in the main study.

an open question regarding their idea of the study's goal.

Hypotheses and analysis plan. The central finding in the original study is the proportion of choices of the disgust-paired beverage and its modulation by thirst. It was analyzed using a logistic regression. CS evaluations were similarly analyzed by regressing evaluative ratings onto thirst. We used the same parallel regression analysis approach for both dependent variables (choice and evaluation).

In addition to thirst, CS visibility was added as a predictor (as well as their interaction). In other words, we investigated the slope of the regression line separately for subliminal CSs that were never identified correctly, subliminal CSs that were identified at least once, and supraliminal CSs. If an effect is independent of awareness, it should operate on subliminal CSs, regardless of whether they are identified correctly or not. If, on the other hand, it depends on conscious visibility of the CS, it should be restricted for correctly identified briefly presented CSs; effects might also be restricted to longer presentation durations more common in EC studies.

The hypotheses (specified below) were addressed by a set of three regression analyses, each involving one of three dependent variable. The dependent variables were computed as independent contrasts from the four US conditions (neutral, fearful face, disgusted face, and disgust-eliciting picture).

Hypothesis (I): If the finding reflects an EC effect (i.e., in the sense that it is based on US valence), then we expect CSs paired with negative USs (fearful and disgusted faces, and disgust-eliciting pictures) to be evaluated more negatively than CSs paired with neutral USs.

To address Hypothesis I we compared evaluations for neutrally paired CSs with those of the three negative CS conditions. The resulting values should be different from zero if there is an EC effect (i.e., if CS evaluations reflect US valence). Furthermore, if EC effects for sub- and supraliminal CSs are of similar magnitude, this effect should not be modulated by CS visibility; if, on the other hand, EC depends on the conscious linking of the US and the CS (as Stahl et al., 2016 and the study in Chapter 2 suggests), we should see an interaction with CS visibility. Reflecting the pattern in the original study, we expect no effect of thirst. We have no specific prediction for choice as dependent variable.

Hypothesis (II): We expect to replicate the finding that, for thirsty participants, water brands paired with fearful faces are preferred and liked better than water brands paired with disgusted faces; this finding should be independent of CS visibility.

To address Hypothesis II, we compared disgust USs (faces and pictures) with fear USs. We expected to replicate the negative regression weight of thirst (i.e., fewer choices and more negative evaluations of disgust-CSs over fear-CSs with increasing thirst); we also expected to replicate the finding for both sub- and supraliminal presentation.

Hypothesis (III): If the finding reflects the simple association of an initially neutral CS with an experience of disgust elicited by the US, then we expect the finding to extend to disgust-eliciting pictures as USs. If, on the other hand, it reflects a relational effect based on gaze cueing, we expect the effect to be restricted to disgusted faces.

To address Hypothesis III, we compared CSs paired with disgusted faces with those paired with disgust-eliciting pictures. If the type of US affects the results as predicted by the gaze-cueing account, we expected an effect of thirst only for the faces but not for the pictures. We have no specific prediction for an effect of CS visibility (although gaze-cueing effects are perhaps more likely to be found for consciously visible CSs).

Data preprocessing. Thirst ratings were standardized prior to regression analyses. Individual evaluative ratings of the CSs were standardized, and contrasts were computed as indicated above. From the 4AFC data, the proportion of choices of each CS will be computed and contrasted as indicated above. Subjects who are familiar with one of the brands used as CS will be excluded. Based on the pilot study this will be approximately 7 % of participants. Additionally, any participant whose correct identification performance was more than 1.5 interquartile ranges below the median of the control items (i.e., those presented for 83 ms to keep up participants' motivation in the identification task) was excluded, as this indicated a lack of motivation to follow the instructions (specifically, if they attend to the stimuli but fail to respond accurately, these participants may be able to consciously identify more stimuli than their identification performance suggests).

Results

Preparing the data. The data from eight participants were excluded due to errors during data collection (one session was aborted due to computer malfunctioning;

four sessions had inappropriate monitor settings; three participants had already participated in pilot studies). These cases were identified during data collection and additional participants were tested to reach our preset goal of 135 participants (101 female, mean age 23.71, range 17 - 60).

According to our pre-specified criteria, we excluded 15 participants with identification performance more than 1.5 interquartile ranges below the median of correct identification in the 83 ms condition. The data of these 15 participants were not used in any analyses as the poor performance indicates a lack in motivation to follow the instructions. We additionally excluded 41 participants who indicated that they were familiar with one of the brands used as CSs. This resulted in a total sample size of $N =$ 79 ($M_{age} = 23.41$; $SD_{age} = 5.80$) for the confirmatory analyses, which—although much smaller than expected—was still greater than the sample ($N = 46$) in Experiment 2 of the original article.

Identification performance of CS stimuli was as expected: The proportion of correctly identified CSs was practically perfect ($M = 0.98$, $SD = 0.04$) for those presented for 1000 ms, and was reduced to 0.88 ($SD = 0.16$) for the control stimuli presented for 83 ms. Finally, for the CSs presented for 17 ms, the proportion of correct identifications was 0.21 ($SD = 0.20$), which was somewhat greater than expected on the basis of pilot tests and was clearly above the .125 chance level, $t(631) = 10.56$, $p < .001$.

In the confirmatory analyses, we distinguished between three types of CSs: CSs presented for 17 ms that were never correctly identified in the visibility test, CSs presented for 17 ms that were identified correctly at least once, and CSs presented for 1000 ms. For a given participant, a CS that was presented for 17 ms was viewed as truly subliminal only when it was never correctly identified in the visibility test. Approximately one correct identifications per CS would be expected by chance, but to allow for a strict test of the subliminal EC hypothesis (i.e., to exclude the possibility that an EC effect may be due to a single consciously processed pairing), all CSs that were correctly identified at least once – whether due to guessing or due to conscious identification – were regarded as potentially consciously visible in the set of confirmatory analyses.

In the 4AFC choice task, the identification performance for each of the four 17 ms CSs that comprised the set of four choice options must be considered, because the conscious processing of only one of the CSs would be sufficient to affect choices. That is, in the absence of subliminal EC effects, if only one of the CSs had been consciously visible during learning, this may lead to the selection of this visible CS—if it had been linked with a neutral US—or to the selection of any of the other CSs, if the visible CS had been linked with a negative US. Accordingly, the visibility criterion was applied at the level of the set of four CSs: An effect on choice was viewed as truly subliminal only if none of the four CSs had ever been correctly identified during the learning phase. This strict criterion yielded zero subliminal cases – all participants had correctly identified at least one of the 17 ms CSs at least once. We were therefore not able to report logistic regression analyses for CSs that were never identified correctly; instead, we conducted the planned logistic regression analyses separately for the subset of CSs presented for 17 ms (i.e., of which at least one CS had been correctly identified at least once in each case, so that choices may have been influenced by conscious learning processes) and the subset of CSs presented for 1000 ms.

Data from the sequential 4AFC choice task were coded as follows: A choice was classified as either in line with a given hypothesis (and coded as 1) or not (coded as 0). For the first hypothesis, a participant's choice was coded as 1 if the CS paired with the neutral US was chosen first and 0 otherwise. For the second hypothesis, a participant's choice was coded 1 if the CS paired with the fearful face was selected before any of the two disgust-paired CSs were selected, and 0 if any disgust-paired CS was selected before the fear CS. For the third hypothesis, a choice was coded 1 if the CS paired with the disgust-eliciting picture was chosen before the CS paired with the disgusted face, and 0 otherwise.

The two thirst ratings had a Cronbach's Alpha of 0.85. An average score was calculated and standardized.

Confirmatory analysis.

Evaluative ratings.

Separate regression analyses were run for evaluative ratings of water and airline CSs as dependent variable, with participant as random factor and visibility and thirst as well as their interaction as independent variables. First, we addressed whether an overall EC effect was present (hypothesis I), that is, whether CSs paired with any of the negative USs were evaluated more negatively than CSs paired with the neutral USs. This was not the case, neither for water CSs, $F(1, 543) = 1.22$, $p = .27$, nor for airline CSs, $F < 1$, $p = .87$; and there was no modulation by visibility or by thirst or their interaction (smallest $p = .17$). This indicates the absence of an overall EC effect on evaluative ratings. Second, we investigated whether water brands paired with fearful faces were preferred over water brands paired with disgust USs (faces and pictures; hypothesis II). This was not the case, $F(1, 385) < 1$, $p = .33$, and there were again no effects of visibility, thirst, or their interaction (smallest $p = .11$). In other words, we did not obtain the expected finding of a thirst-modulated preference for fear-paired water bottles over disgust-paired water bottles. A parallel analysis of airline CSs also yielded no significant effects. Third, we compared water CSs paired with disgusted faces with those paired with disgust-eliciting pictures (hypothesis III). Again, no significant effects were obtained, but in contrast to our hypothesis, CSs paired with disgust-eliciting pictures tended to receive more negative evaluations than those paired with disgusted faces, $F(1, 227) = 2.98$, $p = .086$. This trend is in the opposite direction to that predicted by the gaze-cueing explanation; results therefore suggest the absence of a gaze-cueing artifact.

Choice.

The set of confirmatory analyses aimed at comparing performance for CSs that were never identified correctly with those that were identified at least once. Yet, with CS identification clearly above chance, for each participant at least one of the four CSs present in the 4AFC task was identified correctly at least once. Therefore, the planned set of confirmatory analyses could only be computed for 17 ms CSs, which were correctly identified at least once, and for the 1000 ms CSs. For these subsets, and separately for water and airline brands, we computed logistic regression analyses with each of three choice variables (computed according to our hypotheses) regressed onto thirst.

Contrasting our hypotheses, in none of the analyses thirst predicted choice (all $ps > .28$). There was, however, a preference for water CSs paired with disgusted faces over those paired with disgust-eliciting pictures (i.e., contrary to hypothesis III): Significant intercept terms for both presentation times indicated selection of the CS paired with disgusted faces in more than 50% of cases (17 ms: $B = -0.49$, *Wald* $= 4.46$, $p = .035$; 1000 ms: $B = -0.77$, *Wald* $= 10.12$, $p = .001$).[13] This finding is

[13]Because the null hypothesis of no preference was .25 for the choice of neutral versus all other CSs (hypothesis I) and .33 for the choice between the fear CS and the two disgust CSs, the intercept tests against .5 are not informative and results are not reported

consistent with the above trend towards more negative evaluations for CSs paired with disgust-eliciting pictures as compared to CSs paired with disgusted faces.

Confirmatory analysis: Summary and discussion.

With the possible exception of a trend toward a less favorable evaluation of water CSs paired with disgust-eliciting IAPS pictures when compared with CSs paired with disgusted faces, the confirmatory analyses of evaluative ratings yielded no evidence for an overall EC effect, nor for a preference of fear CSs over disgust CSs, as well as no evidence for a moderation by thirst. We also found no evidence for a gaze-cueing artifact. However, contrary to our hypothesis III but consistent with the trend in evaluations, the choice data showed a preference for water CSs paired with disgusted faces over those paired with disgust-eliciting pictures. This effect was found for CSs presented for 1000 ms, and it was also found for CSs which were presented for only 17 ms (but recall that this effect cannot be interpreted as subliminal because, for all participants, at least one of these 17 ms CSs had been correctly identified on the visibility test at least once).

The pre-defined exclusion criteria were knowingly chosen very conservative, in order to be able to present a strong case for subliminal EC effects. However, there were no EC effects on evaluation – subliminal or supraliminal – after excluding participants as planned. This finding may be due to reduced statistical power but also due to increased error variance (i.e., if EC effects are absent for 17 ms CSs which were correctly identified only by chance, combining these data with those of visible CSs will tend to mask any EC effects for the latter). In fact, more participants had to be excluded than could be expected based on pilot data. First, an unexpectedly large number of people indicated that they were familiar with at least one of the 16 brands. In a pilot study of the brands, only 7 % of the participants indicated familiarity with one of the 16 brands used in the present study, which would have resulted in the expected exclusion of approximately 10 participants. Yet, in the main study, 48 participants reported being familiar with one or more brands. One reason for this could be demand effects—participants may have felt uncomfortable to report that they knew no brand at all—a notion that was supported by informal interviews with participants after the study. We therefore conducted a set of exploratory analyses, reported below, that included these participants.

Second, identification of briefly presented CSs was better than expected based on pilot data. With a total of 40 trials during which CSs were presented for 17 ms (i.e., four CSs presented 10 times each), and given eight choice options in the identification task (i.e., a chance level of 12.5%), five correct identifications – slightly more than one per CS – were expected by chance. Whereas we expected performance to be only slightly above chance (i.e., 6 correct identifications) based on pilot data, the observed identification performance corresponds to almost twice the chance level (i.e., nine correct identifications). In the conservative approach used above, all correctly identified CSs—whether due to guessing or due to conscious identification—were classified as potentially consciously visible. A more liberal criterion would allow for a single correct identification per CS that might have arisen by chance. We used this relaxed visibility criterion in the exploratory analyses of the evaluative rating data reported below; in other words, we distinguished between CSs which were identified at or below chance (i.e., in zero or one cases) and those CSs that were correctly identified at above-chance levels (i.e., in two or more cases).

In the 4AFC choice task, the visibility of each of the four 17 ms CS, which comprised the set of four choice options, may have affected choices. Accordingly, the relaxed visibility criterion was applied to the set of four CSs. However, applying the same chance-level criterion as for the rating data (i.e., a maximum of 4 out of 40 correct identifications) yielded zero cases of chance-level performance. For all participants, identification of the set of four 17 ms CSs was above chance. To nevertheless test whether CS visibility affects EC in the choice task, we therefore relaxed the criterion further by allowing a maximum of 8 out of 40 correct identifications (this yielded 66 participants with 8 or fewer correct identifications for the set of four 17 ms water CSs, and 86 participants for the 17 ms airline CSs). In the analyses of participants' choices, we therefore distinguished between 17 ms CSs with low (but still above-chance) identification performance (i.e., 20% correct identifications) and those with high (>20%) identification performance. In sum, the exploratory analyses reported below repeated the planned analyses with a larger number of participants as well as a relaxed visibility criterion.

Exploratory analysis.

Evaluative ratings. The set of analyses reported above was repeated but with a relaxed visibility criterion (i.e., allowing for a single correct identification per CS that would be expected by chance) and a greater sample size of $N = 120$ as participants were not excluded who reported to know any of the brands (see Figures 8 - 10). The only new finding was that the preference for water CSs paired with disgusted faces over those paired with disgust-eliciting pictures was now clearly confirmed, $F(1, 350) = 13.37$, $p < .001$ (see Figure 10). Given that the CSs paired with face USs were evaluated similarly to those paired with neutral USs, the face CSs effectively represented a neutral baseline. The above finding can therefore be interpreted as an EC effect that is known to be robustly obtained when IAPS pictures are used as USs (e.g., Pleyers et al., 2007; Stahl et al., 2009), but was restricted here to the water CSs.

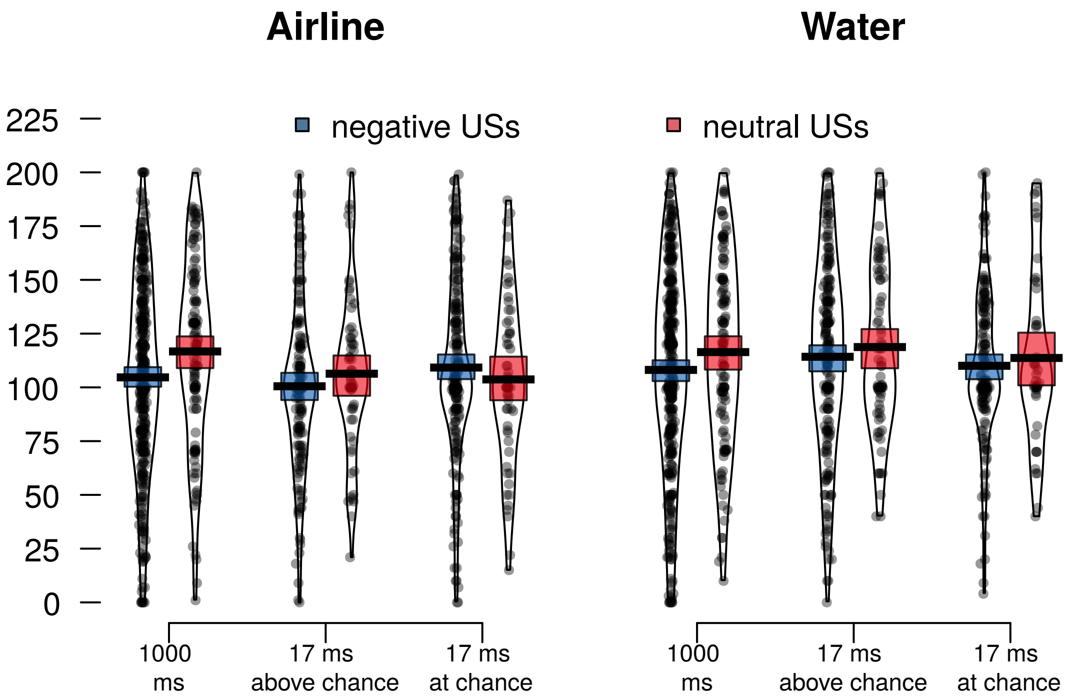

Figure 8. Evaluations are depicted showing the averages (horizontal black lines) with the 95% highest density intervals (in color) displayed around them for Hypothesis I, comparing CSs paired with neutral USs to all other CSs (i.e., those paired with a negative US). The CSs presented for 17 ms are separated by identification performance (at chance: zero or one out of ten correct identifications; above chance: two or more correct identifications). The figure includes data from participants who indicated familiarity with one or more of the brands. Dots represent individual EC effects and additionally the smoothed distribution for each condition is shown.

To further explore the joint influence of all factors of the design, a regression analysis was computed for evaluative rating as the dependent variable, with participant as random factor, and CS type (Water or Airline), US type (disgust-eliciting picture, disgusted face, fearful face or neutral), CS presentation time (17 vs. 1000 ms), and knowledge of brand (yes or no) as independent variables, and including all interactions of these independent variables. A significant effect of CS type was found, $F(1, 1770)$

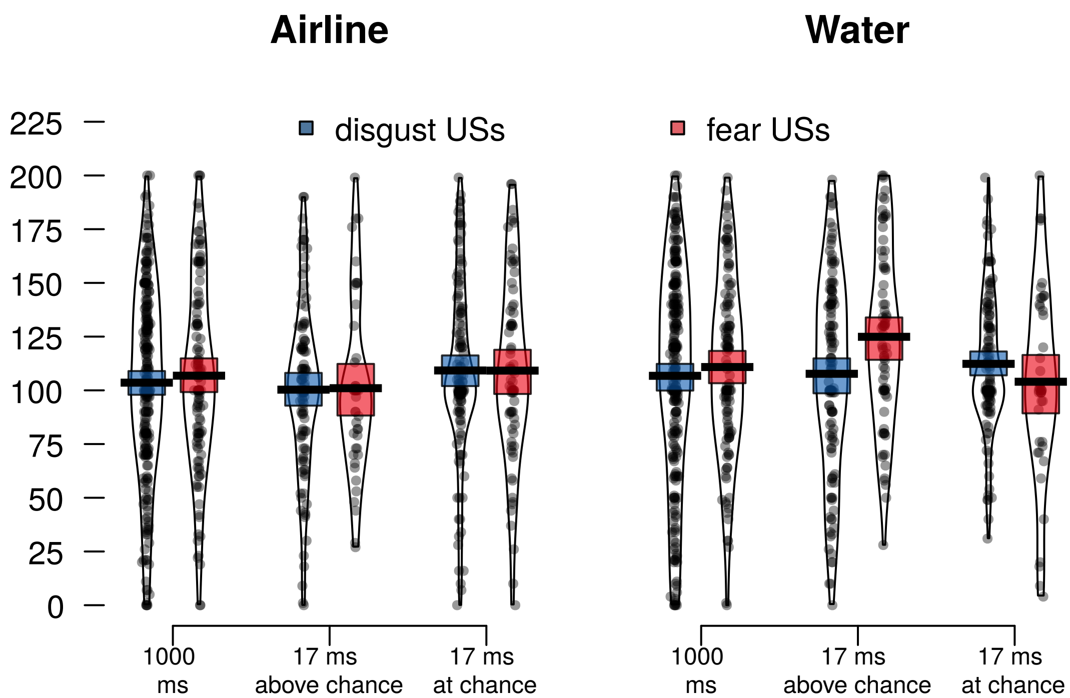

Figure 9. Evaluations are depicted showing the averages (horizontal black lines) with the 95% highest density intervals (in color) displayed around them for Hypothesis II, comparing CSs paired with a fearful USs to CSs paired with a disgust US (disgusted face or disgist-eliciting image). The CSs presented for 17 ms are separated by identification performance (at chance: zero or one out of ten correct identifications; above chance: two or more correct identifications). The figure includes data from participants who indicated familiarity with one or more of the brands. Dots represent individual EC effects and additionally the smoothed distribution for each condition is shown.

= 7.23, p = .007: Airline CSs were rated less likable than water logos. In addition, a significant main effect of US type was found, $F(3, 1770) = 7.44$, $p < .001$. This effect reflected the less favorable rating of the CSs paired with disgust-eliciting US pictures (see Appendix E). No other effects were obtained. Interestingly, there was no two-way interaction of both factors, $F(3, 1770) = 1.26$, $p = .286$; however, when analyzed separately for airlines and water CSs, the effect of US type was restricted to the water CSs, $F(3, 826) = 6.79$, $p < .001$, but disappeared for the airline CSs, $F(3, 826) = 1.93$, $p = .124$.

In addition, there was also no interaction of US type with presentation time, $F < 1$, suggesting that the EC effect extended to the 17 ms condition. To further explore the possibility of a subliminal EC effect in the evaluative ratings of water CSs, separate analyses were performed for CSs presented for 17 ms that were correctly

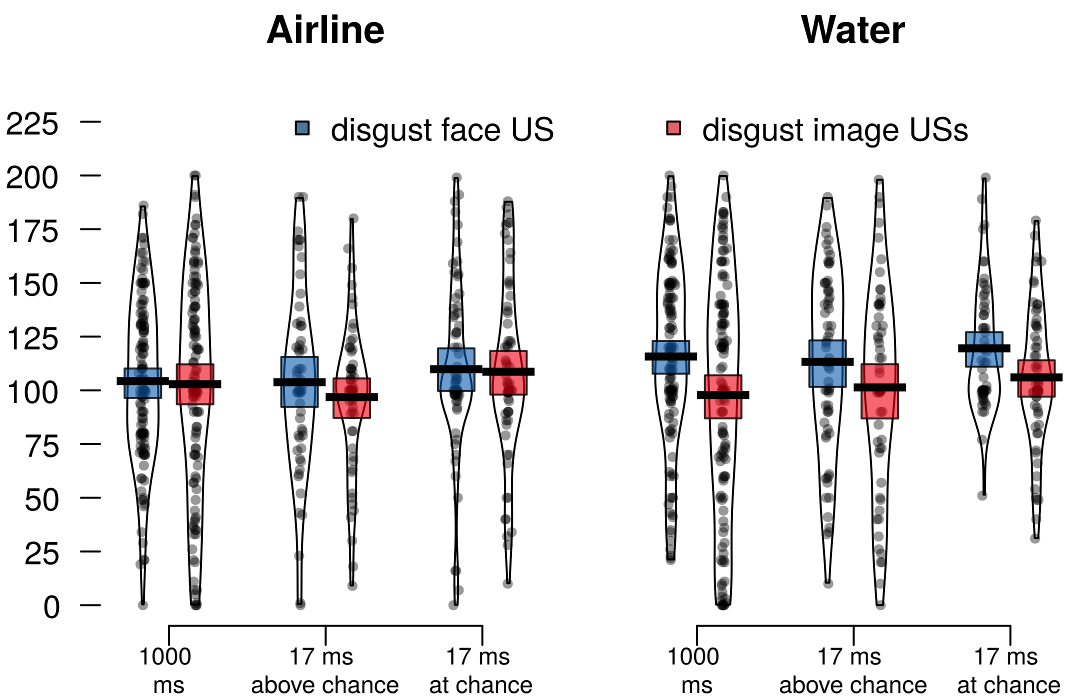

Figure 10. Evaluations are depicted showing the averages (horizontal black lines) with the 95% highest density intervals (in color) displayed around them for Hypothesis III, comparing CSs paired with a disgusted face US to CSs paired with a disgust-eliciting picture US. The CSs presented for 17 ms are separated by identification performance (at chance: zero or one out of ten correct identifications; above chance: two or more correct identifications). The figure includes data from participants who indicated familiarity with one or more of the brands. Dots represent individual EC effects and additionally the smoothed distribution for each condition is shown.

identified at chance levels (i.e., no more than once), for those 17 ms CSs that were correctly identified at above-chance levels (i.e., in two or more cases), as well as for the CSs presented for 1000 ms. A significant main effect of US type was found for 1000 ms CSs, $F(3, 354) = 4.57$, $p = .004$, as well as for those 17 ms CSs with above-chance identification, $F(3, 158) = 3.21$, $p = .025$. Paired comparisons revealed that, in both cases, this effect reflected a less favorable evaluation of CSs paired with disgust-eliciting IAPS pictures, compared to all other CSs (see Appendix E). Importantly, however, there was no EC effect for those 17 ms CSs with at-chance identification, $F(3, 103) = 1.51$, $p = .216$.

A similar analysis for airline CSs yielded a main effect of US type for the 1000 ms CSs only, $F(3, 354) = 2.86$, $p = .037$. Paired comparisons revealed that this overall main effect reflects a preference for the neutral CS over all other CSs (see Appendix E).

Choice.

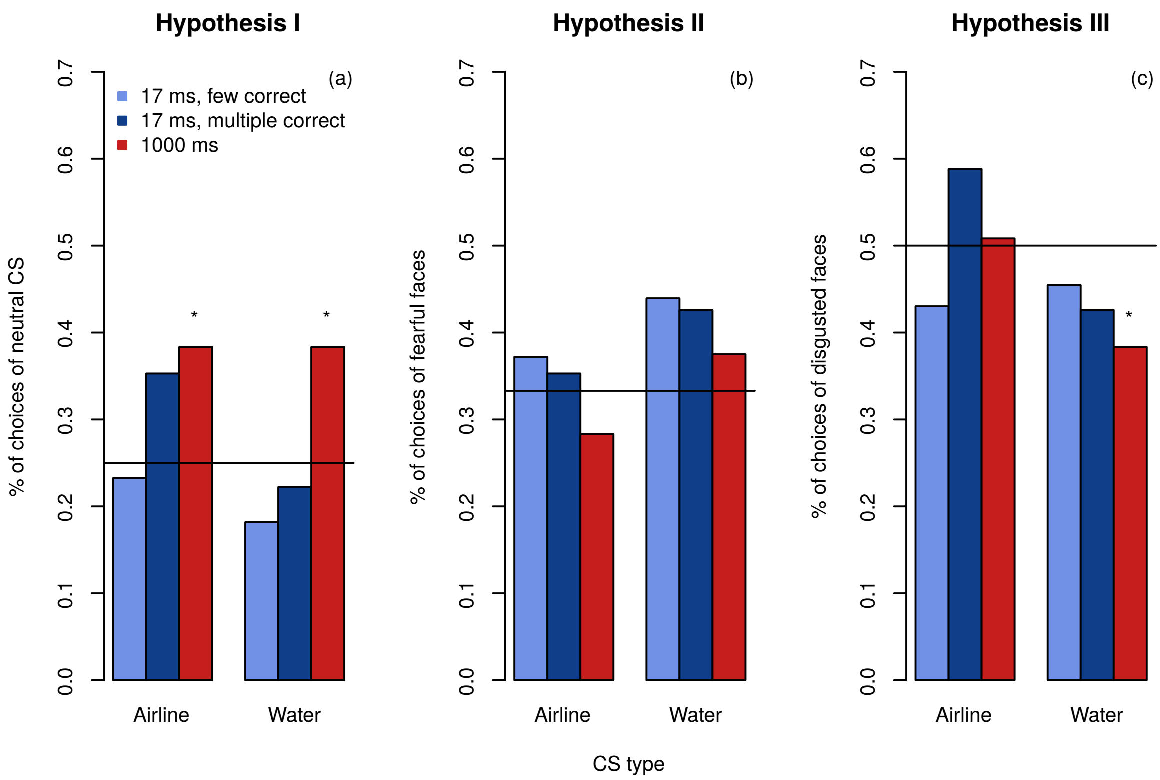

Figure 11. Proportion of participants who preferred a CS paired with a specific US based on the three hypothesis (neutral US vs. all other USs, fearful face US vs. digust USs, and disgusted face US vs. disgust-eliciting image). The horizontal line (at .25/.33/.5) indicates chance level. Asterisks indicate a significant deviation from chance (α = 5%). The CSs presented for 17 ms were all identified at above-chance levels; they are split by identification performance (low: 20% or fewer correct identifications; high: more than 20% correct identifications). The figure includes data from participants who indicated familiarity with one or more of the brands.

For exploratory purposes, the set of originally planned analyses was conducted after re-including those participants who reported being familiar with one or more of the brands and after relaxing the visibility criterion. Logistic regression analyses were conducted with thirst as predictor, separately by CS type and visibility (17 ms low, 17 ms high, 1000 ms). These analyses again yielded no significant effects of thirst, neither for water CSs (smallest p = .379), nor for airline CSs (smallest p = .059). Choice proportions largely reflect the pattern of participants' evaluations, as illustrated in Figure 11. Figure 11a shows that, when CSs were presented for 1000 ms, participants preferred the neutral CS over any of the negative CSs, both for water brands, $\chi^2(1, N = 120) = 11.38$, $p < .001$, and for airline brands, $\chi^2(1, N = 120) = 11.38$, $p < .001$.

Figure 11c shows that, when CSs were presented for 1000 ms, participants preferred the CS associated with the disgusted face over that paired with the disgust-eliciting picture, $\chi^2(1,\ N = 120) = 6.53$, $p = .011$. The remaining choice proportions did not deviate significantly from chance.

Discussion

The present study aimed at replicating and extending the finding that, for thirsty participants, subliminally presented water brands could be evaluatively conditioned by pairing them with relevant disgust stimuli. However, the confirmatory analyses revealed neither an overall EC effect (hypothesis I), nor a preference for fear CSs over disgust CSs (hypothesis II), nor any modulation by thirst. This was true regardless of CS presentation time and identification performance. We will discuss possible causes below. The only significant finding that emerged in the confirmatory analyses was a preference for CSs paired with disgusted faces over those paired with disgust-eliciting pictures, which was contrary to the prediction of the gaze-cueing account (hypothesis III) and was obtained for both 1000 ms CSs as well as the subset of 17 ms CSs which were correctly identified at above-chance levels.

Exploratory analyses based on a larger sample yielded two additional findings. First, the choice data revealed an overall EC effect: For both water and airline brands, participants preferred the neutrally paired CS over any of the negative CSs when the CS was clearly visible (i.e., presented for 1000 ms); this preference was also reflected in evaluative ratings of airline CSs (but was only descriptively present in evaluative ratings of water CSs). If we take this pattern to suggest that the choice task was more sensitive (see also Verwijmeren et al., 2012), the exploratory findings can be tentatively interpreted as evidence for an overall EC effect in line with hypothesis I.

Second, an EC effect for water CSs emerged when the disgust-eliciting IAPS pictures were used as USs. This effect did not interact with CS visibility; however, simple effects showed EC effects only for those CSs that were either clearly visible to participants (i.e., presented for 1000 ms), or were presented briefly (i.e., 17 ms) but identified above chance in the visibility check. This effect on evaluations parallels the finding obtained in the confirmatory analyses of the choice data (i.e., CSs paired with disgusted faces were preferred over CSs paired with disgust-eliciting pictures, but this preference was obtained only for clearly visible 1000 ms CSs as well as for 17 ms CSs which were correctly identified at above-chance levels).

Taken together, these results are consistent with the original finding in two important ways. First, EC effects by disgust USs were obtained only for water brand CSs but not for airline CSs, supporting the point made by Verwijmeren et al. (2012) that EC occurs more reliably when CS and US are related in a meaningful way.

Second, as in the original study, the result has been obtained for CSs presented for 1000 ms as well as for (a subset of) CSs presented for only 17 ms. Setting aside the issue of awareness for a moment, these results replicate the finding that, even with a relatively small number of very brief and suboptimal presentations, novel brands can reliably acquire valence by way of an evaluative conditioning procedure.

With regard to these two central aspects, the present finding can be seen as a partial replication of the finding by Verwijmeren et al. (2012). Two other aspects of the findings diverged from the original pattern. First, in contrast to the two experiments reported in the original study, the EC effect was not modulated by participants' reported thirst. In theoretical terms, we do not replicate the effect of goal relevance on evaluative conditioning. Given that the measure of thirst was practically identical in both studies, and that there was sufficient variability of thirst ratings, we could only speculate about the causes of this discrepancy. Future research is clearly needed to clarify this issue. Second, the EC effect was obtained only when disgust-eliciting IAPS pictures were used as USs; it failed to obtain for the same disgusted faces that were used in the original study. We will discuss possible causes in the next section.

Limitations. One major difference to the study by Verwijmeren et al. (2012) was that we used IAPS pictures as additional USs. The inclusion of the IAPS pictures modified the stimulus context and may thereby have affected participants' attentional focus: In the original study, the difference between facial expressions was the only source of variability, and participants were therefore more likely to attend to facial expression. In the present study, however, the US stimuli varied on different dimensions, such that a broader distinction could be more easily made between faces and IAPS pictures, and comparably subtle differences between fearful faces and disgusted faces might have gone unattended. Alternatively, the addition of IAPS pictures may have induced a different processing mode: Whereas participants in the original study may have processed the faces in a more global impression-formation manner, the presence of the IAPS pictures (and/or the presence of the CS identification task), may have induced a more analytic processing style. This notion of a context-dependent EC effect is consistent with findings showing that the effect of a specific US is modulated by participants' attention (Gast & Rothermund, 2011) as well as processing goals (Corneille, Yzerbyt, Pleyers, & Mussweiler, 2009). Additional research is necessary to determine whether these effects of US context can be replicated, and whether they can be interpreted as effects of attention and/or processing goals.

A second major difference was that all manipulations were done within participants, resulting in eight different CSs per CS category, compared to only two CSs in the original study. This greater number of CSs and conditions reflects an increased complexity of the learning situation and perhaps an increased processing load. This increased load might have interfered with an EC effect by reducing the strength of associative learning (e.g., Pleyers, Corneille, & Luminet, 2009). It is not easy to see, however, why the effect of complexity or load would interfere with the effects of facial expressions but spare the EC effect elicited by disgust-eliciting pictures.

The visibility check might also have interfered with associative learning because, in between unconscious processing occasion of the CSs, people consciously perceived and processed the same CSs during the identification task. This might have caused associations to form among the neutral CSs, and these, in turn, may have weakened the associations between CS and US. Another possibility is that, with the CS being presented before the US, participants might have attempted to ignore the US in order not to forget the CS for the upcoming identification task. Perhaps this top-down influence was overcome in a bottom-up manner by the disgust-eliciting IAPS pictures, but not the disgusted or fearful faces, restricting the EC effect to the disgust-eliciting picture condition. This explanation is, however, inconsistent with findings from brain activation studies comparing responses to emotion-eliciting stimuli with responses to emotional facial expressions, suggesting that neurophysiological reaction to faces are similar to—or even stronger than—those elicited by IAPS pictures (Britton, Taylor, Sudheimer, & Liberzon, 2006; Hariri, Tessitore, Mattay, Fera, & Weinberger, 2002).

CS identification and subliminality. A final feature of the present study was the concurrent visibility test during the learning phase, which assessed participants' ability to identify the briefly presented CS. On average, a CS presented for 17 ms was correctly selected, from among eight options, in one fifth of cases. This is clearly above the level expected by chance, indicating that participants were able to perceive the briefly presented CSs at least partially in at least some cases, and demonstrating that CS presentation in the 17 ms condition was not strictly and objectively subliminal. It is important to note here that participants may nevertheless have been subjectively unaware of the stimuli at the time of presentation. Future research is needed to distinguish between different types of awareness and their role in EC.

For a strong claim that EC operates unconsciously, we sought to establish an EC effect for CSs that were either identified not even once (strict criterion), or identified only at chance levels (in the relaxed criterion). However, the EC effect was found only for those CSs that were identified at above-chance levels, not for those never identified, nor for the CSs identified at chance. This finding is consistent with a series of findings from our lab in which we manipulated CS visibility across materials and learning conditions by varying CS duration and masking, and found EC effects for correctly identified CSs but not for those that could not be identified by participants during the learning phase (Stahl et al., 2016). In a broader sense, the present finding is also consistent with previous findings of EC effects for IAPS pictures that emerged only in the presence of awareness of the CS-US pairing (e.g., Pleyers et al., 2007; Stahl et al., 2009).

In the original study (Verwijmeren et al., 2012, Experiment 2), a joint assessment of visibility and memory for the paired US was made at the very end of the experiment, so it is not clear whether participants' failure to report their conscious perception of CSs presented for 17 ms in this measure is due to true subliminal perception, or instead due to forgetting over the course of the experiment, due to the joint assessment method, or an interaction of both. To the degree that presentation conditions were similar to those in the original study, the present findings can be used as a proxy for the visibility of stimuli in that study: Presentation conditions were similar in that presentation duration was identical, and the same US faces as well as similar CSs were used. However, the CSs were somewhat less likely to be consciously visible in the present study because (1) in contrast to the original study, the present CSs did not feature the name of the brand in black-on-white plain text below the image of the brand, and (2) the high-contrast pre-mask used in the present study disturbed perception more than the blurred water-bottle mask used in the original study. If we use the present above-chance identification performance for the CSs presented for 17 ms as a conservative estimate, it appears that participants in the original study may have been able to identify the CSs at least partially and at least in some cases. If we accept this possibility, the potential discrepancy disappears and both findings can be consistently interpreted as EC effects for briefly presented but not necessarily subliminal CSs.

Outlook. As mentioned above, the increased complexity due to the within-subjects design and the addition of the disgust-eliciting pictures could be an explanation why the original pattern of EC effects could not be fully replicated. One possible way to test this would be to leave out the disgust-eliciting pictures for one set of participants while using them for another set. If our above suggestions hold, we would not find an EC effect of faces in the conditions where disgust-eliciting pictures are used, but the effect would re-emerge in the absence of IAPS pictures. As the inclusion of the IAPS pictures has apparently interfered somehow with the face USs' ability to elicit an EC effect, the present findings do not speak to the question of whether the original EC effect was based on gaze cueing or not. To test this hypothesis more directly, future research could manipulate gaze direction (i.e., as in the study by Bayliss et al., 2007).

In short, this paper has partially replicated the Verwijmeren et al. (2012) study in that it provides additional evidence for a relevance effect in EC, even when the CS is presented very briefly. In the present study, relevance increased the chance of finding an EC effect in the absence of motivational influences. Future research should address the scope and limitations of this effect in terms of stimuli and experimental contexts; in addition, it is of course necessary to further investigate the role of motivation and goals, and to identify the conditions under which motivational influences interact with CS-US relevance.

Chapter 5: Simultaneous CS-US presentations as a necessary condition for automatic EC?

The replication and extension of the study by Verwijmeren and colleagues (2012) yielded a series of interesting results. First, we found an EC effect, as a preference for water bottles (and airline brands) paired with neutral USs over CSs paired with negative USs was observed. Additionally, an EC effect was observed even with stimuli presented for 17 ms. While the visibility test clearly showed that an EC effect for CSs presented for 17 ms only emerged when the stimuli were identified correctly above chance level, an effect with stimuli presented for 17 ms could still be considered interesting. As Stahl and colleagues (2016) demonstrated that above chance identification is not sufficient for an EC effect, the results obtained in the previous study indicate that EC effects with very briefly presented (and above-chance identified) stimuli might be possible. While this finding might not be interesting for the theoretical question of contingency awareness in EC (as contingency awareness might have existed) it might have interesting implications for marketing research (i.e., brief exposure to stimuli might be sufficient for behavioral changes).

The main goal of this experiment was to investigate potential beneficial factors in automatic EC (e.g., relevance, choice measures) while making sure that CSs were indeed presented subliminally. However, we did not find any evidence for EC effects with truly subliminally presented CSs in any condition. Additionally, the (goal) relevance of thirst appeared to play no role. Thirst did not predict EC effects in any of our analyses and we found no evidence that the motivational state to attend relevant CS-US pairs might be of importance in (automatic) EC effects. There were some indications for a larger effect of the IAPS images compared to the disgusted face for water bottles than for airline brands, as this effect emerged only for water brands in the analyses with lower statistical power. However, the effect reported by Verwijmeren et al. (2012) that an effect was only found for water brands and not for airline brands could not be observed in our study. While we would not exclude the possibility that CS-US relevance might be beneficial for (automatic) EC effects, it did not appear to have had a large influence in the present study.

Another step that was taken to find (possibly small) changes in preference after a subliminal EC paradigm was to introduce a measure of choice as the dependent variable. In the confirmatory analysis—with few participants—we only observed a choice effect (admittedly for CSs that were likely consciously observed) but no effect with an evaluative rating. This could be interpreted as an indication that choice measures might be more sensitive to affective reactions on CSs after an EC paradigm than evaluative measures. However, unlike claimed by Verwijmeren et al. (2012) we did not observe any choice effects when the CSs were presented truly subliminally. Another important point is that we used a forced choice paradigm with four choice options (then three options, then two options) instead of a 2AFC as used by Verwijmeren et al. (2012) and suggested by Jones et al. (2010). A 4AFC might not have the characteristics assumed and does not capture spontaneous associations, but participants might be consciously debating which CS to choose first, second, third, last. We will therefore implement a 2AFC in the experiment presented in the next chapter.

A last important finding of the previously discussed experiment is that CS identifications were above chance in the 17 ms condition. Of course it is mere speculation, but the findings by Verwijmeren et al. (2012) might be based on visible—but briefly presented—stimuli. As discussed, we acknowledge the possible influence of the visibility check on the processing of the CS-US pairs in the previous study and address this problem in an experiment presented in the next chapter. One additional problem that will be addressed is that statistical power is automatically reduced when only analyzing stimuli that were identified at or below chance in the 17 ms condition. Therefore finding an EC effect in a clearly visible presentation condition (e.g., for a CS presented for 1000 ms) but not in a subset of stimuli or participants in the 17 ms condition might be criticized as not being a fair test of an automatic EC effect. We will therefore pay attention that stimuli are identified at or below chance in the next studies, in order to have more statistical power to find an EC effect in the brief presentation time (i.e., "subliminal") conditions.

Both studies discussed in Chapter 4 and Chapter 2 made use of a sequential presentation paradigm of the CS-US pair. The US was followed by the CS in the experiments in Chapter 2, while the CS was followed by the US in the experiment presented in Chapter 4. As discussed in Chapter 1, studies indicated that automatic EC effects might only be found when CS and US have been presented simultaneously. This characteristic could explain why we did not observe any EC effects in the previous three experiments. Thus, we will attempt to realize simultaneous presentation conditions in the experiments discussed in the next chapters.

Chapter 6: Subliminal influence on preferences? A test of evaluative conditioning for brief visual conditioned stimuli using auditory unconditioned stimuli

Additional materials are stored in an online repository: osf.io/cx5eh

The acquisition of preferences plays an important role in our daily life: Companies want us to prefer their products over their competitors', politicians want us to prefer them over their opponents, and governmental bodies want us to adopt a healthy lifestyle. Given the multitude of motivations to sway our preferences it is important to establish whether a person's attitudes can be shaped without her becoming aware of the procedure. Evaluative conditioning (*EC*) is one way in which preferences can be acquired: When an initially neutral stimulus (the conditioned stimulus or *CS*) is paired with a positive or negative stimulus (the unconditioned stimulus or *US*), it is subsequently evaluated in accordance with the valence of the US. We argue that understanding processes underlying evaluative conditioning could have a major impact on the understanding of preference acquisition in general. There are currently two dominant families of theories trying to explain the mechanisms that underlie the EC phenomenon (see Sweldens et al., 2014 for a review). Single-process theories propose that the acquisition of preferences can only happen in a conscious, deliberate, propositional manner (e.g., Lovibond & Shanks, 2002; Mitchell et al., 2009). Dual-process models, in contrast, propose that preferences can also be acquired in an automatic, non-conscious manner (e.g., Gawronski & Bodenhausen, 2006). Whereas dual-process models posit an "indirect influence on propositional reasoning mediated by direct influence on associative evaluation" (Gawronski & Bodenhausen, 2006, p. 703) for which contingency awareness is not necessary, propositional single-process models posit that contingency awareness is a necessity for changes in preference to occur (Mitchell et al., 2009). Whether a change in preferences can be achieved when people are unaware of the contingency between the CS and US is, therefore, a central question for theories of EC.

In support of associative processes, Olson and Fazio (2001) report EC effects for participants who were unaware of the CS-US contingency. Others however, have argued that awareness is a necessary condition for EC effects to occur: Pleyers and colleagues have demonstrated that unaware EC effects are found only for those CS-US pairs for which participants could recall the specific US that had been paired with the given CS (Pleyers et al., 2007). Similarly, Stahl et al. (2009) found that EC effects are absent when participants cannot recall the valence of the US that was paired with a given CS. An alternative explanation for these findings is that participants relied on their automatic affective response to the CS to answer questions about the paired US valence rather than reporting their actual memory (Gawronski & Walther, 2012; Stahl et al., 2009). Such an "affect-as-information" account can explain the correlation of EC effects with participants' report of US valence from a dual-process perspective in which EC is in fact independent from awareness (Hütter et al., 2012). Consistent with the latter view, when using a method that avoids this problem, Hütter and colleagues found EC effects in the absence of contingency awareness (Hütter et al., 2012).

Because these previous studies investigating contingency awareness as a necessary precondition for EC have relied on retrospective reports by participants, they have been the target of two methodological criticisms: First, such reports are merely correlational and thus causal inferences cannot be drawn; second, they are susceptible to forgetting. When awareness of the CS-US pair is assessed only after the learning phase (Lovibond & Shanks, 2002; Sweldens et al., 2014), it is possible that a stimulus pair has been perceived during the learning phase but the episode cannot be retrieved during the contingency-awareness assessment at the end of the study. Awareness is therefore systematically underestimated if assessed by retrospective reports (Lovibond & Shanks, 2002). To the same effect, a recent meta-analysis also showed that awareness tests of unconscious learning effects are often underpowered (Vadillo et al., 2015). The authors of this meta analysis show that a joint analysis of awareness checks yields clear evidence for above-chance levels of awareness, a finding that contradicts the null findings reported in the investigated individual studies that were due to insufficient statistical power.

Experimental manipulations of contingency awareness may provide stronger evidence and further illuminate the discussion. One way to manipulate contingency awareness is by way of presentation duration (see Dedonder et al., 2014 for a different approach): Briefly presented and masked stimuli may sometimes be processed without reaching consciousness (Dehaene et al., 2006; Van den Bussche, Van den Noortgate, & Reynvoet, 2009). By showing either CS or US very briefly, the complete CS-US pair cannot be perceived consciously, thereby interfering with contingency awareness. Observing EC under such conditions would support the notion that EC effects can be formed "independent[ly] of conscious awareness" in associative learning (Gawronski

& Bodenhausen, 2014b, p. 193). A dual-process model would provide the most plausible explanation for such a finding, given that propositional theories propose that "associative learning is never automatic and always requires controlled processes" (Mitchell et al., 2009, p. 187).

Previous experiments have demonstrated EC effects with subliminal stimuli (De Houwer et al., 1994, 1997; Fulcher & Hammerl, 2001; Krosnick et al., 1992; Niedenthal, 1990). In those studies, however, it was the US (not the CS) that was presented subliminally. This is critical because it has been shown that the perception of valenced information may be possible even with very short presentation times [Nasrallah, Carmel, and Lavie (2009); Zeelenberg, Wagenmakers, and Rotteveel (2006); cf. Lähteenmäki et al., 2015]. If, in the above studies, the valence of the US has been processed consciously, then any EC effects in these studies can be explained by a propositional single-process model and do not require dual processes. Only very few studies have so far reported EC effects for briefly presented CSs (e.g., Dijksterhuis, 2004), and these studies' methodologies have been criticized, for example, for employing between-subject manipulations of the CSs or of US valence (see Stahl et al., 2016 for a brief review; Sweldens et al., 2014). Hofmann and colleagues conclude in their meta analysis on EC in humans that more research is needed before any conclusions can be drawn about a subliminal EC effect (Hofmann et al., 2010).

A recent study in our lab addressed the possibility of subliminal EC in a series of experiments that attempted to overcome the above-discussed methodological problems. In this study, CS-US contingency was manipulated by varying CS presentation duration and/or masking, and awareness of brief and masked visual CSs was assessed immediately after their paired presentation with visual US images. Across six experiments no EC effects were found, despite the fact that CS identification was slightly above chance (Stahl et al., 2016). The present study aims to test whether EC effects can be found under slightly above-chance CS identification conditions, when three potential shortcomings of the study by Stahl et al. (2016) are addressed.

The present study. The goal of the present set of experiments is to provide an even more stringent and fair test of EC with briefly presented CSs than Stahl et al. (2016). As in those studies, our goal was to improve upon the methodological shortcomings of previous studies discussed by Sweldens et al. (2014). In line with previous work, therefore, we manipulated positive and negative USs within participants to rule out any effects of mood induction; and we manipulated presentation conditions of the CSs rather than the USs to avoid that— even if USs are not reliably identified—conscious processing of US valence could be invoked as a single-process explanation for an EC effect.

First, note that, in the study by Stahl et al. (2016), both the CS and the US were visual stimuli, presented concurrently and next to each other on the computer screen. This was done to increase the chances of EC effects by way of an implicit misattribution process (Jones et al., 2009), which assumes that participants first experience an affective response (elicited by the US) that is then misattributed to the CS because it happens to be the stimulus that is attended at the time the affective response occurs. Factors that supposedly increase the chances of implicit misattribution with visual stimuli are (1) relative salience of the CS, (2) simultaneous onset of CS and US stimuli, (3) shifts of attention (eye gaze) between CS and US, and (4) moderately valent ("mildly evocative") US stimuli (Jones et al., 2009). One problem that arises with the above method is that, when limiting presentation duration of the CS, participants' attention must be directed towards the CS during its brief presentation for it to have any chance of affecting cognition and behavior. This may interfere with the effect of a simultaneous stimulus onset in the sense that it requires that participants process the US and CS sequentially and in that order. If that is the case, it was perhaps less likely in this procedure that participants experienced an affective response elicited by the US while they attended the CS.

On a related note, recent studies have investigated the effect of sequential versus simultaneous CS-US presentation on implicit EC effects. Results indicated that EC by way of an automatic process requires simultaneous pairing of stimuli, whereas an EC effect by propositional processes can also be found with sequential pairings (Hütter et al., 2012; Sweldens et al., 2010). This finding might contribute to explaining the absence of an EC effect in the experiments by Stahl et al. (2016).

To address these problems, we used a cross-modal EC procedure in which the visually presented CS is paired with a simultaneously presented auditory US. This allows not only for a simultaneous CS-US presentation but also for a simultaneous *processing* of CS and US. Auditory USs paired with visual CSs have been used in conditioning procedures with children (Neumann, Waters, Westbury, & Henry, 2008) and product preference studies (Gorn, 1982; Schemer et al., 2008). The previously mentioned meta analysis by Hofmann and colleagues (Hofmann et al., 2010) estimated that the size of a cross-modal EC effect does not differ significantly from unimodal

EC effects. Using this cross-modal approach of CS-US presentation allows for a brief CS presentation and ensures that CS and US can be attended at the same time, an issue which has received only little attention in previous studies on subliminal EC.

Second, we assessed whether EC for near threshold visible CSs can be obtained not only in the presence but also in the absence of an online CS identification task. To ensure that briefly presented stimuli are not clearly visible, in two initial studies we will measure visibility during the learning phase, thereby using an online visibility criterion as in the studies by Stahl et al. (2016). Following every CS-US pairing, participants will be asked to identify the previously presented CS from a selection of all CSs. Using this online visibility criterion, we can assess an average identification rate for each CS for each participant. We use this estimate as a proxy for the perceptual awareness of the CS (while recognizing that this measure may be influenced by unconscious influences on familiarity as well as guessing). If we observe an EC effect for stimuli that were not identified at above-chance levels during the learning phase, this would constitute strong evidence that this EC effect was caused by associative processes and not by conscious propositional processes. If we observe no EC effect, even for CSs that were correctly identified at slightly above-chance levels, a single-process model of evaluative learning would be preferred as the more parsimonious account.

The visibility-check task has the additional benefit of directing participants' attention toward the CS, but it may also induce an analytic task set that may not be conducive for automatic EC. In addition, participants are presented with a set of several CSs as choice options in close temporal proximity with the US; this may dilute automatic EC effects because (a) affective response may be attributed to different CSs, and/or (b) the additional CSs presented on the visibility-check task may also be associated with USs of different valence over the course of the learning phase. In sum, it may be argued that the visibility-check task may interfere with the formation of subtle CS-US associations during learning. To address this possibility, in a final study participants will work on a different task during the learning phase.

Third, in order to provide optimal conditions for obtaining even subtle EC effects, we introduced an additional, potentially more sensitive dependent measure: In a two alternative forced choice (2-AFC) task, we are pitting two CSs against one another that were paired with USs of opposite valence. In at least one study on subliminal influence a 2-AFC task showed a significant effect while an evaluative rating did not show an effect (Verwijmeren et al., 2012). It seems plausible, therefore, that small and subtle EC effects might not be reflected in evaluative ratings because they were not perceived as large enough to justify selection of a different point on, say, a 7-point scale—but may nevertheless tip the scale when being forced to choose between two CSs that were paired either with a positive or a negative US.

As in Stahl et al. (2016), we wanted to be certain that automatic processes have a fair chance to operate and to produce EC effects. In the study at hand we therefore realized near-threshold but slightly above-chance (instead of fully subliminal) presentation conditions. We can therefore be certain that the brief visual CS stimuli could in fact be processed, and that automatic processes were given a fair chance to operate on them.

Taken together, across three experiments we investigated EC effects for clearly visible as well as for near-threshold visual CSs that were paired with auditory USs. In the unregistered Experiments 1 and 2, we used an online visibility-check task as an index of contingency awareness; the pre-registered Experiment 3 tested whether near-threshold cross-modal EC can be found in the absence of an online visibility check. In Experiments 1 and 3, a 2-AFC choice measure was administered to test whether EC with briefly presented CSs can be found on this potentially more sensitive measure.

Experiment 1

The first experiment tested if we can observe a cross-modal EC effect with CSs presented for 1000 ms as well as with CSs presented 17 ms, while strictly assessing CS visibility during the learning phase.

Method.

Design.

In Experiment 1 we manipulated the presentation time of the CS (17 ms vs. 1000 ms) as well as US valence (positive vs. negative) within participants. We manipulated the order of dependent measures (2-AFC or rating first) between participants.

Sample.

An a-priori power analysis for a paired one-sided t-test with $\alpha = \beta = .05$ and $d = 0.3$, which is approximately the effect size found by Olson and Fazio (2001), yielded a required sample size of $N = 122$. We recruited 123 participants for this study; three participants aborted the study, leaving 120 participants for the final analysis (Age $M = 22.98$, $SD = 5.40$). Participants were mostly University of Cologne students who received partial course credit for their participation.

Material.

Eight grey-scale drawings of cars were used as CSs. The contrast of the images were equated to ensure comparable visibility of all cars under brief presentation conditions. Large differences between images could result in high identification rates in the online measure of awareness, overestimating the visibility of the image. Based on a small pilot study ($N = 28$) the most neutral images with low standard deviations were selected as CSs for this experiment. Two additional images were selected for filler trials. 10 positive, 10 neutral, and 10 negative sound files were selected from the IADS data base (see Appendix F; Bradley & Lang, 1999). All experimental scripts, data files, and analysis scripts are available at https://osf.io/cx5eh/.

Procedure.

Participants were seated in front of a 60-Hz CRT monitor and instructed to attend to and memorize image-sound pairs. Furthermore, participants were told that they would be asked to identify the presented image from a set of 8 images after every trial.

Each CS was randomly assigned to the positive or negative valence condition for each participant anew. CS-US pairs were shown 10 times, resulting in 80 critical trials. Each trial started with a 2100 ms blank screen, followed by a 700 ms forward mask. Afterwards, the CS was shown for either 17 ms or 1000 ms. The sound of the US had the same onset as the CS image and an average length of 3462 ms ($SD = 359$ ms). Each CS was followed by the same pixel mask for 2000 ms and a subsequent 2000 ms blank screen. The order of trials was randomized.[14]

In addition to the 80 critical trials, there were 20 filler trials in which two additional images of cars were presented for 85 ms and paired with 10 neutral sounds. These trials served the sole purpose of motivating the participants to attend to the briefly presented CSs (Pratte & Rouder, 2009). The same two images of cars were used for these trials, and no rating data was collected for these images.

[14] The experiment was programmed using PsychoPy (Peirce, 2007).

Following every critical trial, participants performed the identification task by selecting one CS from an array of all 8 CSs by pressing a corresponding number on the keyboard. For filler trials, participants selected an image from an array of 6 random critical CSs and the two filler images. Participants were instructed to guess if they did not know which image had been presented.

After the learning task, participants indicated how much they liked each CS (presented in random order) on a slider ranging from -100 (*not at all*) to 100 (*very much*; saved as 0 to 200). In the 2-AFC task, a positively-paired and a negatively-paired CS from the same presentation time condition were pitted against each other at random. Each CS was shown only once in the 2-AFC, yielding four choices in total (i.e., two for CSs presented for 17 ms and two for CSs presented for 1000 ms). Participants were instructed to choose the car they would actually want to buy. The order of dependent variables was counterbalanced across participants. In the end participants provided demographic data (age, gender, and occupation), speculated about the goal of study, and could provide additional comments.

Data analysis.

We report p values and Bayes Factors as inference criteria for all analyses. Bayes Factors are readily interpretable as changes in model odds. Thus, Bayes factors can be interpreted as a measure of evidence for an alternative hypothesis relative to the null hypothesis given the observed data (e.g, Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). We use BF_{10} to denote the evidence for the alternative hypothesis relative to the null hypothesis (i.e., $BF_{10} > 1$ indicates support for the alternative hypothesis) and BF_{01} to denote the evidence for the null hypothesis relative to the alternative hypothesis (i.e., $BF_{01} > 1$ indicates support for the null hypothesis). We report Bayes factors rather than posterior model odds because they can be used to determine the rational posterior belief in a hypothesis based on any subjective prior belief in the hypotheses. The α-level for all frequentist analyses was .05.

We performed repeated measures ANOVA to analyse evaluative ratings and paired t tests to examine interaction effects. To compute ANOVA Bayes factors we used default multivariate Cauchy priors as described by Rouder et al. (2012) with a scaling parameter of $r = 0.5$; for all t tests we used Cauchy priors with a scaling parameter of $r = \sqrt{2}/2$ (Rouder et al., 2009). All Bayes factors are estimated with errors less than 1%.

To analyse 2-AFC responses we used logistic mixed effects regression models. For the frequentist analyses we specified maximal random participant effects with intercepts and slopes for presentation time (Barr, Levy, Scheepers, & Tily, 2013). We tested each effect by comparing the full model with a model without the effect of interest by means of likelihood ratio tests. For specific contrasts we compared least square means (i.e., predicted marginal means) from the full model. To compute Bayes factors we used independent Cauchy priors with scaling parameters $r = 0.91$ for all experimental factors and $r = 1.28$ for the intercept (see Appendix G1; Gelman, Jakulin, Pittau, & Su, 2008). We chose these scaling parameters by transforming the scaling parameters used in the Bayesian ANOVA and t tests to logits (Eq. 7.3, Borenstein, 2009). In this analysis the intercept corresponds to the main effect of US valence, i.e., the inclination to select the positively paired CS rather than the negatively paired CS. We chose a wider prior for the intercept in correspondence with our priors in t tests on liking responses. Finally, we used the default uninformative priors specified by the *brms*-package (Bürkner, 2016) for all other (nuisance) parameters. In contrast to the frequentist analysis, we added a crossed maximal random effect for CS pairs with intercepts and slopes for all within-item effects (this additional random effect was omitted in the frequentist analysis to overcome convergence problems). We estimated Bayes factors as Savage-Dickey density ratios of prior distributions and maximum likelihood Gaussian density estimates of the posterior distributions (Morey, Rouder, Pratte, & Speckman, 2011).

We performed all analyses in R (Version 3.4.4; R Core Team, 2016)[15] and Stan (Carpenter et al., in press).

Results. Prior to our main analysis, we inspected correct responses in the online visibility check for CSs that had been presented for 85 ms or 1000 ms. We expected both stimulus durations to be long enough to allow supraliminal processing and substantially above chance identification performance. We, thus, examined performance in these conditions to identify inattentive or unmotivated participants. With eight CS options to choose from, random guessing would yield 12.5% correct responses. Using the Tukey boxplot outlier criterion, we excluded 9 participants from further analyses whose identification performance was below 12.50% at 85 ms or below

[15] We, furthermore, used the R-packages *afex* (Version 0.20.2; Singmann et al., 2016), *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2015), *papaja* (Version 0.1.0.9735; Aust & Barth, 2016), *RCurl* (Version 1.95.4.10; D. T. Lang & CRAN team, 2016), *rvest* (Version 0.3.2; Wickham, 2016), *tidyr* (Version 0.8.0; Wickham & Henry, 2017), and *yarrr* (Version 0.1.5; Phillips, 2017).

57.50% at 1000 ms.

Visibility.

The remaining 111 participants on average correctly identified 87.66% ($SD =$ 32.90) of CSs presented for 1000 ms and 64.50% ($SD = 47.86$) of CSs presented for 85 ms. Correct identification for CSs presented for 17 ms was on average 21.37% ($SD =$ 41.00), which was significantly above chance, $t(110) = 7.73$, $p < .001$, $BF_{10} > 1000$, $d = 1.75$, 95% HDI [1.45, 2.05].

Evaluation.

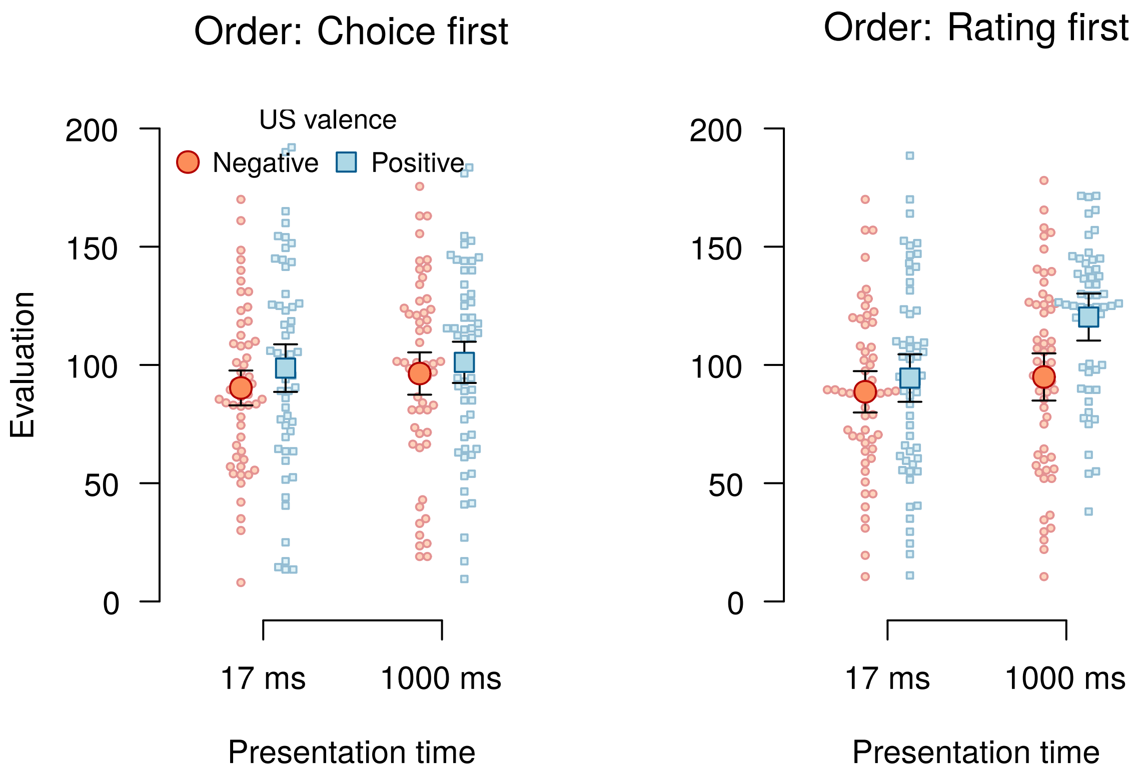

Figure 12. Evaluations of CSs in Experiment 1, split by the order of dependent variables, CS presentation time, and valence of the US paired with the CS. Error bars represent 95% within-subjects confidence intervals, dots represent participants' individual data points.

As can be seen in Figure 12, CSs paired with positive USs were preferred to CSs paired with negative USs ($BF_{10} = 28.59$, $F[1, 109] = 13.54$, $MSE = 1,003.34$, $p < .001$, $\hat{\eta}^2_G = .021$), and CSs presented for 1000 ms were preferred to CSs presented for 17 ms, $BF_{10} = 11.88$, $F(1, 109) = 8.33$, $MSE = 1,372.89$, $p = .005$, $\hat{\eta}^2_G = .018$. Descriptively, there was some indication of an interaction of US valence, presentation duration, and the order of dependent variables, but the evidence was ambiguous, $BF_{01} = 1.19$, $F(1, 109) = 3.13$, $MSE = 1,189.78$, $p = .080$, $\hat{\eta}^2_G = .006$.

We explored the data using two separate follow-up ANOVA to analyze each order-of-dependent-variables condition separately. When participants chose between CSs before rating them (i.e., choice-first), we found no significant effects of our experimental manipulations on evaluative ratings ($2.43 < BF_{01} < 4.61$ and $p \geq .12$). When participants rated CSs before choosing between them (i.e., rating-first), CSs paired with positive USs were preferred to CSs paired with negative USs ($BF_{10} = 25.33$, $F[1, 55] = 12.86$, $MSE = 1,055.84$, $p = .001$, $\hat{\eta}^2_G = .043$), and CSs presented for 1000 ms were preferred to CSs presented for 17 ms, $BF_{10} = 34.71$, $F(1, 55) = 9.15$, $MSE = 1,573.02$, $p = .004$, $\hat{\eta}^2_G = .045$. Most importantly, the effect of US valence depended on CS presentation duration, albeit the evidence was weak, $BF_{10} = 1.41$, $F(1, 55) = 4.10$, $MSE = 1,299.90$, $p = .048$, $\hat{\eta}^2_G = .017$. Follow-up tests did not indicate an effect of US valence when CSs were presented for 17 ms ($t[55] = -0.94$, $p = .350$, $BF_{01} = 2.75$, $d = 0.12$, 95% HDI [-0.14, 0.38]), but showed clear indication of an effect when CSs were presented for 1000 ms, $t(55) = -3.73$, $p < .001$, $BF_{10} = 112.39$, $d = 0.48$, 95% HDI [0.20, 0.75].

Choice.

CSs that were paired with positive USs were chosen more frequently than CSs paired with negative USs in the 2-AFC task, $\chi^2(1) = 14.41$, $p < .001$, $BF_{10} = 24.44$, $d = 0.25$, 95% HDI [0.10, 0.41]. We found weak evidence that the inclination to choose positively paired CSs was independent of the order of dependent variables, (i.e., there was positive evidence for the absence of an effect of order; see Figure 13), $BF_{01} = 4.34$, $\chi^2(1) = 0.00$, $p = .979$. The data did not provide evidence for or against an effect of presentation time on choice behavior, ($BF_{01} = 2.05$, $\chi^2[1] = 1.60$, $p = .206$); a possible effect of presentation time appeared to be independent of the order of dependent variables, albeit the evidence was weak, $BF_{01} = 3.28$, $\chi^2(1) = 1.34$, $p = .247$.

We further explored the effects of CS-US pairings on choice behavior in the condition where the 2-AFC task was administered first. Here, Bayesian and frequentist analyses yielded conflicting results. The Bayesian analysis testing whether positively paired CSs were chosen more frequently than negatively paired CSs was inconclusive, both for CSs presented for 17 ms ($BF_{01} = 1.88$, $d = 0.23$, 95% HDI [-0.06, 0.53]) and 1000 ms ($BF_{01} = 1.67$, $d = 0.25$, 95% HDI [-0.05, 0.56]). The frequentist analysis, however, indicated significant effects in both conditions, $z = 1.85$, $p = .032$ and

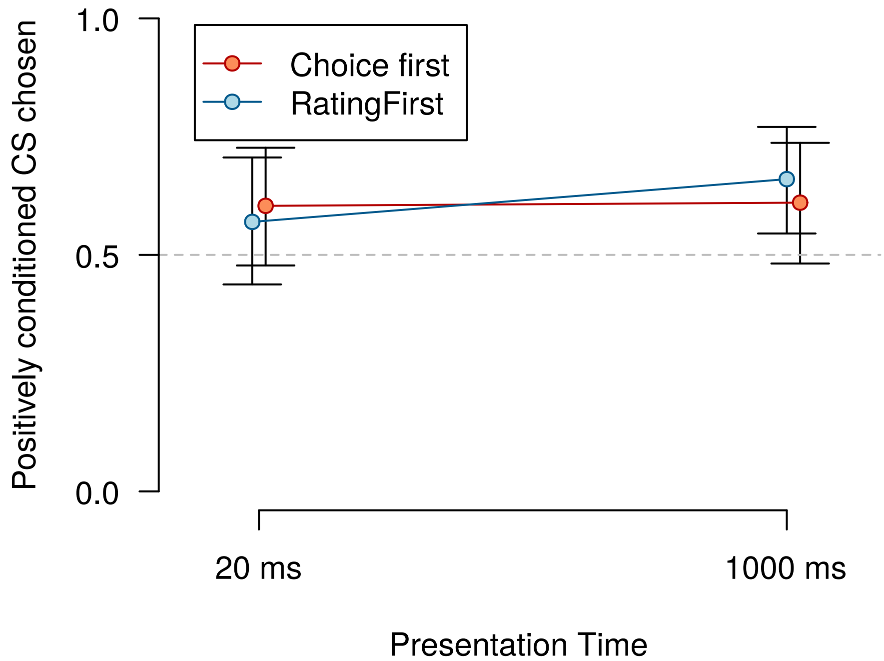

Figure 13. Rate of two-alternative forced-choice responses between positively and negatively conditioned CSs in Experiment 1. Higher values indicate a preference for the positively conditioned CS. Points and error bars represent Bayesian estimates of condition means and the corresponding 95% highest density intervals from the logistic mixed effects regression model.

$z = 2.07$, $p = .019$ for 17 and 1000 ms, respectively. We found no difference between the presentation time conditions, $BF_{01} = 5.53$, $z = -0.09$, $p = .927$.

Discussion Experiment 1. Extending the findings by Stahl et al. (2016), we did not find evidence for automatic processes in EC in the present cross-modal procedure. For the visual CS presented for 1000 ms, our findings replicate previous studies that have demonstrated cross-modal EC.

Our results indicate that the order of our dependent variables may have affected participants' evaluative ratings of the CSs (note though that the statistical evidence was ambiguous): Only when participants evaluated CSs before choosing between them did we find an EC effect for stimuli presented for 1000 ms. Yet, irrespective of the order of dependent variables, we observed no EC effect for CSs presented for 17 ms, despite the fact that they were clearly identified better than chance. The statistical evidence of the Bayesian analyses for the absence of EC for briefly presented CSs was, however, merely suggestive. Thus, we have addressed a potential limitation of previous studies that showed the absence of an EC effect for briefly presented CSs (i.e., Stahl et al., 2016), and have replicated this finding. Yet, because of the relatively weak evidence for this absence, our results need to be replicated before strong conclusions can be drawn.

The analysis of 2-AFC responses revealed an overall EC effect: Positively paired CSs were chosen more frequently than negatively paired CSs. This result indicates that the implemented 2-AFC task in which participants choose between positively and negatively paired CSs is suitable to assess EC. The preference for positively paired CSs appeared to be independent of the presentation time of CSs during the learning phase and the order in which the dependent variables were measured. However, when exploring choice responses of participants who performed the choice task before evaluating CSs, our data, again, did not provide conclusive evidence. Bayesian and frequentist analyses yielded conflicting results and, thus, should be interpreted cautiously. Nevertheless, the 2-AFC task appeared to be a valid measure of EC. A replication in a larger sample could shed more light onto the possibility that choice might be more sensitive than evaluative rating to detect potentially subtle EC effects for briefly presented CSs.

Limitations of Experiment 1.

Our findings challenge the existence of an automatic associative conditioning process underlying EC. One might, however, criticize the continuous assessment of CS visibility during the learning task. Whereas the visibility check provides strict control over awareness of CSs, it could be argued to interfere with the automatic conditioning process. Specifically, the visibility check may induce highly focused and deliberative stimulus processing and thus obscure the influence of unaware learning experiences. The deliberative and analytic mindset may have been further promoted by the instructions to memorize the CS-US pairs. In short, it is possible that our visibility check and instructions obstructed the automatic EC process.

Three additional aspects of our experimental design may constitute suboptimal conditions for obtaining an automatic EC effect for briefly presented CSs. First, our CSs were highly similar in shape and contrast to ensure equivalent visibility under brief viewing conditions. Similar stimuli may result in a confusability that could affect automatic associative processing of the CSs. Thus, the high similarity between briefly presented CSs may mitigate a potentially weak EC effect.

Second, the duration of USs used in Experiment 1 was relatively long (M =

3462 ms, *SD* = 359 ms) compared to the briefly presented CS. This mismatch in presentation duration may have resulted in a higher salience of the US compared to the CS, which, as suggested by Jones et al. (2010), might have inhibited automatic EC effects. Third, a simultaneous presentation onset of CS and US might not have resulted in a simultaneous processing of both stimuli: Specifically, the valence of the auditory US may have been extracted only after a delay that may have been longer than the entire presentation duration of the CS, such that the processing of the affective response may have been delayed to after the processing of the CS.

To sum up the findings of Experiment 1, while the paradigm is well-suited to study EC, it might not yet provide optimal conditions for automatic processes in EC by way of implicit misattribution. Thus, the results of Experiment 1 can be considered inconclusive for several reasons. In Experiments 2 and 3, we aimed at replicating and extend our findings to address methodological criticism.

Experiment 2

In Experiment 2, we used different materials and modified several aspects of the procedure. Experiment 2 aimed at replicating the pattern of EC effects across CS presentation duration levels on evaluative ratings as the sole dependent measure. Experiment 2 also implemented an on-line CS identification task; data from this task was to serve as reference for a final Experiment 3 in which the on-line CS identification task was omitted.

We used new CSs for which mean visibility under the presentation conditions was unknown, so we again used an online visibility check to estimate CS identification performance for these new CSs across critical presentation duration levels, and to find presentation conditions under which the new CSs were identified at slightly above-chance levels. Based on previous findings from our lab (Stahl et al., 2016), we selected duration levels of 20 ms and 30 ms, in addition to a 1000 ms condition.

Furthermore, we compared the simultaneous CS-US onset, which we used in Experiment 1, to a 400 ms delayed onset of the CS relative to the US. Using this onset asynchrony, the briefly presented CS was presented in the middle of the auditory US in order to create a stronger sense of simultaneous CS and US processing as compared to a mere simultaneous CS-US presentation onset. More specifically, we estimate that the affective response to be misattributed to the CS according to the implicit-misattribution account may take approximately 400 ms to develop: Research has shown that the pupillometric response to a valent sound starts after 400 ms (Partala & Surakka, 2003), indicating that the valence of the sound is actively processed after this interval. By using a 400ms stimulus onset asynchrony (SOA), we ensured that the CSs presented during that time are especially likely to be the target of implicit misattribution, and therefore, likely to show an automatic EC effect.

Method.

Design.

In Experiment 2 we manipulated the presentation time of the CS (20 ms vs. 30 ms vs. 1000 ms) as well as the valence of the US (positive vs. negative) within participants. Half of the CSs presented for 1000 ms had a simultaneous CS-US onset, the onset was delayed by 400 ms for the other half, and all CSs presented for 20 ms or 30 ms were delayed by 400 ms. We assumed that a delayed onset would be better suited to induce a simultaneous experience of CS and US when the CS is presented briefly. To test our intuition and ensure that a delayed onset yields an EC effect of at least the same magnitude, we compared the delayed and simultaneous onset in the long CS presentation condition. To focus our resources on the more critical evaluative ratings measure, we did not administer the 2-AFC task in Experiment 2.

Material.

CS.

As CSs, we used Pokémon cartoons similar to those used in previous studies (Stahl & Heycke, 2016). The images were changed to greyscale and contrasts were slightly reduced. 24 stimuli were pretested in a small pilot study (N = 27). Eight cartoons were selected based on mean evaluative ratings within 5% above or below the mean of the scale and the lowest standard deviation. The selected set had a mean evaluative rating of 98.69 and standard deviation of 36.33 on a 201 point slider scale (see Experiment 1).

For each participant, two different CSs were shown for 20 ms, for 30 ms, for 1000 ms with a delayed CS-US onset, and for 1000 ms with a simultaneous CS-US onset. Each CS was randomly assigned to one of the four conditions for each participant anew.

US.

We used a set of ten harmonious and ten disharmonious tones (Topolinski & Deutsch, 2012, 2013), which were cut into 800 ms segments, as USs. Eight of these tones have been successfully used in a previous cross-modal EC study (Gast et al., 2016) and we selected twelve additional similar tones from the set. Finally, we created ten neutral sounds that consisted of three metronome ticks in varying pitches.

Procedure.

Participants were seated in front of a 100 Hz CRT monitor. In contrast to Experiment 1, participants were not instructed to memorize CS-US pairs. They were instructed to pay attention to the sounds and images, and told that they would be asked to identify the images after every trial.

Each trial started with a blank screen for 500 ms, followed by a fixation cross shown for 500 ms. Then a black-and-white pixel mask was shown for 500 ms (which was randomly rotated by 0, 90, 180 or 270 degrees), followed by the CS presentation (duration was 20 ms, 30 ms, 1000 ms, or 80 ms for filler trials as in Experiment 1). As mentioned above, US presentation started 400 ms before the onset of the CS in the 20 ms, 30 ms, and half of the 1000 ms trials. For the other half of the 1000 ms trials, the onset of CS and US was simultaneous. The CS was immediately followed by a post-mask shown for 100 ms (which was again randomly rotated by 0, 90, 180 or 270 degrees, but at a different angle than the forward mask). Afterwards a blank screen was shown for 1000 ms.

Similar to Experiment 1 we included a visibility check after every trial. However, in Experiment 2 we gave participants only CSs that were shown for a similar presentation time (either for 20 ms and 30 ms or for 1000 ms) to choose from, in order to better control for guessing strategies. Participants may have correctly reasoned that a CS that was previously presented and correctly perceived at a longer presentation time may not be a plausible option in trials with a brief presentation time (such guessing strategies may have inflated previous estimates of stimulus visibility). Participants were again asked to select the CS they had seen during the trial from a list of these four CSs. If the CS was shown for 80 ms, two additional Pokémon images that were not used in the learning phase were given as distractors.

After evaluating the critical CSs as in Experiment 1, participants indicated how well they knew the Pokémon on a 4-point scale (*not at all*, *looks familiar*, *know it*, *know it very well*) and answered the same demographic questions as in Experiment 1.

Additionally, we asked participants whether they had worn the headphones during the entire procedure.

Sample.

We used a sequential Bayesian analysis to analyze the incoming data (Rouder, 2014). We started analyzing the data after collecting 30 participants to reduce the chance of false positive evidence (Schönbrodt et al., 2015) and analyzed the data after every day of data collection. Our main analyses consisted of two t tests for EC effects in the 1000 ms condition separately for the simultaneous and delayed CS-US onset. We planned to collected data until the Bayes factor for both EC effects was larger than 5 or until a maximum of 55 participants was collected.

These criteria were met, and data collection was therefore stopped, after 46 participants. Participants were mostly University of Cologne students and received partial course credit or two Euro for their participation. 5 participants were excluded due to poor performance on the visibility check at 80 and 1000 ms (see Experiment 1 for details). Three additional participants were excluded because they reported that they did not wear the headphones during the entire procedure. Two additional participants reported for at least one CS, that they knew it very well and were therefore excluded. We, thus, included 37 participants in the final analysis (Age $M = 23.51$, $SD = 3.83$, 26 female).

Data analysis.

We used Bayes Factors for inference in all analyses as described in Experiment 1. We again performed repeated-measures ANOVA to analyse liking responses and paired t tests to examine interaction effects. We only report Bayesian analyses because p values are uninterpretable when the sampling plan is not clearly defined a priori, as was the case in our design (e.g. Chapter 11, Kruschke, 2015).

Results.

Visibility.

One goal of Experiment 2 was to ensure that the new stimulus material was identified at slightly above-chance levels. Participants were asked to identify the CS from an array of four CSs, which resulted in a chance level of .25. One stimulus was correctly identified in 90 % of cases in the 20 ms condition and in 93.33% of cases in the 30 ms condition. We excluded this stimulus for all further visibility analyses.

The average correct identification rate in the 20 ms condition was 0.45, ranging from 0.36 to 0.54, which was clearly above chance, $BF_{10} = 932.30$, $d = 1.77$, 95% HDI [1.24, 2.30]. The CSs showed highly similar standard deviations, ranging from 0.48 to 0.50. Similarly, in the 30 ms condition the average correct identification rate was 0.57, ranging from 0.46 to 0.80, which was significantly above chance, $BF_{10} > 1000$, $d = 2.22$, 95% HDI [1.62, 2.86] (standard deviations in the 30 ms conditioned ranged from 0.40 to 0.50).[16]

Evaluation.

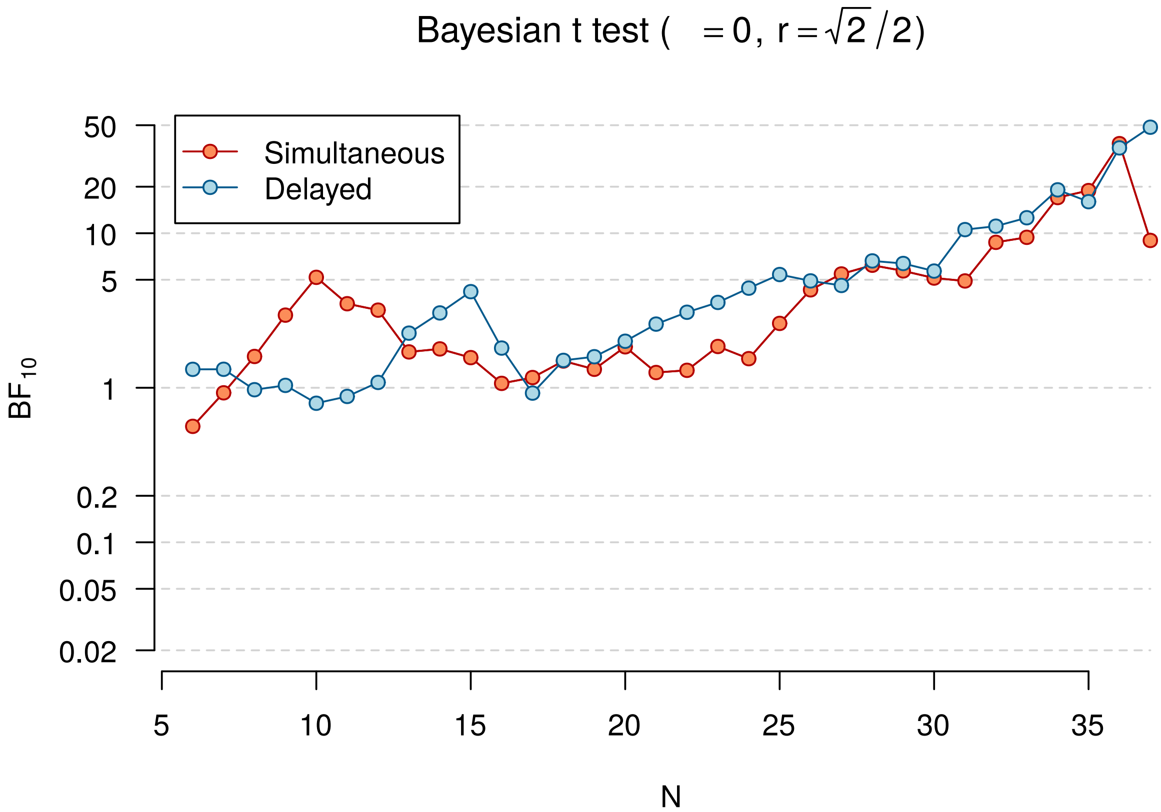

Figure 14. Sequential Bayes factor of t tests of the 1000 ms condition in Experiment 2 (with the visualization starting after five participants).

As shown in Figure 14, Bayes factors for the simultaneous ($BF_{10} = 9.86$, $d = 0.43$, 95% HDI [0.10, 0.76]) and delayed CS-US onset ($BF_{10} = 49.94$, $d = 0.54$, 95% HDI

[16]One of the CSs was highly visible even with brief presentations. In an attempt to reduce its visibility, we modified the image file to reduce the contrast of that stimulus. A small pilot study ($N = 7$), where we presented all 8 CSs for 20 ms ten times and asked participants to identify the CS from a list of all 8 CSs, showed that the visibility of this stimulus was substantially reduced but still elevated compared to the other CSs (mean visibility = 0.66).

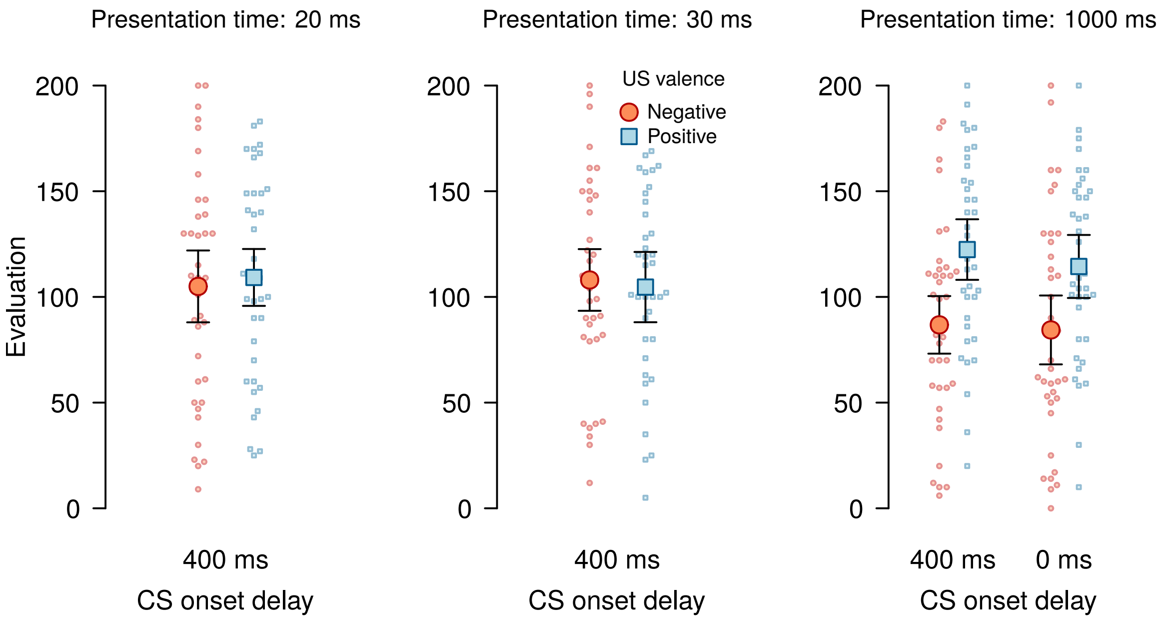

Figure 15. Evaluations of CSs in Experiment 2, split by the presentation time (20 ms, 30 ms or 1000 ms) and the SOA of CS and US (delayed onset by 400 ms of the CS or simultaneous onset of CS and US for one 1000 ms condition only). Error bars represent 95% within-subjects confidence intervals, dots represent participants' individual data points.

[0.20, 0.88]) were larger than 5, which was our stopping rule, and both developed similarly over the course of data collection. The data, moreover, indicated that there was no difference between the delayed and simultaneous onset in the EC effect for CSs presented for 1000 ms, $BF_{01} = 3.95$, $\hat{\eta}^2_G = .003$.

We computed an ANOVA for presentation time (20 ms, 30 ms, 1000ms) $\times$ US valence (positive, negative) and found evidence for an interaction between the two factors, $BF_{10} = 5.17$, $\hat{\eta}^2_G = .029$, see also Figure 15. As in Experiment 1, we found an EC effect only for CSs presented for 1000 ms and not for briefly presented CSs: There was a robust EC effect only for CSs presented for 1000 ms, $BF_{10} = 394.81$, $d = 0.66$, 95% HDI [0.30, 1.02]. In contrast, there was some evidence for the absence of an EC effect for CSs presented for 20 ms, $BF_{01} = 4.12$, $d = 0.06$, 95% HDI [-0.25, 0.36], as well as for CSs presented for 30 ms, $BF_{01} = 7.01$, $d = -0.05$, 95% HDI [-0.35, 0.27].

Discussion Experiment 2. Replicating Experiment 1 with different CS and US materials, Experiment 2 found a cross-modal EC effect for CSs presented for 1000 ms, but not for those presented for 20 or 30 ms. Additionally, there was no detrimental effect of the 400 ms SOA on EC for supraliminal CSs: Based on the findings of Experiment 2, we conclude that a slight CS-US onset asynchrony does not affect the EC effect at long presentation times compared to a simultaneous CS-US onset. We speculate that using a 400 ms-delayed CS-US onset may induce a stronger experience of simultaneous presentation for briefly presented CSs compared to a simultaneous onset. We propose that the delayed CS-US onset may provide more favorable conditions for EC effects with briefly presented CSs, as it might increase the chances of affective responses to the US being experienced simultaneously with the CS presentation.

The visibility of the briefly presented CSs was clearly above chance even for stimuli presented for 20 ms. This was expected, given that the CSs used in Experiment 2 were more easily discriminable than those used in Experiment 1, and given that CS duration was slightly longer than the 17 ms implemented in Experiment 1. CSs were identified above chance, but the correct identification rates were much lower than for stimuli presented for 80 ms or 1000 ms. Presenting the CSs for 20ms therefore yields a suboptimal yet clearly not invisible method of presentation given the stimuli and procedure at hand.

Finally, there was no indication for an EC effect for briefly presented CSs. Both in the 20 ms and in the 30 ms condition, the data suggested that a near-threshold EC effect was absent. Although we made procedural changes to create more favorable conditions for automatic processes in EC compared to Experiment 1 (i.e., we eliminated instructions to memorize CS-US pairs, high similarity between CSs, and simultaneous CS-US onsets), we were not able to find an EC effect with briefly presented CSs. But, again, caution is warranted when drawing conclusions about automatic processes in EC based on the findings of Experiment 2 alone. The evidence for the absence of an EC effect was moderate and, as mentioned above, we cannot preclude that the online visibility check interfered with an automatic processing of the pairings. We therefore aimed at a replication of Experiment 2 with a larger sample and including again the potentially more sensitive 2-AFC measure.

Registered Experiment 3

The goal of Experiment 3 was to replicate the findings of Experiment 1 and 2 in a sample large enough to detect medium to small EC effects for near-threshold CSs, or conversely to accumulate conclusive statistical evidence for the absence of such an effect. The only major difference to Experiment 2 was that we did not administer the visibility check throughout the learning task to avoid inducing an analytic deliberative mindset that may have obstructed automatic processes in Experiments 1 and 2 (see also Stahl et al., 2016). As in Experiment 1, we additionally assessed CS preferences in a 2-AFC choice task. In a high-powered study, this inclusion will contribute to answering the question whether choice is a more sensitive measure to detect small changes in preference as compared to evaluative ratings. From the visibility data of Experiment 2, we can conclude that the visibility of the briefly presented CSs was above chance even at 20 ms. Given that CSs are identified above chance, an absence of an EC effect for briefly presented CSs would be highly informative because we have attempted to create optimal conditions under which dual-process theories of EC would predict an automatic EC effect.

If, on the other hand, we observed an EC effect with briefly presented CSs, the results would contradict findings by Stahl and colleagues (Stahl et al., 2016) who found no EC effect for near-threshold but above-chance visual CSs. Such a result would call into question their conclusion that there is no automatic EC effect, and it would lend credibility to the critique that the absence of EC in previous studies may have been caused by the experienced delay between CS and US, by a focused, analytic stimulus processing mindset induced by instructions to memorize CS-US pairings, or by interference from the online visibility check. Thus, Experiment 3 will provide a stringent and fair test of the predictions of single- versus dual-process models of EC.

Procedure. The procedure was the same as in Experiment 2 except that all briefly presented CSs were displayed for 20 ms and all CS onsets were delayed by 400 ms relative to USs. The study, thus, followed a 2 (*presentation time*: 20 ms or 1000 ms) $\times$ 2 (*US valence*: positive or negative) $\times$ 2 (*DV order*: 2-AFC or rating first) design; the first two factors were manipulated within participants. The order of dependent variables was counterbalanced across participants.

We substituted the visibility task with a target-detection task similar to the one used by Olson and Fazio (2001) in the surveillance paradigm. In Experiment 3, one of the two filler CSs was randomly selected to serve as the target stimulus for this task. It was presented for 200 ms (while the other filler stimulus was presented for 80 ms), and participants were instructed to press the space bar as quickly as possible whenever it appeared on screen during the learning phase. This task achieves the goal of focusing participants' visual attention on the CSs in a manner comparable to the CS identification task while being much less resource-demanding and considerably easier to perform. Most importantly, the task has repeatedly been shown to successfully support EC effects in incidental EC procedures that are supposedly due to the operation of automatic associative processes (Olson & Fazio, 2001).

Material.

Based on the combined visibility data from Experiment 2 and the small pilot test administered afterwards, we decided to select the four most clearly visible CSs to be shown for 1000 ms only and the four CSs with lowest visibility to be presented for 20 ms only (for an overview of the visibility rates see Appendix H). Note that all CSs were identified at above-chance levels in both data sets.

Analysis plan. Before conducting our analyses, we excluded participants who (a) reported for at least one of the Pokémon CSs that they “know it very well” (because pre-experimental preferences towards the stimuli are typically not affected by EC); (b) reported that they did not wear the headphones during the learning task; (c) missed too many targets in the ongoing task (participants were excluded based on Tukey’s outlier criterion, but only if that criterion reaches 3 misses, thereby allowing for 1 or 2 lapses during the learning phase); (d) aborted the experiment prematurely; or (e) explicitly reported other major (but unforeseeable) issues during the experiment that hindered them to participate in an orderly fashion (e.g., loud noises during experiment, difficulty understanding the instructions; based on our experience from previous studies we expect at most one of such cases to occur).

After the initially set N was reached, we analyzed the data using frequentist and Bayesian data analyses. All further incoming data were analyzed using a sequential Bayesian analysis (Rouder, 2014; Schönbrodt et al., 2015).

Planned sample size.

We first collected a fixed number of 122 participants based on the same a-priori power analysis as described for Experiment 1, which ensured adequate power for a one-sided paired t-test with $\alpha = \beta = .05$ and medium-to-small effects of $d = 0.3$ that are at the lower end of typical incidental EC effect sizes. A minimal $N = 122$ should also help to minimize the chance of a false positive finding in the sequential Bayesian analyses (Schönbrodt et al., 2015). We continued data collection until all targeted Bayes factors had reached (or exceeded) 10 or greater (or less than 1/10 when the inverse Bayes factor is considered), or until we ran out of money (i.e., participants were paid 2€ for the short study, and funds were available to pay a maximum of $N = 300$ participants). The total N might be higher than 300 because some participants may opt to participate in exchange for partial course credit. After the minimal sample size was reached, the data were analysed after every day of data collection.

Stopping rule.

We planned to stop data collection when the evidence for both EC effects (20 ms and 1000 ms) was conclusive, both for evaluative ratings and the choice measure. That is, we monitored four Bayes Factors, and we stopped data collection when, in each of the four cases, either $BF_{01} > 10$ or $BF_{10} > 10$.

Confirmatory analysis.

In our confirmatory Bayesian analyses, which were also the basis of our data collection stopping rule, we first tested if the order of dependent measures affects the outcome of our experimental manipulations. Only when finding compelling evidence that the order of dependent measures did not affect results ($BF_{01} > 10$) we ignored the factor of dependent measure order. If we observed inconclusive statistical evidence, or if we found evidence that the order of dependent measures affected our results (BF_{10} > 10), we would analyze the evaluation data only for the subset of participants who first evaluated the CSs, and vice versa for the choice data analysis. For the frequentist analyses planned to be conducted on the initial fixed sample size of $N = 122$, we would ignore the factor of dependent measure order if we did not find a significant effect of the order on our results.

Initial confirmatory analysis. To probe for the presence or absence of EC effects within each level of the presentation time factor (20 ms and 1000 ms), we calculated one-tailed (Bayesian and frequentist) paired t-tests. We used the same priors and settings as in Experiment 1 and 2. For the initial analysis, we used the first 122 valid data sets (i.e., continued data collection if we had to exclude participants for any of the reasons given above until $N = 122$ was reached).

Evaluation.

Since the results of the frequentist and Bayesian analyses diverged, we will report them separately.

Frequentist analysis.

There were no indications for an interaction of the measurement order and the US valence, $F(1, 120) = 2.47$, $MSE = 1,174.34$, $p = .118$, $\hat{\eta}^2_G = .004$, nor the measurement order and the US valence and CS presentation time, $F(1, 120) = 0.63$, $MSE = 1,335.92$, $p = .429$, $\hat{\eta}^2_G = .001$. We therefore used the data of all participants (i.e., participants who did the evaluation task first and participants who did the choice task first) for further analyses of the evaluation. Of the CSs presented for 1000 ms, those shown with positive USs were evaluated more positively than those presented together with negative USs, $t(121) = 4.72$, $p < .001$, $d = 0.43$. There were no indications for an EC effect for CSs presented for 20 ms, $t(121) = 0.51$, $p = .307$, $d = 0.05$.

Bayesian analysis.

The Bayesian analyses on the potential influence of the order of measurement did not yield sufficient evidence to ignore the factor of measurement order. This was true for the three way interaction of order × presentation time × US valence, $BF_{01} = 3.99$, $\hat{\eta}^2_G = .001$, and the interaction between US valence and order, $BF_{01} = 2.58$, $\hat{\eta}^2_G = .004$. For the following analyses, we therefore only used the data of participants who performed the evaluation task before the choice task ($N = 59$). As in the frequentist analyses, for the CSs presented for 1000 ms a clear EC effect was found, $BF_{10} = 498.39$, $d = 0.53$, 95% HDI [0.26, 0.80]. There was however, no evidence for nor against an EC effect for CSs presented for 20 ms, $BF_{10} = 1.20$, $d = 0.22$, 95% HDI [-0.03, 0.47]. Since both of these Bayes factors were part of the sequential stopping rule, we continued data collection after the initial data analysis.

Choice.

We analyzed 2-AFC responses using logistic mixed effects regression models, using the same specifications as in Experiment 1 for both frequentist and Bayesian analyses. In these analyses, in which the proportion of choices of the CS paired with positive USs is the dependent variable, the intercept can be interpreted as a main effect of US valence, corresponding to an inclination to select positively paired CSs over negatively paired CSs. A main effect of presentation time would be equivalent to a modulating effect on EC, indicative of a stronger EC effect for one of the two levels of the presentation-duration factor. We planned to assess the respective intercept terms to probe for the presence of EC effects on choice in each level of the presentation-duration factor.

Frequentist analysis.

No significant interaction of order and US valence, $\chi^2(1) = 0.03$, $p = .858$, nor a significant three-way interaction of the order of dependent measures $\times$ US valence $\times$ presentation time of the CS was found, $\chi^2(1) = 0.36$, $p = .549$. We therefore ignored the order factor and analyzed the data of all participants. Of the CSs presented for 1000 ms, those conditioned with positive USs were chosen more often than those paired with negative USs, $z = 3.87$, $p < .001$. For CSs presented for 20 ms, there was no indication for a preference for CSs paired with positive USs, $z = -0.41$, $p = .685$. The results of the choice variable therefore reflected the result of the evaluation variable reported above.

Bayesian analysis.

Similar to the analyses of the evaluation data, the Bayes Factors for the interaction of the US valence and the order of dependent measures, $BF_{01} = 4.40$, as well as the interaction of US valence $\times$ order of dependent variables $\times$ CS presentation time, $BF_{01} = 3.88$, did not reach our pre-set level of a Bayes Factor larger 10 for the null hypothesis, which would have allowed us to ignore the order of dependent variables. We therefore only analyzed the data of participants who performed the choice task before the rating task ($N = 63$). Even through, descriptively, CSs shown for 1000 ms and conditioned with positive USs were more likely to be selected, the Bayes Factor was inconclusive, $BF_{10} = 2.64$, $d = 0.37$, 95% HDI [0.06, 0.69]. The Bayes Factor was more conclusive for stimuli presented for 20 ms: The data supported the null hypothesis of no preference for CSs conditioned with positive USs, $BF_{01} = 12.01$, $d = 0.03$, 95% HDI [-0.23, 0.28].

Since we did not find conclusive evidence for all four target Bayes Factors in our initial analysis, we continued data collection and analyzed the data after every day of data collection.

Confirmatory analysis after sequential testing. During the sequential data collection, one of the target Bayes Factors did not reach our pre-set level (i.e., > 10 or $< 1/10$), and we therefore collected the previously set maximum data of 300 paid participants (363 participants in total, including participants who took part for partial course credit). Two participants took part twice and the second data set was removed. The data of one participant was removed because she reported that she did not wear the headphones and two participants were excluded due to technical difficulties. Additionally, 42 participants were excluded because they did not press the space bar at least 8 times when the target was shown and additional 56 participants were excluded because they reported to know at least one Pokémon very well. Consequently 260 participants are included in the analysis (age M = 23.58, SD = 6.03; 208 female).

Evaluation.

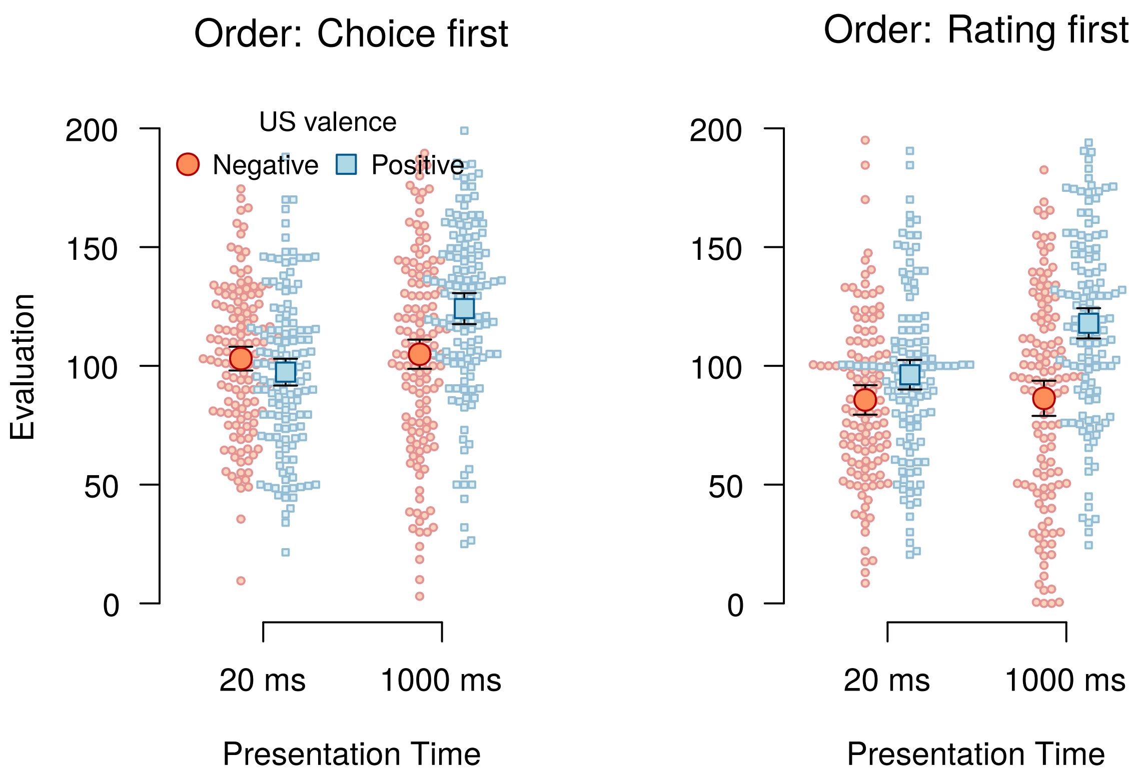

Figure 16. Evaluations of CSs in Experiment 3, split by the order of dependent variables, CS presentation time, and valence of the US paired with the CS. Error bars represent 95% within-subjects confidence intervals, dots represent participants' individual data points.

Since we observed an interaction between US valence and the order of dependent measures, $BF_{10} = 17.74$, $\hat{\eta}^2_G = .009$, we analyzed only the data of participants who performed the evaluation task before the choice task (N = 125). As in the initial analysis, we observed an EC effect for CSs presented for 1000 ms, $BF_{10} > 1000$, $d = 0.59$, 95% HDI [0.40, 0.78]. Interestingly, we also observed an indication for the presence of an EC effect for briefly presented CSs, $BF_{10} = 6.88$, $d = 0.24$, 95% HDI [0.06, 0.41], see Figure 16 (and see Figure 17 for a visual comparison of all EC effects). Even though this Bayes Factor did not reach our pre-set thresholds, and the effect

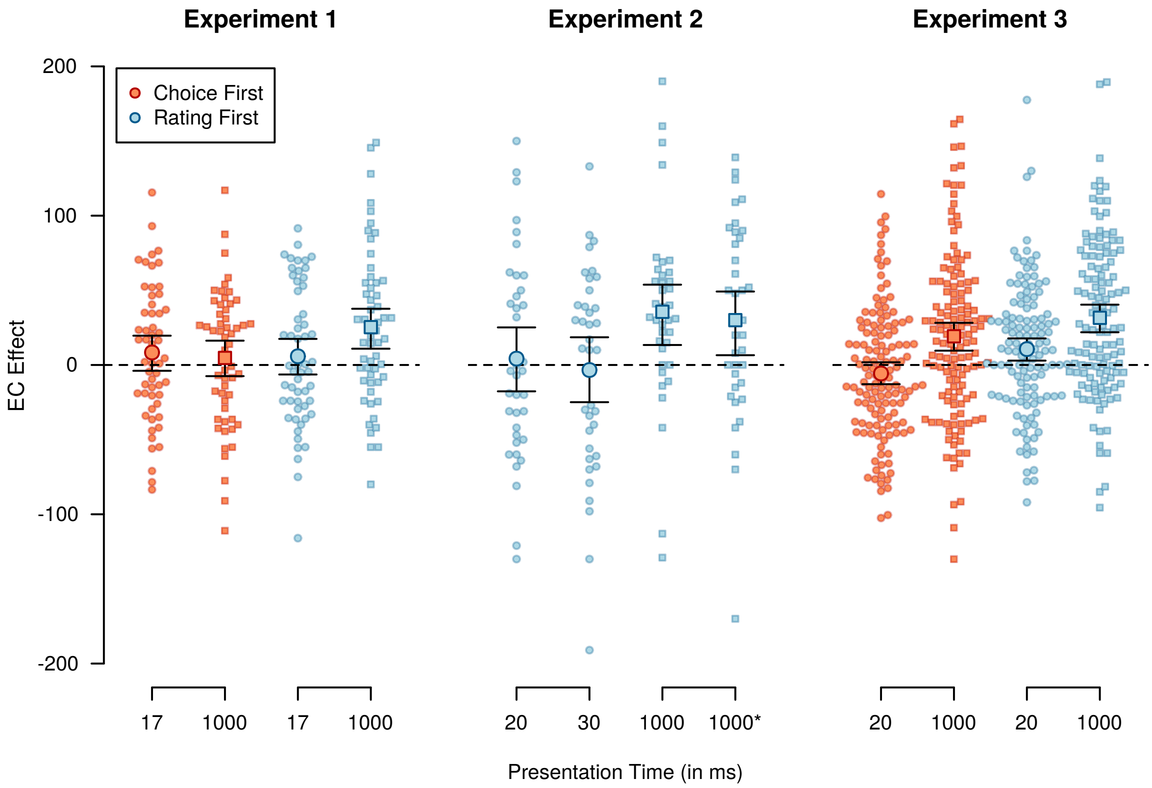

Figure 17. EC effects from all three Experiments in the evaluative rating, split by the order of dependent variables and CS presentation time. Values above 0 indicate a more positive evaluation of CSs paired with positive USs than paired with negative USs. Large dots represent the mean EC effect and the corresponding error bars the 95% highest density intervals. Small dots represent participants' individual EC effects. * = simultaneous CS-US onset.

size is small, the results indicate that even when CSs were presented for only 20 ms, participants evaluated CSs presented with positive USs as more positive than CSs presented with negative USs.

Choice.

As with the initial analysis, we did not find sufficient evidence—according to our pre-set thresholds—for the absence of an influence of the order of dependent measures (Order × US valence: $BF_{01} = 6.68$, $d = -0.02$, 95% HDI [-0.10, 0.07]; Order × US valence × CS presentation time: $BF_{01} = 6.70$, $d = 0.02$, 95% HDI [-0.07, 0.10]) and therefore only analyzed the participants who completed the choice task before the evaluation task (N = 135). After the sequential analysis, we observed statistical evidence for a preference for CSs conditioned with positive USs, when the CSs were

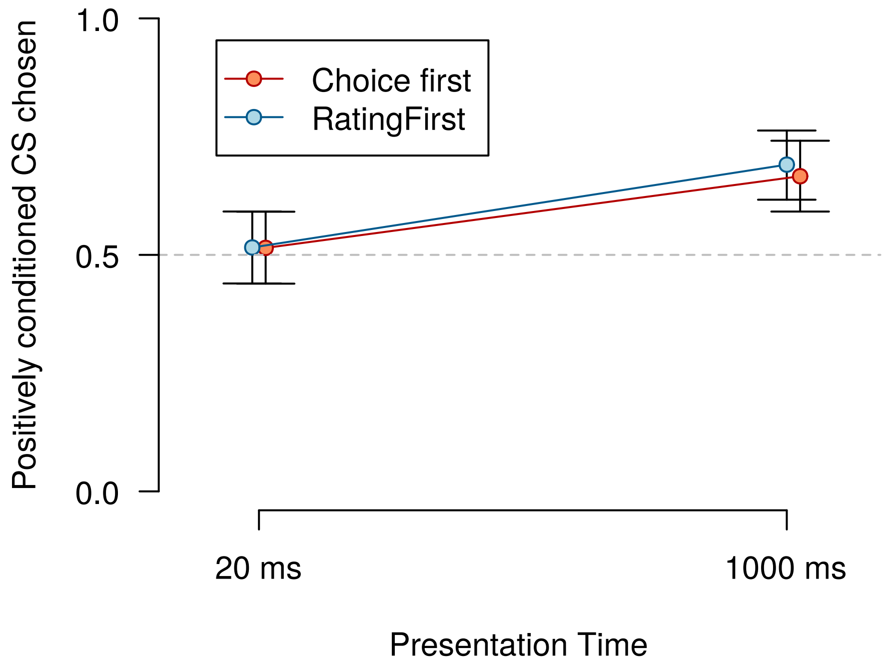

Figure 18. Rate of two-alternative forced-choice responses between positively and negatively conditioned CS in Experiment 3. Higher values indicate a preference for the positively conditioned CS. Points and error bars represent Bayesian estimates of condition means and the corresponding 95% HDI from the logistic mixed effects regression model.

presented for 1000 ms, $BF_{10} = 339.31$, $d = 0.38$, 95% HDI [0.20, 0.57]. When the CSs were presented for 20 ms, we found evidence for the absence of a preference for CSs conditioned with positive USs, $BF_{01} = 15.12$, $d = 0.03$, 95% HDI [-0.13, 0.20], see Figure 18. Note that the EC effects on the choice variable were consistently smaller than the EC effect on the evaluation variable.

Exploratory analysis. In a first set of additional—not registered—analyses, we explored the sensitivity of our results to our pre-registered participant exclusion criteria. In a second set of analyses, we explored the interaction between our two dependent variables.

Different participant exclusion criteria.

In a first analysis we focused on the small EC effect on evaluative ratings obtained for CSs presented for 20 ms; this effect was based on the subset of participants who completed the evaluation task before the choice task. Here, we again analysed only the evaluation data for those participants, and we again excluded all participants who did not press the space bar at least 8 times. The only difference to our confirmatory

Table 2

Effect of different exclusion criteria on the evaluation (evaluation first) and choice (choice first) dependent variables

Presentation Time	Registered	Include Space	Include Knowing	Include Both
Evaluation				
1000 ms BF_{10}	> 1000	> 1000	> 1000	> 1000
20 ms BF_{10}	6.88	2.77	2.45	1.48
N	125	144	154	179
Choice				
1000 ms BF_{10}	339.31	> 1000	> 1000	> 1000
20 ms BF_{10}	1/15.12	1/14.66	1/19.01	1/18.16
N	135	149	162	179

Note. In the 'Include Space' analyses only participants who reported to know at least one of the Pokemon very well were excluded. In the 'Include Knowing' analyses only participants who did not press the space bar at least 8 times were excluded.

analyses was that we only excluded those participants who reported to know any of the *Pokémons shown for 20 ms* very well (in the confirmatory analysis, a participant was excluded as soon as she indicated knowing *any* Pokémon very well). In effect, this analysis included 137 participants (instead of 125 in the confirmatory analysis). Including participants who report to know any of the Pokémon shown for 1000 ms very well—but none of the Pokémon presented for 20 ms—should not have an effect on the evaluation of CSs in the 20 ms condition. However, the results of the evaluation analysis for CSs presented for 20 ms (evaluation-first only) was now inconclusive, $BF_{10} = 1.40$, $d = 0.17$, 95% HDI [0.00, 0.33] (i.e., there was no evidence for a preference for positively paired CSs over negatively paired CSs anymore). Even though adding these participants to the analysis should not influence the ratings for CSs presented for 20 ms, the Bayes Factor was reduced and we did not find any statistical evidence for the alternative hypothesis over the null hypothesis.

We additionally explored the effect of relaxing all previously set exclusion criteria (i.e., include participants who did not press the space bar often enough, include participants who knew at least one of the Pokémon very well, or both) on the choice measure as well as the evaluation measure (see Table 2), with the following results: For the choice measure, relaxing any of the exclusion criteria did not affect the pattern of results. Similar, the conclusions remain unaltered for EC effects on evaluative ratings of CSs presented for 1000 ms. Results changed only for evaluation of

CSs presented for 20 ms; here, the Bayes Factor was smaller—and inconclusive—when additional participants (other than those in the registered analysis) were included. It should however be noted that, in theses analyses, we also found no statistical evidence for the absence of an EC effect for CSs presented for 20 ms.

Influence of order of the dependent variables.

Choice.

First we checked whether the choice results were affected by the order of dependent measures. We analyzed the data of the evaluation-first group and found evidence for the absence of an EC effect in the 20 ms condition, $BF_{01} = 13.78$, $d = 0.06$, 95% HDI [0.04, 0.09], but evidence for EC in the 1000 ms condition, $BF_{10} > 1000$, $d = 0.43$, 95% HDI [0.25, 0.62]. This pattern of results was equivalent to the choice-first group. Unsurprisingly, when we analyzed the entire sample and ignored the order of dependent variables, we found evidence for the absence of an EC effect in the 20 ms condition, $BF_{01} = 18.45$, $d = 0.03$, 95% HDI [-0.09, 0.15], but evidence for an EC effect in the 1000 ms condition, $BF_{10} > 1000$, $d = 0.41$, 95% HDI [0.27, 0.55].

Evaluation.

As reported above, we observed a clear indication for an interaction between US valence and the order of dependent measures in the evaluation task, $BF_{10} = 17.74$, $\hat{\eta}^2_G = .009$. There was an indication for the absence of the interaction between US valence, the order of dependent measures, and the presentation time of CSs, $BF_{01} = 6.81$, $\hat{\eta}^2_G = .000$. Interestingly, for the CSs shown for 1000 ms the interaction between US valence and order of dependent variables was inconclusive, $BF_{01} = 1.37$, $\hat{\eta}^2_G = .006$, but we found clear indications for an US valence × order of dependent variables interaction for CSs presented for 20 ms, $BF_{10} = 12.08$, $\hat{\eta}^2_G = .015$. This interaction for CSs presented for 20 ms reflects the finding that the EC effect was larger when the evaluation was administered before the choice task, than when it was administered after the choice task, $BF_{10} = 9.91$, $d = 0.36$, 95% HDI [0.12, 0.60] (a parallel comparison for CSs presented for 1000 ms yielded inconclusive results, $BF_{01} = 1.48$, $d = 0.22$, 95% HDI [-0.02, 0.45]).

Discussion Experiment 3. The main goal of the third experiment was to investigate whether we can find EC effects with brief (20 ms) and longer (1000 ms) CS presentation times when participants do not have to answer to a visibility check after every trial. For an evaluation of CSs after the learning phase as well as a 2-AFC task, we found that CSs presented for 1000 ms and paired with positive USs were preferred over CSs presented for 1000 ms and paired with negative USs. On the evaluative rating measure, we also found some evidence that CSs presented for 20 ms and paired with positive USs were rated as more positive as CSs presented for 20 ms and paired with negative USs. Interestingly, this pattern was not found in the choice task: Here, CSs presented for 20 ms were as likely chosen when they were paired with positive USs as when they were paired with negative USs. We additionally observed that the EC effect on evaluative ratings in the 20 ms condition was modulated by the choice task, in that the EC effect was smaller after the choice task compared to the effect found when the evaluation was administered before the choice task. In the following section, we will discuss the finding of an EC effect for CSs presented for 1000 ms, the diverging findings of EC effects for CSs presented for 20 ms, and the potential interference of the choice task on the rating task.

The finding of an EC effect in the evaluation of CSs presented for 1000 ms replicated the previous two experiments, as well as previous studies on cross-modal EC with stimuli that can be consciously perceived. The result therefore shows that the present paradigm is capable of reliably demonstrating cross-modal EC effects in evaluative ratings. Results were somewhat more mixed with regard to the choice measure: In Experiment 1, the evidence for an EC effect for CSs presented for 1000 ms in the 2-AFC task was not conclusive. In Experiment 3, however, we found compelling evidence that CSs presented for 1000 ms and paired with positive USs were preferred over CSs presented for 1000 ms and paired with negative USs. This finding shows that evaluative conditioning might be a useful tool to influence decision-making behavior, and it lends ecological validity to the EC phenomenon. Future studies in an applied setting could further build on this finding.

In contrast to CSs presented for 1000 ms, we did not find a preference for positively paired CSs presented for 20 ms in the 2-AFC task. We can only speculate whether this is due to the fact that either it is simply not possible to influence participants' decision-making behavior with briefly presented CSs, or whether the choice task was not sufficiently sensitive to detect small effects. Based on previous findings (Verwijmeren et al., 2012), we had originally speculated that choice might be a more sensitive measure of preferences than evaluations. In both Experiment 1 and Experiment 3—which used both choice and evaluation as dependent measures—effect sizes in the evaluation task were larger than effect sizes in the choice task. In light of these findings, we would argue that choice might not be a more sensitive measure than evaluation.

Consistent with this interpretation, there was some evidence for an EC effect in the 20 ms condition in the evaluation (but not the choice) task. For the interpretation of this effect, it should be noted that—as in the Study by Stahl et al. (2016)—the briefly presented CS stimuli were not truly subliminal but in fact visible at above-chance levels. We therefore refrain from claiming that we have found an EC effect for stimuli that were truly subliminal. Nevertheless, the present indication for an EC effect for stimuli presented only very briefly contradicts previous findings (Stahl et al., 2016) suggesting that EC effects require much longer CS presentation durations. There are a few factors that might explain this discrepancy: In contrast to the experiments by Stahl et al. (2016), the present studies (1) did not include a visibility check after each trial, and (2) they implemented a cross-modal EC procedure which might have allowed for a more simultaneous experience of CS and US (Jones et al., 2009). Combined, these changes might have resulted in the formation of an EC effect even with a very brief presentation time. It has to be noted, however, that the effect size found in this study was very small; this could explain why previous studies (e.g., Experiment 1 and Experiment 2) did not find the effect. Another important point is that, in contrast to the effect for CSs presented for 1000 ms, the effect in the 20 ms presentation time condition was not robust to additional—exploratory—analyses. Including only a few additional participants (i.e., those who knew one or more of the CS stimuli presented for 1000 ms but not those presented for 20 ms) sharply reduced the statistical evidence for an EC effect in the 20 ms condition (i.e., the Bayes Factor indicated merely anecdotal evidence). Nevertheless, the EC effects on the evaluation of CSs presented for 20 ms are interesting and give reason to invest additional efforts into the question whether a cross-modal setting to achieve simultaneous presentation of CS and US—as used in this paper—might be well-suited for automatic EC effects.

It is additionally worth discussing the order effect of our dependent variables. For CSs presented for 20 ms, the evaluation EC effect was smaller when evaluation was administered after (as compared to before) the choice task. One possible explanation for this pattern of results is a consistency effect in the choice-first group: participants may have evaluated the CSs in line with their previous choices; if, as our results indicate, there was no EC effect on choice, these choices were likely to be random (i.e., equally likely to be consistent as inconsistent with US valence), and this may have masked, or interfered with, the small EC effect on evaluations. This is in line with research showing that choices can influence the preference of initially equally valued items (Brehm, 1956). Note, however, that this finding has recently been critically discussed, and possible boundary conditions were suggested for the influence of choice on preferences (Voigt, Murawski, & Bode, 2017).

General Discussion

We set out to test the hypothesis that simultaneous CS-US presentation might be beneficial for EC with briefly presented CSs. To achieve a simultaneous CS-US presentation that guaranteed that the visual attention could be directed towards the briefly presented CS, we implemented a cross-modal paradigm with auditive USs and visual CSs. In the first two studies, we did not observe an EC effect for briefly presented CSs: If anything, we observed some evidence for the *absence* of an EC effect for briefly presented CSs. However, interference arising from the CS identification task—which was prompted after every trial in the learning phases of Experiment 1 and 2—, as well as lack of statistical power, could be reasons for not finding an EC effect in these conditions. We therefore omitted the CS identification task in a high-powered Experiment 3, and we additionally introduced a slightly delayed CS onset to create a more simultaneous experience of CS and US. In this third study, we found some evidence for an EC effect for briefly presented CSs, which appeared to be contingent on our—registered—exclusion criteria.

One important limitation is that CSs were not presented truly subliminally in any of the presented studies. CS identification was above chance for all the realized conditions, such that the observed EC effects might have been driven by some participants' conscious perception of some of the CS presentations. Awareness of the contingency between these CSs and the auditive USs may therefore have formed consciously and might underlie the EC effects found in the brief presentation condition. It therefore does not follow that the process underlying the EC effect with briefly presented CSs in Experiment 3 must have been an automatic one. The notion that only a few stimuli were consciously perceived by perhaps only a small subset of participants could also explain the small effect size found for CSs presented for 20 ms in Experiment 3, which could arise from a mixture of an average-sized EC effect contributed by participants who saw the briefly presented stimuli and a null EC effect of participants who did not see the stimuli. Nevertheless, the results of Experiment 3 give reason to further investigate the possibility of an EC effect with truly subliminal CS presentation.

The core proposition of this paper was that a simultaneous presentation of CS and US might be beneficial for EC effects with briefly presented CSs. Through a cross-modal EC procedure with visually presented CSs and auditive USs, we attempted to achieve a simultaneous experience of CS and US. Using this procedure an EC effect with briefly presented—but above chance visible—CSs was obtained, which is in contrast to the findings by Stahl et al. (2016). However, the present study does not address whether the procedural changes indeed had the discussed causal effects, due to the non-experimental manipulation of the (presence versus absence of the) CS identification task, and due to the lack of evidence regarding the effect (or its absence) of slightly delaying the CS onset on EC with briefly presented CSs.

Future research should more directly address the possibility that a simultaneous experience of CS and US might be beneficial for EC effects to occur (Jones et al., 2009), and that a CS identification task to assess visibility, administered after every trial, might interfere with EC for briefly presented CSs. Perhaps most importantly, the present work found a small EC effect with briefly presented CSs, suggesting that the search for a set of enabling conditions for subliminal EC effects might prove worthwhile, and that the cross-modal paradigm proposed here might be a useful method for future studies.

Chapter 7: No evaluative conditioning effects with briefly presented stimuli

Additional materials are stored in an online repository: osf.io/3dn7e

The evaluative conditioning (EC) effect is a change in the evaluation of an initially neutral conditioned stimulus (CS) after (repeated) pairing with a negative or positive unconditioned stimulus (US) in the direction of the valence of the US (De Houwer, 2007). While there is no doubt about the existence of the effect (Hofmann et al., 2010), there are currently two opposing views about the underlying processes explaining EC effects; namely propositional single-process views (Mitchell et al., 2009) and dual-process views (Gawronski & Bodenhausen, 2006, 2014b). One important difference between the two views concerns the role of CS-US contingency awareness and its necessity for EC effects to occur (for a review see Sweldens et al., 2014). Proponents of propositional single-process views claim that contingency awareness is a necessary precondition for EC effects to occur (Mitchell et al., 2009). Put differently, participants have to be aware of both the CS, the US, and their co-occurrence during the learning phase in order for a change in CS valence (in line with US valence) in a subsequent evaluation to occur. Support for this claim comes from studies demonstrating that EC effects were only found when participants were aware of the US valence with which CSs had been shown (Stahl et al., 2009).

Proponents of dual-process views (for example as described in the associative-propositional evaluation model; Gawronski & Bodenhausen, 2006, 2014b) claim that in addition to a propositional learning process, associative learning processes exist that require no contingency awareness for associations to form. A change in preference—in an associative framework—can be the result of a co-occurrence of a CS and US in the absence of the explicit knowledge of participants that these two stimuli appeared together. Dual-process theories are supported by findings of EC effects in experimental conditions under which contingency awareness is highly unlikely or even impossible: When one of the two stimuli in an EC learning phase is presented subliminally, it should be processed unconsciously, and cannot be consciously linked to the US (Dehaene et al., 2006). If EC effects can be found with a subliminal presentation schedule, contingency awareness between CS and US can be ruled out. Such effects could therefore not be explained by propositional single-process views (Mitchell et al., 2009).

A number of studies empirically support the notion of EC effects with subliminally presented stimuli (e.g., Dijksterhuis, 2004; Niedenthal, 1990; Rydell et al., 2006). Most of these studies, however, have been criticized on methodological grounds (i.e., manipulation of US valence between participants allowing for mood differences between groups as an explanation for evaluation differences; Sweldens et al., 2014) or could not be replicated by an independent lab (Heycke, Gehrmann, Haaf, & Stahl, 2018).

Additional empirical evidence for the absence of EC with briefly presented stimuli was recently introduced by Stahl et al. (2016). In a series of studies, no EC effects were found with stimuli that were identified slightly, but substantially, above chance. In most of the experiments, Stahl and colleagues measured the visibility of the briefly presented words using an online visibility check. Specifically, after every trial during the learning phase, participants had to select the CS (from a selection of CSs) that was presented during that trial. This online check of visibility enabled the researchers to obtain a more accurate estimation of the actual visibility of stimuli during the experiment (Lovibond & Shanks, 2002) as compared to asking for memory of briefly presented stimuli only at the end of the study (e.g., Rydell et al., 2006). One could speculate, however, that this online visibility check might interfere with the learning procedure in two ways: (I) participants might be in a deliberative and analytic mindset during the task as they are trying to identify the briefly presented stimuli, and therefore associations between the briefly presented CS and the US might not form if they depend on holistic processes, (II) assuming that a new evaluative response toward the CSs is acquired during the learning phase, showing multiple CSs that were paired with USs of different valence after every trial during the learning phase could be considered an additional trial of the learning phase, and could therefore interfere with the intended CS-US pairing schedule. A recent experiment by Heycke, Aust, and Stahl (2017) attempted to tackle this potential interference by omitting the online visibility check, while also addressing another potential limitation: the presentation schedule of the CS-US pairs.

One potentially important factor in EC with briefly presented stimuli—central to the studies reported by Heycke and colleagues—is the simultaneous presentation of CS and US. Recent findings support a claim by the implicit misattribution theory (Jones et al., 2009) that EC effects acquired in an associative automatic fashion require a simultaneous presentation of CS and US (Hütter & Sweldens, 2013; Sweldens et al., 2010). If this characteristic of the learning phase is indeed necessary for automatic EC effects to occur, it could help explain why recent replication studies did not find an EC effect in a sequential learning paradigm (Heycke et al., 2018). Studies showing the need for a simultaneous CS-US presentation for associative automatic EC effects, however, did not use a subliminal presentation schedule. In a subliminal presentation paradigm, participants' visual attention might be focused on either the briefly presented CS, or the US, but not both. It therefore seems difficult to realize a simultaneous visual presentation of CS and US when either one is presented subliminally.

Heycke and colleagues speculated that a cross-modal EC paradigm, with (subliminal) visual CSs and positive/negative sounds as USs, might allow for a simultaneous presentation (or, more importantly, simultaneous processing) of CS and US and therefore might provide a better test of subliminal EC effects acquired via implicit misattribution (Jones et al., 2009). In a high-powered study using this paradigm—that avoided the potential problem of the online visibility check—indications for an EC effect with briefly presented stimuli were indeed found by Heycke and colleagues (2017). It should be noted, however, that pilot studies, which made use of an online visibility check, showed that CSs were recognized at above-chance levels under the conditions realized in that study. Nevertheless, the results were in contrast to previous findings of no EC effect with briefly presented (but visible slightly above chance) CSs (Stahl et al., 2016).

Taken together, Heycke et al. (2017) found an EC effect for briefly presented—but slightly supraliminal—CSs, when using a cross-modal presentation paradigm and no online visibility check. However, the Bayes factor found by Heycke and colleagues in support of this finding ($BF_{10} = 6.88$) was not convincing if one might hold a strong prior believe against the existence of EC effects with briefly presented stimuli. The effect was also only present when a substantial amount of participants was excluded (based on pre-specified exclusion criteria). Nevertheless, the cross-modal approach appeared promising and an important question is therefore whether the EC effect found by Heycke et al. (2017) can be replicated, and whether it relies on above chance CS identification or extends to CSs detected at chance level.

Method

As the current study aims at replicating the findings by Heycke et al. (2017), we used the same experimental script, with some minor adjustments: (1) In order to ensure that the effect found by Heycke et al. (2017) can be generalized to different stimuli, we used a new set of CSs that were randomly assigned to a brief and a longer CS presentation time condition. (2) We additionally added an extensive visibility check, which was conducted after the measurement of the dependent variable. The study can therefore be considered a close replication of the study by Heycke et al. (2017), using different participants, different CSs, and adding an extensive visibility check. (3) We dropped the choice task (which appeared to have had an effect on the evaluation task and would therefore require counterbalancing) in order to focus the statistical power on the more promising dependent variable (in the original study, no EC effect on choice was found for briefly presented CSs).

Design. The study followed a 2 (US Valence: positive vs. negative) $\times$ 2 (CS duration: 30 ms vs. 1000 ms) within-subjects design.

Material. We used the same harmonic and disharmonious sounds (Topolinski & Deutsch, 2012, 2013) as USs that were used in the study by Heycke et al. (2017), but replaced the Pokemon images that we used as CSs in previous studies by a new set of less familiar stimuli.

The CS selection procedure consisted of two parts: first, images were rated in order to select neutrally valenced images, second, a visibility task was administered for this subset of images to select images that could not be identified above chance in a procedure similar to the learning paradigm of the main study.

We downloaded 100 images of toy figures (Gogo's Crazy Bones), transformed them to black and white and removed all photo artifacts. In an online study, all images were rated on a scale ranging from -100 to 100 by 119 participants (who were shown 50 semi-randomly selected images each, resulting in 44 to 52 evaluations per image). We selected the 16 most neutral images (i.e., the 16 images with mean ratings closest to the midpoint of the scale) based on this pilot study.

In a second pilot study, participants came to the lab and were presented with images in a procedure highly similar to the main study (see below): This second pilot study–in which each of the 16 images was shown 10 times for 30 ms in a cross-modal learning procedure—served as a test for the visibility of stimuli. The images were presented along positive/negative sounds (randomly determined for each participant anew) to keep the presentation conditions as close as possible to the main study. After every trial, participants had to indicate which stimulus was shown in the preceding trial. Two additional images (that were not part of the selected 16 neutral images) were presented for 80 ms or 200 ms (10 times each) to raise the motivation to detect brief and masked stimuli (Pratte & Rouder, 2009), which we will refer to as motivational stimuli. The intertrial interval, fixation cross, and timing of image-sound pairs were identical to the main study. In contrast to the main study, after every trial all 18 stimuli (16 CSs plus the two filler stimuli) were shown on the screen and participants were asked to select the image that had appeared during the trial.

In a first part of this second pilot study with 30 participants we identified four images that were identified at chance level (with $BF_{01} > 5$). We reduced the contrast of all other images in order to reduce the visibility of the stimuli and ran another pilot study with the four images identified in the first part and all other images with a reduced contrast (16 CSs plus two filler stimuli). In total 16 new participants took part in the second part, and results implicated eight stimuli that were not identified above chance (all $BF_{01} > 5$). Those eight stimuli served as CSs for the main study.

Procedure. The learning phase followed the procedure of the experiment by Heycke et al. (2017). Participants first received a surveillance task instruction (i.e., press the space bar as soon as a specific target character appears). One (of two) stimuli (which were the same motivational stimuli as in the pilot studies) was randomly selected to be the target stimulus for each participant anew. Participants then completed a 10-trial practice phase for the surveillance task. In this practice phase, the actual target (shown 4 times) plus three stimuli (shown twice each) were shown (targets for 200 ms and the other images for 30 ms or 1000 ms). The three non-targets were only used for this practice phase, while the target was also the target in the following learning phase. During the practice phase, only neutral sounds were played (see below). Additionally—and contrary to the subsequent learning phase—feedback was given when the space bar was pressed but no target animal was shown and feedback was given after trials in which a target was shown but the space bar was not pressed. We included these practice trials as we had observed in previous studies that the surveillance task was not performed by all participants. As we used the performance on this task as an outlier criterion (see below), we wanted to make sure that participants were aware that their performance on the task was measured. After the practice phase, the same target stimulus was shown again, and participants were instructed that they would not receive feedback anymore on their performance in the following detection task.

Each trial in the subsequent learning phase was set up as follows: A fixation cross with a duration of 500 ms in the center of the screen announced the trial and was followed by a random-checkerboard pattern mask presented for 500 ms. The mask was randomly rotated by either 0, 90, 180 or 270 degrees. Afterward, either a CS was shown for 30 ms or 1000 ms; or a surveillance target was presented (for 200 ms); or the (non-target) motivational stimulus (see above) was shown for 80 ms. The US onset was always 400 ms before the CS onset, in order to approximate a simultaneous experience of CS and US valence. The trial was followed by a 100 ms backward mask (also randomly rotated, but with a different angle than the forward mask) and a 1500 ms intertrial interval with a blank screen.

The learning phase consisted of 4 CSs presented for 30 ms and 4 CSs shown for 1000 ms, which were randomly assigned to the presentation time for each participant anew. Half of the CSs were paired with positive sounds, the other half with negative sounds, which was also randomly determined. Additionally, as described above, a target image was randomly selected for each participant anew out of the two motivational stimuli. Both motivational stimuli were always paired with neutral ticking sounds taken from Heycke et al. (2017). During the learning phase, each image-sound pair was shown 10 times in a random order, which resulted in 100 trials in total.

After the learning phase, evaluative ratings of the CSs were administered. Participants were instructed to give their spontaneous evaluation of the previously shown figures. Each CS was presented in the center of the screen (in a random order), and evaluation on a slider ranging from -100 to 100 was administered.

In the subsequent visibility test, all CS-US pairs (including the target and motivational stimulus) were shown again in the same order as in the learning phase. However, no target instructions were given; instead, after each trial, participants were presented with all 10 CS stimuli (8 CSs, 1 motivational stimulus, 1 surveillance target stimulus). Participants were instructed to click on the stimulus that had just been shown in that trial. At the end of the study, participants answered demographic questions (age, study, gender, was the headphone worn, goal of experiment, and comments).[17]

[17]The data of the pretest and the main study were collected under a Born Open Data protocol (Rouder, 2016) in which they were automatically logged, uploaded, and made freely available after every day of data collection (github.com/methexp/rawdata/tree/master/croco4, github.com/methexp/rawdata/tree/master/croco4b, github.com/methexp/rawdata/tree/master/croco5).

Data analysis. We used the same outlier and exclusion criteria as in the previous study (Heycke et al., 2017). Before conducting any data analyses, we excluded participants who (a) reported that they did not wear the headphones during the experiment, (b) did not press the space bar often enough when a target image was shown (using Tukey's outlier criterion but only if the criterion reached 3 misses out of 10 trials, allowing for 1 or 2 misses), (c) aborted the experiment or (d) explicitly reported major (unexpected) problems during the experiment (e.g., distractions or difficulties with the instructions).

For the data collection, we used a sequential Bayesian analysis paradigm (Rouder, 2014; Schönbrodt et al., 2015), analyzing the data after every day of data collection and stopping after a pre-defined criterion was reached (see below). We use BF_{10} to denominate the evidence for the alternative hypothesis relative to the null hypothesis (i.e., $BF_{10} > 1$ indicates support for the alternative hypothesis over the null hypothesis) while BF_{01} denominates the evidence for the null hypothesis relative to the alternative hypothesis (i.e., $BF_{01} > 1$ indicates support for the null hypothesis). All data analyses were performed in R (Version 3.4.4; R Core Team, 2016)[18].

Repeated measures ANOVAs were performed to analyze the interaction between US valence and presentation time in the evaluative ratings and paired t tests to examine interaction effects. In the Bayes Factor ANOVA analyses, the following method of estimation whether a given interaction or factor should be considered to have an influence on the evaluative rating was used: A full model with all main effects and interactions was compared to a model without the main effect or interaction (i.e., the predictor) of interest. A Bayes factor for the alternative hypothesis over the null hypothesis (i.e., $BF_{10} > 1$) could be interpreted as relative evidence for one model (i.e., the full model including our predictor term of interest) over the competing model (without the predictor interest). As in the previous studies by Heycke et al. (2017), we used default multivariate Cauchy priors (Rouder et al., 2012) with a scaling parameter of $r = 0.5$ in the Bayes factor ANOVA and Cauchy priors with a scaling parameter of $r = \sqrt{2}/2$ for all t tests (Rouder et al., 2009). All errors for Bayes factor estimates were less than 1%. For all t tests we also report the median of the posterior distribution as an effect size estimate and its 95% highest density interval

[18]We, furthermore, used the R-packages *afex* (Version 0.20.2; Singmann et al., 2016), *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2015), *papaja* (Version 0.1.0.9735; Aust & Barth, 2016), *RCurl* (Version 1.95.4.10; D. T. Lang & CRAN team, 2016), *rvest* (Version 0.3.2; Wickham, 2016), *tidyr* (Version 0.8.0; Wickham & Henry, 2017), and *yarrr* (Version 0.1.5; Phillips, 2017).

(HDI; with a probability of 95%, the true population effect size is within the 95% HDI). Additionally, for all analyses, the total number of individual participants included in the analysis is reported in brackets.

Sample size rationale.

As part of the sequential Bayesian analyses paradigm, we started the data analysis after initially collecting the data of 30 participants, with the same stopping rules used by Heycke and colleagues (2017) in Experiment 3. Specifically, we decided a priori to collect data until either the Bayes factors for both t tests of interest (i.e., for an EC effect in the 1000 ms condition and an EC effect in the 30 ms condition) were larger than 10 (for the null or alternative hypothesis) or until a maximum of 150 (paid) participants was reached. We ran the analyses after every day of data collection and decided that signed-up participants were allowed to participate, even if we had already stopped the data collection based on the above-specified analyses.

Participants.

We stopped data collection after 166 participants took part in the study ($N_{paid} =$ 94, all others received partial course credit), based on the results of the t tests (see below). Six participants were excluded, as they either took part in previous studies of this series or because they were told about the procedure of this study by others before taking part. One participant reported that she did not wear the headphone during the entire procedure and seven additional participants did not detect the surveillance target stimulus at least 7 times during the learning phase. These participants were also excluded before running any of the analyses. The total sample size in the following analyses was therefore N = 152 ($M_{age} = 23.51$, $SD_{age} = 6.93$; 105 female).

Results

We first report the same analyses as in the previous experiment by Heycke et al. (2017) in order to test whether an EC effect could be found in the 30 and 1000 ms condition. Afterwards, we report additional analyses using the data from the visibility check.

Evaluation. In the overall ANOVA, we found an interaction of the presentation time of the CSs and the US valence, $BF_{10}(152) = 83.25$; $\hat{\eta}_G^2 = .018$, see Figure 19). The interaction was characterized by the fact that, when stimuli were presented for

1000 ms, an EC effect was present, $BF_{10}(152) = 3,355.25$, $d = 0.37$, 95% HDI [0.21, 0.53]. We therefore replicated previous findings, showing that images shown while a positive (harmonic) sound was played were evaluated more positively than images that were shown while a negative (disharmonic) sound was played. Contrary to our previous finding, we found statistical evidence for the absence of an EC effect when stimuli were presented for 30 ms, $BF_{01}(152) = 10.47$, $d = 0.01$, 95% HDI [-0.15, 0.16].

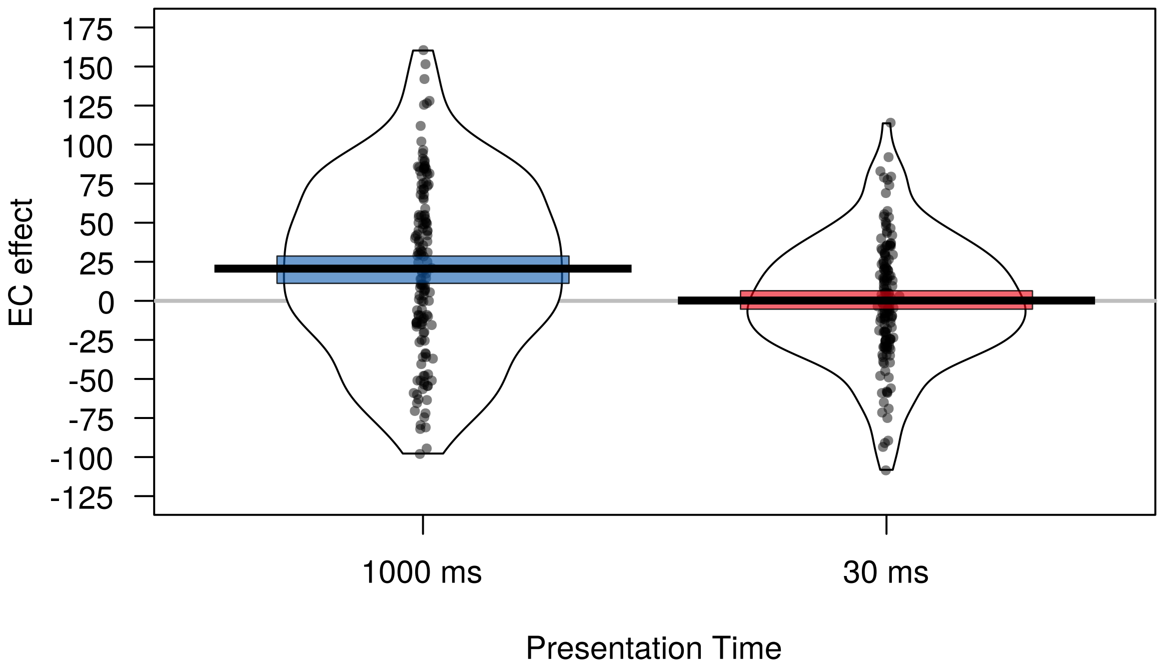

Figure 19. EC effects, split by CS presentation time. Large horizontal bars represent the mean EC effect, with the 95% highest density intervals (in color) displayed around them. Dots represent individual EC effects and additionally the smoothed distribution for each presentation time condition is shown.

Visibility. We tested whether stimuli were identified above chance in the visibility testing procedure which followed the learning and evaluation phase. Stimuli that were presented for 30 ms were indeed identified above chance, $BF_{10}(152) = 1,504.92$, $d = 0.37$, 95% HDI [0.20, 0.53] (chance level = 0.1, average visibility = 0.13, $SD = 0.09$).

Exploratory analysis. To investigate whether an EC effect might exist for stimuli that were not consciously perceived, we excluded individual ratings of stimuli that were identified three or more times in the visibility check (i.e., we included only stimuli identified below, at, or slightly above chance).[19] Similar to the result when using all evaluations, we found evidence for the absence of an EC effect for stimuli presented for 30 ms, $BF_{01}(143) = 17.87$, $d = -0.06$, 95% HDI [-0.23, 0.10].

Second, we also analyzed only evaluations of stimuli presented for 30 ms in the learning phase that had been selected at least 3 times during the subsequent visibility check. There were only 18 participants that had correctly identified at least one positive and one negatively paired CS more than three times, allowing for their inclusion in the paired *t* test analysis. Nevertheless, we found strong indications for an EC effect for briefly presented CSs that had been identified at least three times in the visibility test, $BF_{10}(18) = 31.54$, $d = 0.73$, 95% HDI [0.21, 1.27].[20]

Discussion

The main goal of this experiment was to replicate our previous finding of an EC effect for briefly presented stimuli, using a cross-modal design. Replicating previous work, we found compelling evidence for EC effects with stimuli shown for 1000 ms, demonstrating that the general finding of cross-modal EC is robust and generalizes to a new set of CSs. However, and contrary to our previous study, we did not find any evidence for an EC effect for briefly presented stimuli, but strong evidence for the *absence* of EC effects for these stimuli. If indeed spontaneous associations between CS and US underlie EC effects in the absence of contingency awareness, one could speculate that perceiving CSs consciously might prevent EC effects. We therefore analyzed only the evaluation of stimuli that had been identified at or around (but not above) chance level, but the conclusion remains unaltered: We again found statistical evidence for the absence of an EC effect.

In contrast, we observed an EC effect in the subset of stimuli that were briefly presented but recognized at least 3 times in the subsequent visibility test. Considering this substantial EC effect for briefly presented stimuli ($d = 0.74$) in this subgroup, one could speculate that the EC effect found in our previous study was driven by a similar subgroup of individuals, who saw some of the stimuli. This speculation

[19] As we had two stimuli per experimental cell (e.g., two different CSs for briefly presented stimuli paired with a positive US), we had at least one rating per US valence in the 30 ms condition for 143 participants when applying the exclusion criterion.

[20] We additionally tested whether these 18 participants, who showed a better visibility for the 30 ms stimuli, also showed a larger EC effect for stimuli presented for 1000 ms than all other participants. There were no indications that this was the case, $BF_{01}(152) = 1.65$, $d = 0.30$, 95% HDI [-0.15, 0.77].

should be taken into consideration when looking at previous studies that claim to have found subliminal EC but did not check for actual visibility in a stringent manner (as criticized by Sweldens and colleagues, 2014). The results of the present study therefore demonstrate that (I) a cross-modal EC effect with stimuli presented for a longer duration appears to be robust; (II) when stimuli are presented briefly, but can still be identified, a cross-modal EC effect can also be obtained (cf. Stahl et al., 2016); and (III) when stimuli are presented briefly but cannot be identified, EC effects are absent even in the cross-modal paradigm that allows for simultaneous presentation.

Limitations. The visibility check of the main study showed that stimuli presented for 30 ms were identified at above-chance levels. This finding is surprising, as our pre-tests indicated that the same stimuli were identified at chance level. There was, however, an important difference between the visibility checks in our main study and the pre-tests: In the pre-test, the majority of trials included briefly presented stimuli, whereas more than half of the stimuli were presented for a longer duration in the main study. The fact that most stimuli were difficult to see in the pre-tests could have had an influence on the motivation to detect the stimuli, leading to a lower visibility rate (Pratte & Rouder, 2009). Additionally, the same characteristic could have had an influence on participants' response strategy: As half of the CSs were presented for 1000 ms (and two additional stimuli for 80 ms or 200 ms), those stimuli could easily be identified during the task. When a stimulus was shown briefly and participants were asked to identify the stimulus from the set of ten stimuli, participants might have discarded the six stimuli that they knew were shown for longer durations. This would lead to a guessing rate of 1/4 instead of the rate of the 1/10 we assumed. Note that the visibility of the subset of stimuli that showed EC (i.e., those correctly identified on 3 or more trials) would still be above the thus-corrected chance level. Future research should assess the plausibility of these speculations.

If we take these results at face value, we replicated the finding by Stahl et al. (2016) that above chance CS identification might be insufficient for EC effects with briefly presented stimuli. However, given our speculations about the visibility check above, it might be possible that stimuli were indeed presented subliminally and we overestimated the actual visibility. In any case, the main conclusion remains unaltered: EC effects were not found for briefly and masked CSs, while EC effects were found when the CS was presented for a longer duration (i.e., 1000 ms).

Implications. The main theoretical goal of this study was a test of the necessity of contingency awareness for EC, as predicted by propositional single-process views. The results of this study clearly support this view as contingency awareness was found necessary for EC effects to form. It should be taken into account that contingency awareness could be manipulated by means other than brief presentation; future studies should look at different options of manipulation contingency awareness directly. However, recent studies using parafoveal presentation (Dedonder et al., 2014) or continuous flash suppression (Högden et al., 2018) to manipulate contingency awareness experimentally, also failed to find any evidence for EC effects in the absence of contingency awareness. In addition, it would be interesting to test whether EC effects with briefly presented and masked CSs follow other regularities than EC effects with clearly visible CSs, as recently suggested by Greenwald and De Houwer (2017). Furthermore, single-and dual-process views have different predictions that go beyond contingency awareness (e.g., relational information about CS and US) that should be considered before evaluating the merits of both views in explaining the mechanisms underlying EC (for a review see Corneille & Stahl, 2018).

Taken together, the results of this study challenge the finding by Heycke et al. (2017) that a cross-modal EC paradigm might be beneficial for EC effects with briefly presented stimuli and support the claim made by Stahl and colleagues (2016) that above-chance CS identification may be necessary but insufficient for attitude learning.

Chapter 8: General Discussion

We set out to test whether contingency awareness is necessary for EC effects to form and therefore experimentally manipulated contingency awareness using subliminal stimulus presentation paradigms. A series of previous studies had reported such EC effects with subliminally presented stimuli. However, as discussed in the introduction, a *p*-curve analysis of these studies revealed no particularly convincing results. The three studies of Rydell and colleagues appeared to be of particular importance for the evidential value of the *p*-curve, which could indicate that the results of Rydell and colleagues might be based on a robust paradigm. However, in two close replications—using the original stimulus material—we did not replicate the findings reported by Rydell et al. (2006).

We then tested an additional set of possibly beneficial factors in subliminal EC. First, we tested whether the relevance between US and CS as well as the goal relevance for the participants would moderate an EC effect. We did not find any evidence that these factors might be beneficial for automatic EC effects and did not observe any EC effects with truly subliminally presented stimuli. Second, we attempted to measure preference with choice measures, as it had been proposed that choice might be a more sensitive measure for automatic EC effects than explicit evaluations. However, we did not observe any EC effects using a choice measure when CSs were presented subliminally. In contrast, we did observe an EC effect with briefly presented—but visible—CSs with an evaluative but not with a choice measure in Experiment 3 in Chapter 6. Third, we speculated whether a simultaneous presentation of CS and US using the visual and auditory modality might be beneficial for subliminal EC effects, as previous studies had shown that automatic EC effects might rely on a simultaneous presentation schedule. While we found an EC effect with briefly presented CSs with a simultaneous cross-modal presentation schedule, an additional experiment suggested that the observed EC effect was only found when participants consciously perceived the stimuli. We set out to test the question whether EC effects require contingency awareness. The results of the series of empirical investigations suggest that, indeed, contingency awareness is necessary for EC effects to form.

One important additional aspect of the studies presented in the previous chapters was that we wanted to ensure that stimuli were not perceived consciously during the learning phase (i.e., were presented truly subliminally). While a number of previous experiments on subliminal EC did not test whether the briefly presented stimuli were visible (see Chapter 1), we used different methods to assess visibility: We tested the visibility during the learning phase (Chapter 4 and Chapter 6) as well as after the learning phase (Chapter 7) by asking participants what they saw on a trial-by-trial basis. Two important findings can be deduced from the results of the empirical studies presented in the previous chapters: First, EC effects can form even when stimuli are presented very briefly (e.g., for 17 ms) as long as the stimuli can be perceived consciously. It therefore appears necessary that each study investigating the influence of subliminally presented stimuli needs to deploy a visibility check. Only then one can draw the conclusion that the behavioral change was the result of subliminal stimulus processing and not the result of a conscious perception of the CS-US pair. Merely presenting stimuli for a brief period of time is not sufficient to claim that the stimuli were presented subliminally. Second, measuring visibility is not as straightforward as it might appear. As we saw in Chapter 7, the correct identification of the same stimuli (with the same presentation time) might depend on the number of stimuli to select from or on the number of stimuli presented briefly. Guessing strategies and personal differences (e.g., conservative vs. liberal decision criteria; see Lovibond and Shanks, 2002) should be taken into account when estimating stimulus visibility in studies using subliminal presentation schedules.

To summarize the findings from the previous chapters, we found no evidence for EC when stimuli were presented truly subliminally and therefore did not find any evidence for EC without contingency awareness.

Meta-Analysis

To investigate the influence of the presented studies (and the results by Stahl and colleagues, 2016) on an overall estimation of EC effects with briefly presented stimuli, the meta-analysis was calculated again, now including these studies. A random effects analysis revealed a mean effect size of $d = 0.32$ (Z = 0.16) with a 95 % confidence interval from $d = 0.22$ to $d = 0.42$, see Figure 20. The analysis yielded that the study results were not homogeneous $Q = 88.66$, $df = 49$, $p < .001$, which was mainly due to the reversed effects reported in Chapter 2 (and the reversed effects found by Versluis et al., 2017).[21]

[21] Removing these results from the analysis resulted in a homogeneous set ($Q = 58.13$, $df = 45$, $p = .091$) and a mean effect of $d = 0.38$ (Z = 0.19).

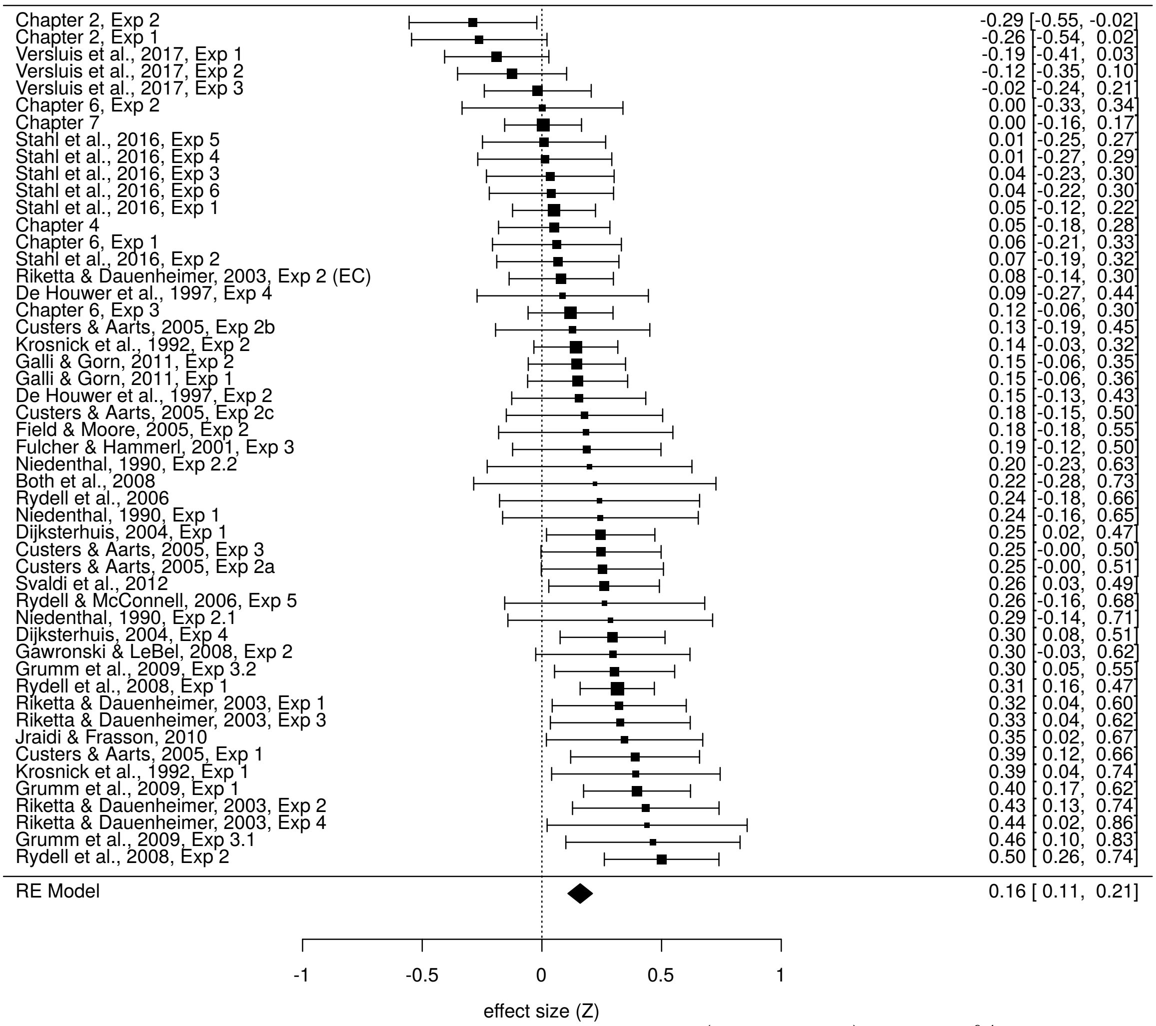

Figure 20. A forest plot showing the effect sizes (Fisher's Z) and 95% confidence intervals for each study included in the literature review and the studies discussed in the previous chapters, based on a random effects model.

Our studies show that it is difficult to produce EC effects with briefly presented stimuli and suggest that EC effects are absent when stimuli are presented truly subliminally. The meta-analysis also demonstrates a substantial decrease of the estimated average effect size when including the set of studies conducted in our lab (from $d = 0.45$ to $d = 0.32$). While we can only speculate, it appears possible that the small effect of $d = 0.32$ might be a mixture of studies (or participants) with truly subliminal stimuli who report no effect and studies (or participants) with slightly visible stimuli reporting a mean effect observed with supraliminally presented stimuli (e.g., $d = 0.52$ as observed in the meta-analysis by Hofmann et al., 2010).

Theoretical implications

The main goal of this set of studies was to investigate the need for contingency awareness in evaluative learning. In a series of studies we therefore experimentally manipulated contingency awareness by means of subliminal stimulus presentation. As discussed in Chapter 1, the propositional single-process view (Mitchell et al., 2009) postulates that EC effects depend on contingency awareness of the CS-US pair during the encoding of the stimuli. Dual-process accounts such as the APE model (Gawronski & Bodenhausen, 2006), in contrast, postulate that EC effects can occur automatically and therefore without contingency awareness. Studies demonstrating subliminal EC effects have been used by theorists to show support for the APE model (Gawronski & Bodenhausen, 2011). However, as discussed in Chapter 1, the empirical evidence for subliminal EC effects was not convincing. Therefore, finding an EC effect without contingency awareness in a methodological solid experiment would lend support to dual-process models and could not be explained by propositional single-process views.

Our results clearly indicate that EC effects are not possible when stimuli are presented truly subliminally, which is in line with the prediction of a propositional single-process model and not in line with the prediction of dual-process models. These results generally match the results by Stahl and colleagues (2016) who also found no EC effects with briefly presented stimuli, even if those stimuli were identified slightly above chance. In contrast to the studies by Stahl et al. (2016), we found some indications that EC effects can occur with a brief presentation schedule, but only when stimuli are visible above chance. Both our investigations and the studies by Stahl et al. (2016) support the claim that evaluative conditioning effects are not possible when stimuli are presented truly subliminally. Of course, presenting stimuli only for a brief moment is just one way to experimentally manipulate contingency awareness and it is therefore difficult to generalize the absence of an EC effect in this paradigm to all possible experimental manipulations of contingency awareness. However, recent studies by Högden and colleagues (2018) using continuous flash suppression to prevent contingency awareness between CS and US did not obtain any evidence for EC effects in the absence of awareness. Additionally, the findings by Dedonder et al. (2010) are in line with these results, demonstrating the absence of an EC effect when stimuli were presented parafoveally and were therefore not consciously perceived. Given these consistent null findings by a number of different researchers, the possibility should be taken into consideration that EC might depend on contingency awareness. If contingency awareness is indeed a necessary precondition of evaluative conditioning, the necessity for an additional associative process would be reduced. Specifically, the possibility of the acquisition of an EC effect without contingency awareness is one of the key assumptions of the associative process that cannot be explained by propositional single-process models. If future research would confirm the findings presented in the previous chapters and the mentioned findings by Högden et al. (2018) and Dedonder et al. (2014), properties of dual-process models should be revisited and the possibility should be entertained to combine the dual- and single-process model into one comprehensive (single-process) model.

If EC effects without contingency awareness, as assumed by the associative account, are not possible, it might be worth investigating other diverging predictions of single- and dual-process models. Future research could, for example, address other proposed differences in associative vs. propositional learning, such as the absence of extinction and the influence of relational information between US and CS on EC. Specifically, investigating the influence of information qualifying the relation between CS and US might give insights into working mechanisms underlying EC. Imagine for example that a pharmaceutical product serves as the CS and an image of a skin disease as the US. When telling participants that the CS causes the negative US it will be evaluated as less positive than when the CS is preventing the negative US (Hu et al., 2017). These findings are in line with a propositional process. However, if spontaneous associations can underlie EC one might find that the (negative) US and the CS are associated, ignoring the relational qualifier. The empirical findings of studies investigating these effects are mixed (Hu et al., 2017; Moran & Bar-Anan, 2013) and additionally, findings of unqualified associations might also be explained by (partial) forgetting in a single-process view (De Houwer, 2014a). Future research should investigate whether effects can be found that can only be explained by associative processes in which spontaneous associations arise.

To sum up, given the results of the investigations presented in the previous chapters and given the results of the *p*-curve analysis of studies claiming to have found subliminal EC, the claim by Jones and colleagues (2010) that subliminal EC should be taken as clear evidence for EC effects in the absence of contingency awareness is challenged. Theorists should be cautious building models on the assumption that EC effects can form without contingency awareness, especially when building upon studies using subliminal presentation procedures.

Practical implications

As summarized above, influencing evaluations by means of subliminal EC appears to be very difficult and we did not manage to influence participants' behavior when they were not aware of the stimuli. It is therefore safe to say that bans of subliminal advertisement—while not harmful—might not be necessary (Alcohol and Tobacco Tax and Trade Bureau, 2014). We did, however, find some indications for beneficial circumstances for EC effects with briefly presented—but visible—CSs (Chapter 4 and 6). Yet, it appears that the increase in the possibility to find an EC effect with briefly presented CSs was related to an increase in CS-US contingency awareness. While these findings are not directly relevant when evaluating the predictions of single- and dual-process models (as both models would predict an EC effect when contingency awareness is possible), the results might be interesting for marketing research. Acquiring more knowledge on how changes in preference can be induced with briefly presented stimuli might give insights into boundary conditions of EC as a possible application in marketing. To change participants' preferences using an EC learning paradigm with briefly presented stimuli, factors that raise contingency awareness need to be present. These factors might include the (1) simultaneous presentation of CS and US, (2) absence of distracting tasks such as a visibility test (see also Pleyers et al., 2009) and (3) CS-US relevance. These possibly beneficial factors should receive attention in research in marketing and consumer science to confirm their profitable role.

Outlook

A series of factors that might benefit automatic EC effects were discussed and empirically tested in the previous chapters. In the next paragraph, I will address each factor and provide an outlook on the future of research on subliminal EC in general.

One factor that was discussed is the possibility that a choice measure might be more sensitive than an explicit evaluation. While we did not find empirical support for this notion in the present studies, the ecological validity should be an important factor in future studies investigating EC in applied settings. The question whether a change in preference (that is the result of an EC learning phase) translates to behavioral changes is an important question that needs to be addressed by future studies (see Kruglanski et al., 2015).

A second factor which was hypothesized to play an important role in subliminal EC is the simultaneous presentation of CS and US. We consistently observed EC effects in a cross-modal paradigm with IADS sounds (e.g., "a woman crying" or "a drink being poured in") as well as with simple harmonic and disharmonic sounds, when CSs were presented for 1000 ms. If the simultaneous CS-US presentation is indeed beneficial for (automatic) EC effects the cross-modal paradigm should be used in further paradigms investigating "automatic" EC effects, such as the process-dissociation procedure (Hütter & Sweldens, 2013) or the surveillance paradigm (Olson & Fazio, 2001). The cross-modal approach might be particularly interesting to use in a surveillance paradigm, in order to disguise actual CS-US pairs. The findings of a reliable EC effect with CSs presented for 1000 ms in a cross-modal paradigm have interesting implications for the application of EC in marketing research: The effect of valent sounds in advertising has been discussed for decades (Gorn, 1982) and the results of our investigations show that even a repeated brief CS-US presentation of about one second is sufficient for changes in preferences. Summarizing, the cross-modal approach appears promising and might prove to be of use in different EC learning paradigms such as the surveillance paradigm (Olson & Fazio, 2001).

In Chapter 1 characteristics of studies showing subliminal EC were discussed, see Table 1. Two factors that were almost equally distributed over the discussed studies concerned the question of the stimulus presentation schedule (i.e., was the CS or US briefly presented) and the dependent variable used to measure the effect (i.e., implicit or explicit measure). We therefore tested for subliminal EC in a series of combinations of these two factors as there appeared to be no clear setting beneficial for EC. In the seven experiments discussed in the chapters above, we have used briefly presented USs combined with an implicit measure and briefly presented CSs combined with explicit measures (and choice as a possible implicit measure). We therefore covered a series of possible combinations but were not able to find any evidence for automatic EC effects. The only combination that has not been investigated in the presented studies are investigations with a briefly presented US and explicit measures. While this combination might be worth investigating, it bears the problems that are attached with subliminal US presentation (e.g., valence might not be encoded from briefly presented USs; Lähteenmäki et al., 2015).

Going beyond the investigated aspects, three additional factors could be considered in the investigation of subliminal evaluative conditioning in the future: individual differences, focus on feelings, and subjective awareness. In the next paragraph, these three factors are briefly discussed.

First, a factor that has not received much attention in the investigation of subliminal EC concerns the question of individual differences between participants. To gain a deeper insight into the working mechanisms of EC and to raise the quality of the observed data, future studies should focus more on the individual participant. One observation we made throughout the series of studies was that the visibility rates of briefly presented stimuli differed greatly between participants. While some participants appeared to see none of the briefly presented stimuli, other participants were aware of a number of stimuli. Future research could attempt to use individual presentation times for briefly presented stimuli so that all participants perceive the stimuli subliminally. Using monitors with a high refresh rate (which can produce small differences in the presentation times), stimuli presentation durations could be adapted individually before the learning phase. Additionally, other individual factors might be taken into account to test for the possibility of automatic EC effects. The previously discussed result by Fulcher and Hammerl (2001), which showed that only participants with low reactance scores showed a subliminal EC effect, might serve as a good starting point. Thus, future studies could investigate individual presentation times and additionally investigate possible mediating roles of personality traits (see Vogel, Hütter, & Gebauer, 2017).

A second factor discussed in the introduction, which was not directly addressed in the conducted empirical investigations, is the "focus on feeling" instruction used by Gawronski and LeBel (2008). While we did not explicitly instruct participants to introspect their feelings, we used the results from the studies of Gawronski and LeBel (2008) in that we instructed participants to "react spontaneously" and added that "there are no right or wrong answers" during the evaluation (e.g., in the Experiment in Chapter 7). Nevertheless, asking participants to introspect on their feelings before evaluating a stimulus after an EC learning phase might help discover automatic EC effects, when using explicit measures of preference.

A third important factor relates to the question how subjective awareness of briefly presented stimuli might relate to the objective visibility criterion we used in the discussed studies. Were participants actually consciously aware of the CSs they identified correctly or did they merely guess correctly due to subliminal activation of the stimulus? Perhaps more interesting is the question how subjective awareness relates to the presence or absence of an EC effect: Is the EC effect restricted to truly consciously visible CSs, or does the EC effect extend to CSs for which correct identification is based on subjective feelings or guessing? In future research, certainty of identification of the CS could be used as a proxy for this subjective awareness.

Findings related to subliminal EC. Other findings using subliminal presentation paradigms and learning phases similar to EC might be informative for future directions in subliminal EC. In the next paragraph, a series of findings are discussed that did not investigate EC but showed an influence of subliminally presented stimuli on behavior. The methods used in these investigations might be informative for future studies in evaluative conditioning.

One method that could be informative is the (subliminal) approach-avoidance task (AAT; K. Kawakami, Phills, Steele, & Dovidio, 2007), in which a target (e.g., a picture of a black or white person) was presented subliminally and followed by the instruction to make an approach or avoid movement. When images of black people were paired with approach movements and images of white people with avoidance movements, implicit and explicit evaluations of black people were more positive compared to white people (K. Kawakami et al., 2007). Additional research found empirical support for subliminal AAT effects (Jones, Vilensky, Vasey, & Fazio, 2013) and Jones et al. (2010) suggest that results showing subliminal AAT results should be considered as evidence for preference learning in the absence of contingency awareness. However, a recent investigation did not show any AAT effects with subliminally presented stimuli in three experiments (with one constituting a direct replication of the experiment by K. Kawakami et al., 2007; Van Dessel, De Houwer, Roets, & Gast, 2016). It is therefore not clear whether findings of subliminal AAT can serve as an insight into possible mechanisms in EC.

Another promising study, using subliminal stimuli, was recently published by Greenwald and De Houwer (2017). In a series of experiments, unpronounceable letter strings (CSs) were shown briefly and followed by a clearly visible word with either a positive or negative valence (USs, in other experimental conditions male or female first names were used instead). Participants had to press one of two keys quickly to classify the USs as negative or positive during the learning phase (or as male/female). In an evaluative priming task, the trial sequence and the classification task was the same with the exception that CSs were randomly paired with positive or negative words. This resulted in congruent (e.g., a CS previously paired with positive USs shown before a positive target word) and incongruent (e.g., a CS previously paired with negative USs shown before a positive target word) pairs. Analyzing error rates in this task, Greenwald and De Houwer (2017) found that associations between CSs and the positive/negative valence categories appeared to exist as participants made more errors in incongruent compared to congruent trials. Using the data of an extensive visibility check, it was found that the effect was present when participants were not aware of the CSs. While the effect is promising, it should be noted that it was found with positive/negative USs as well as female/male first names as USs. As no evaluative measure—but a priming task—was used as dependent variable, future research needs to confirm that the effect extends to evaluative ratings. Until then the effect can also be explained by (1) a learned motor response, (2) a facilitated identification of the target stimulus, or (3) a facilitated identification of the semantic category of the target (see Corneille & Stahl, 2018). While the procedure does not allow for any conclusions about the possibility of subliminal EC, the method appears promising and should be adapted for future studies in subliminal EC.

A third line of research that might be informative for future directions in subliminal EC are investigations diverting participants' attention away from the briefly presented stimuli (i.e., preconscious conditioning). In the studies discussed throughout the last chapters, we always used a weak stimulus and the attention of participants was directed towards the stimulus. Another possibility is a stimulus presentation of a longer duration while ensuring that attention of participants is not directed towards the stimuli. This technique was for example used in the studies by Högden et al. (2018) where no EC effect was found with continuous flash suppression (see Chapter 1). A recent study investigated inattentional blindness and showed that the decision whether a number was larger or smaller than 5 was influenced by other numbers that were not consciously perceived (Schnuerch, Kreitz, Gibbons, & Memmert, 2016). In the paradigm, participants focused on a cross in the center of the screen which was surrounded by 118 black characters ("W", "X", "Y"). Eight hash symbols surrounded the square of symbols in a virtual circle. After a brief waiting phase, one of the eight hash symbol turned red and each black character was randomly reset, sampling from the same set of characters ("W", "X", "Y"). After 300 ms the red hash symbol was replaced by a number between 1 and 9 and the task of the participant was to judge whether the number was larger or smaller than 5. Importantly, when the red hash symbol was replaced by a number the characters were reset again, this time however in one third of the trials 38 characters were replaced by numbers incongruent to the target number (i.e., if the number replacing the hash symbol was larger 5 the numbers were smaller and vice versa). In two experiments, reaction times towards the target numbers were increased when incongruent numbers were presented (compared to congruent numbers). As the studies by Schnuerch et al. (2016) were pre-registered and a visibility test was conducted (after the test phase), this paradigm appears to be promising to be adapted for an investigation in subliminal EC. Future studies in subliminal EC might be well advised to adapt this paradigm for use in EC research.

Limitations of subliminal EC research. One problem inherent in investigations of subliminal EC is that "subliminal presentation" is not a clearly defined concept (De Houwer et al., 1997). When investigating the conscious perception of stimuli they must be presented briefly enough so they cannot be consciously reported. However, as soon as a stimulus is identified at (or below) chance we do not know whether the stimulus is (practically) invisible or whether a sub-conscious processing of the stimulus is possible. Put differently, using a visibility check we can only differentiate between conscious and not-conscious perception of stimuli, but we cannot differentiate between sub-conscious processing and no processing nor between sub-conscious and conscious processing. Specifically, feelings of familiarity arising from subliminal processing of stimuli might influence the estimation of visibility of the stimuli, which would mean that an above chance identification might arise from subliminal processing of a stimulus. For the future of subliminal EC it appears necessary that subliminality is defined on two levels. First, agreement needs to be reached what (neuronal) processes are active in subliminal processing and how these processes translate into behavior. Second, agreement on manipulation and measurement of a subliminal stimulus presentation needs to be found.

A second problem with subliminal EC is that the learning phases are artificial and effects might not translate into situations encountered outside of a psychology laboratory (Bargh & Morsella, 2008). As subliminal EC might therefore not be considered as ecologically valid, research using a surveillance paradigm might shed more light on (automatic) associative learning in a setting assembling real world preference learning more realistically. Future research should pay attention to (1) other means to manipulate contingency awareness next to subliminal stimulus presentation and (2) the problem of real-world implications of the research.

Theories underlying EC

> "*Good theorizing is more important than anything else. If the theory you're testing is weak or logically unwarranted, best designs, methods and statistics cannot solve the fact that your predictions can't answer the questions you are asking.*" Fiedler (2018)

Going beyond the role of contingency awareness in theories underlying EC, in the next paragraph I will briefly discuss dual- and single-process models in general and possible future directions in the development and improvement of theories explaining the underlying mechanisms in EC.

Dual-process models. When discussing the additional associative process proposed by the APE model, one could critically argue that "no approach that needs two systems can be more parsimonious than an approach that proposes only one of those systems, no matter how parsimonious the second system might be" (Mitchell et al., 2009, p. 195). Contrary to that view, De Houwer (2014c) notes that a single-process model might become less parsimonious than a dual-process model when it needs auxiliary assumptions to account for effects that were previously not explained by a single—but only by a dual—process account. As an assumed second process might not necessarily mean that a model is less parsimonious it is important to take a closer look at the models in general. I will, therefore, discuss two general problems of associative processes added to the propositional process in the APE model: explanatory power and unspecificity.

Explanatory power.

When discussing an additional associative process one should keep in mind that there is clear and undisputed evidence for propositional learning in EC. In particular, the finding that the most important predictor of EC effects appears to be awareness of the CS-US contingency (Hofmann et al., 2010) demonstrates that propositional learning appears to be dominant in EC. Therefore, an additional associative learning process can only contribute little to an explanation of the underlying processes in EC (De Houwer, 2014c). However, if a phenomenon cannot be explained by propositional single-process models, a dual-process model needs to be adopted even if added value is limited. Such a phenomenon could be EC without contingency awareness, which demonstrates why the conducted studies are of great importance in the discussion between single- and dual-process models. Considering the results of the studies in the previous chapters, the explanatory power of an additional associative process might be even more limited than assumed. While a second process might be necessary, currently it only has a highly limited contribution explaining the underlying processes in evaluative conditioning. To ensure that the possibly small contribution of a second—associative—process can be assessed, the model needs to make specific predictions. In the next paragraph, I will therefore briefly look at possible problems of predictions of associative processes in the APE model.

Specificity and constraints.

As discussed in the introduction, only a theory with specific predictions can facilitate scientific progress as otherwise post-hoc alternative explanations can be given for any found effect. One general problem with current theories in psychology is that they are rarely encoded in precise terms and no explicit outcomes are predicted (Fiedler, 2004). The problem of unspecificity can also be identified in dual-process models (De Houwer, 2014c, 2014a). Gawronski et al. (2017) consider the APE model superior to single-process models as predictions are made a priori and single-process models can occasionally only account post-hoc for specific findings. While this statement might be correct, the ability to explain a variety of effects stems from the flexibility of the APE model. The following descriptions of the working mechanism of automatic processes in the associative processes demonstrates the flexibility of the theorized associative processes:

> "In line with this contention, the APE model assumes that the activation of evaluative associations in memory can indeed occur unintentionally, thereby meeting the second criterion of automaticity. **However**, it is important to note that evaluative associations can also be activated intentionally." (Gawronski & Bodenhausen, 2014b, p. 195, emphasis added)

> "Associative learning can be described as unintentional in the sense that the learning process itself does not require the goal to form a new association. **However**, associative learning can certainly have intentional antecedents" (Gawronski & Bodenhausen, 2014b, p. 196, emphasis added)

> "The APE model generally agrees with the contention that associative processes are highly efficient. **However**, this efficiency does not imply that evaluative associations cannot be activated in an effortful manner." (Gawronski & Bodenhausen, 2014b, p. 196, emphasis added)

> "According to the APE model, the formation of mental links through associative learning is resource-independent, **although** attentional distraction may sometimes disrupt associative learning" (Gawronski & Bodenhausen, 2014b, p. 197, emphasis added)

> "We argue that the activation of evaluative associations is controllable to some extent. **However**, the overall success in controlling the activation of evaluative associations is assumed to depend on the nature of the adopted control strategy." (Gawronski & Bodenhausen, 2014b, p. 197, emphasis added)

One could say that the associative processes proposed in the APE model subscribe to the working mechanisms of automaticity (Bargh, 1994), but can also explain when behavior does not abide by these principles. Additionally, the APE model proposes that implicit measures of attitude generally capture associative processes, while explicit measures of attitude capture propositional processes. This duality is challenged by findings that people can introspect on their "implicit attitudes" and are able to predict their scores on implicit measures (Hahn et al., 2014). The APE model could therefore be characterized as a combination of a very flexible associative model on top of a complex propositional model. Due to this extreme flexibility and lack of constraints the APE model cannot be falsified in practice (De Houwer, 2014c, 2014a). Only through constraints of the APE model, specific predictions can be derived and tested. Adding constraints to the APE model would constitute as an important step towards better theories in evaluative conditioning.

Single-process views. The studies presented in the previous chapters show clear support for propositional single-process views as EC was only observed when contingency awareness was present. The findings are therefore in line with a central prediction of propositional single-process views (Mitchell et al., 2009). However, similar to dual-process models, the current single-process models are unspecific and difficult to falsify, which will be discussed in the next paragraph.

Specificity and constraints.

The largest problem of propositional single-process views—as briefly mentioned in Chapter 1—is that no specific model has been proposed. Only a series of articles describe a single-process view in general and make some specific predictions (e.g., De Houwer, 2009, 2014b, 2014c, 2014a; Mitchell et al., 2009). One of the few specific predictions made by Mitchell et al. (2009) is that contingency awareness is necessary for EC (from a propositional process view), which is the prediction tested in the previous chapters. Other descriptions of the propositional single-process view are less constrained and vague and therefore do not allow for a straightforward falsification (De Houwer, 2014a).

To progress in the understanding of evaluative conditioning, a propositional single-process model needs to be introduced, that makes specific predictions, above and beyond contingency awareness. Only when setting boundary conditions and making restricted a priori predictions, a propositional single-process model can serve as a strong contender in the investigation of models explaining preference acquisition. Cognitive models of episodic memory could be one possible starting point to constrain current single-process views in an effort to create a falsifiable single-process model (Stahl & Heycke, 2016). A promising first start is the prediction of EC effects using the formalized memory model MINERVA 2 (Aust et al., 2017).

As current single- and dual-process models are too unspecific and consist merely of semantic descriptions of assumed processes, both models might, in fact, be highly similar and could be combined in one overarching model (Gawronski et al., 2017).

Conclusion

In a series of experiments with different settings, we demonstrated that preferences can be acquired towards neutral stimuli through an evaluative conditioning procedure. However, this finding was restricted to stimuli that were consciously perceived by the participants. We did not obtain any evidence for subliminal EC effects, even though the aim of the set of experiments was to provide optimal circumstances for such EC effects. The results therefore clearly suggest that contingency awareness between CS and US needs to be present for EC effects to form. It appears unlikely that EC effects can be found when one of the stimuli is presented subliminally. This finding can be considered very important when evaluating the merits of theories explaining underlying processes in EC. The findings of our experiments support propositional single-process models, which postulate that contingency awareness is necessary for EC effects to occur. The findings contradict dual-process models that postulate that EC effects can form automatically.

As discussed in the previous sections, both single- and dual-process models need to be improved in order to progress the research in the field of preference learning. Theories should be specified to predict specific outcomes resulting in a possibility to falsify them.

A second necessity related to this point is that there needs to be consensus on definitions and terminology. The field will highly benefit if researchers would agree on precise definitions of terms such as “associative” and “subliminal”. Only if researchers can agree on the same terminology and precise definitions, future theories will be less dependent on individual interpretations of concepts. Even if currently theorists might not be able to agree on the same model, the field of evaluative learning would benefit highly from a precise single-process and a precise dual-process theory with clear—and falsifiable—predictions.

This dissertation opened with a quote by Kant in which he postulated that “our knowledge begins with sense, proceeds thence to understanding, and ends with reason [...]”(Kant, 1787, p. 228). While one can be certain that Kant was not describing evaluative conditioning processes, the quote nicely fits with the empirical findings presented in this dissertation: The results of the empirical work presented here suggest that in evaluative learning, only once perceived information is understood (e.g., by propositional processes), the information will be used in reasoning and evaluations.

References

Alcohol and Tobacco Tax and Trade Bureau. (2014). *Title 27 of the Code of Federal Regulations.*

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of Social Psychology* (pp. 798–844). Worcester, MA, USA: Clark University Press.

Aust, F., & Barth, M. (2016). *Papaja: Create apa manuscripts with rmarkdown.* Retrieved from https://github.com/crsh/papaja

Aust, F., Haaf, J., & Stahl, C. (2017). A memory-based judgment account of expectancy-liking dissociations in evaluative conditioning. *PsyArXiv.* doi:10.17605/OSF.IO/TKX7B

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. doi:10.1016/0146-6402(88)90014-8

Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning. *Cognition & Emotion*, *4*(1), 3–18. doi:10.1080/02699939008406760

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. doi:10.1016/0005-7967(92)90136-5

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1990). Flavor-flavor and color-flavor conditioning in humans. *Learning and Motivation*, *21*(4), 434–455. doi:10.1016/0023-9690(90)90025-J

Baeyens, F., Wrzesniewski, A., De Houwer, J., & Eelen, P. (1996). Toilet rooms, body massages, and smells: Two field studies on human evaluative odor conditioning. *Current Psychology*, *15*(1), 77–96. doi:10.1007/BF02686936

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (pp. 1–40). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. doi:10.3758/BF03210812

Bayliss, A., Frischen, A., Fenske, M., & Tipper, S. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition*, *104*(3), 644–653. doi:10.1016/j.cognition.2006.07.012

Berridge, K. C. (2009). Wanting and Liking: Observations from the Neuroscience and Psychology Laboratory. *Inquiry*, *52*(4), 378–398. doi:10.1080/00201740903087359

Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, *42*(2), 163–176. doi:10.1016/j.jesp.2005.03.004

Borenstein, M. (Ed.). (2009). *Introduction to meta-analysis*. Chichester, U.K: John Wiley & Sons.

Both, S., Laan, E., Spiering, M., Nilsson, T., Oomens, S., & Everaerd, W. (2008). Appetitive and Aversive Classical Conditioning of Female Sexual Response. *The Journal of Sexual Medicine*, *5*(6), 1386–1401. doi:10.1111/j.1743-6109.2008.00815.x

Bradley, M. M., & Lang, P. J. (1999). International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, *52*(3), 384–389. doi:10.1037/h0041006

Britton, J. C., Taylor, S. F., Sudheimer, K. D., & Liberzon, I. (2006). Facial expressions and complex IAPS pictures: Common and differential networks. *NeuroImage*, *31*(2), 906–919. doi:10.1016/j.neuroimage.2005.12.050

Bürkner, P.-C. (2016). *Brms: Bayesian regression models using stan.* Retrieved from https://CRAN.R-project.org/package=brms

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (in press). Stan: A probabilistic programming language. *Journal of Statistical Software.*

Corneille, O., & Stahl, C. (2018). Associative Attitude Learning: A Closer Look at Evidence and how it Relates to Attitude Models. *Personality and Social Psychology Review.*

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. doi:10.1016/j.jesp.2008.08.020

Custers, R., & Aarts, H. (2005). Positive Affect as Implicit Motivator: On the Nonconscious Operation of Behavioral Goals. *Journal of Personality and Social Psychology*, *89*(2), 129–142. doi:10.1037/0022-3514.89.2.129

Davey, G. C. L. (1994). Is evaluative conditioning a qualitatively distinct form of classical conditioning? *Behaviour Research and Therapy*, *32*(3), 291–299. doi:10.1016/0005-7967(94)90124-4

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*(02), 230–241.

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

De Houwer, J. (2014a). A Propositional Model of Implicit Evaluation. *Social and Personality Psychology Compass*, *8*(7), 342–353. doi:10.1111/spc3.12111

De Houwer, J. (2014b). A propositional perspective on context effects in human associative learning. *Behavioural Processes*, *104*, 20–25. doi:10.1016/j.beproc.2014.02.002

De Houwer, J. (2014c). Why a propositional single-process model of associative learning deserves to be defended. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 530–541). New York, NY, USA: Guilford Press.

De Houwer, J., Baeyens, F., & Eelen, P. (1994). Verbal evaluative conditioning with undetected US presentations. *Behaviour Research and Therapy*, *32*(6), 629–633. doi:10.1016/0005-7967(94)90017-5

De Houwer, J., Baeyens, F., & Field, A. P. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion*, *19*(2), 161–174. doi:10.1080/02699930441000265

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

Dedonder, J., Corneille, O., Bertinchamps, D., & Yzerbyt, V. (2014). Overcoming Correlational Pitfalls: Experimental Evidence Suggests That Evaluative Conditioning Occurs for Explicit But Not Implicit Encoding of CS-US Pairings. *Social Psychological and Personality Science*, *5*(2), 250–257. doi:10.1177/1948550613490969

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. doi:10.1016/j.tics.2006.03.007

Del Re, A. C., & Hoyt, W. T. (2014). *MAd: Meta-analysis with mean differences. R Package*. Retrieved from http://cran.r-project.org/web/packages/MAd

Dickinson, A., & Brown, K. J. (2007). Flavor-evaluative conditioning is unaffected by contingency knowledge during training with color-flavor compounds. *Animal Learning & Behavior*, *35*(1), 36–42. doi:10.3758/BF03196072

Dijksterhuis, A. (2004). I Like Myself but I Don't Know Why: Enhancing Implicit Self-Esteem by Subliminal Evaluative Conditioning. *Journal of Personality and Social Psychology*, *86*(2), 345–355. doi:10.1037/0022-3514.86.2.345

Díaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition and Emotion*, *19*(2), 245–268. doi:10.1080/02699930441000300

Fazio, R. H., Chen, J.-m., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. *Journal of Experimental Social Psychology*, *18*(4), 339–357. doi:10.1016/0022-1031(82)90058-0

Fiedler, K. (2004). Tools, Toys, Truisms, and Theories: Some Thoughts on the Creative Cycle of Theory Formation. *Personality and Social Psychology Review*, *8*(2), 123–131. doi:10.1207/s15327957pspr0802_5

Fiedler, K. (2018). "We are ready to move!" An interview with Daniel Lakens and Klaus Fiedler on the current challenges in the field of psychological research. *in-mind.org*. http://www.in-mind.org/blog/post/we-are-ready-to-move-an-interview-with-daniel-lakens-and-klaus-fiedler-on-the-current.

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition & Emotion*, *25*(4), 639–656. doi:10.1080/02699931.2010.513497

Field, A. P., & Davey, G. C. (1997). Conceptual Conditioning: Evidence for an Artifactual Account of Evaluative Learning. *Learning and Motivation*, *28*(3), 446–464. doi:10.1006/lmot.1997.0980

Field, A. P., & Moore, A. C. (2005). Dissociating the effects of attention and contingency awareness on evaluative conditioning effects in the visual paradigm. *Cognition & Emotion*, *19*(2), 217–243. doi:10.1080/02699930441000292

Förderer, S., & Unkelbach, C. (2012). Hating the cute kitten or loving the aggressive pit-bull: EC effects depend on CS–US relations. *Cognition & Emotion*, *26*(3), 534–540. doi:10.1080/02699931.2011.588687

Fulcher, E. P., & Hammerl, M. (2001). When All Is Revealed: A Dissociation between Evaluative Learning and Contingency Awareness. *Consciousness and Cognition*, *10*(4), 524–549. doi:10.1006/ccog.2001.0525

Galli, M., & Gorn, G. (2011). Unconscious transfer of meaning to brands. *Journal of Consumer Psychology*, *21*(3), 215–225. doi:10.1016/j.jcps.2010.12.004

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

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

Gast, A., Langer, S., & Sengewald, M.-A. (2016). Evaluative conditioning increases with temporal contiguity. The influence of stimulus order and stimulus interval on evaluative conditioning. *Acta Psychologica*, *170*, 177–185. doi:10.1016/j.actpsy.2016.07.002

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. doi:10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the Processes Underlying Evaluation: Attitudes from the Perspective of the Ape Model. *Social Cognition*, *25*(5), 687–717. doi:10.1521/soco.2007.25.5.687

Gawronski, B., & Bodenhausen, G. V. (2011). The Associative-Propositional Evaluation Model: Theory, Evidence, and Open Questions. In *Advances in Experimental Social Psychology* (Vol. 44, pp. 59–127). Elsevier. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/B9780123855220000020

Gawronski, B., & Bodenhausen, G. V. (2014a). Implicit and Explicit Evaluation: A Brief Review of the Associative-Propositional Evaluation Model. *Social and Personality Psychology Compass*, *8*(8), 448–462. doi:10.1111/spc3.12124

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

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology*, *44*(5), 1355–1361. doi:10.1016/j.jesp.2008.04.005

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. doi:10.1016/j.jesp.2012.01.002

Gawronski, B., Balas, R., & Creighton, L. A. (2014). Can the Formation of Conditioned Attitudes Be Intentionally Controlled? *Personality and Social Psychology Bulletin*, *40*(4), 419–432. doi:10.1177/0146167213513907

Gawronski, B., Brannon, S. M., & Bodenhausen, G. V. (2017). The associative-propositional duality in the representation, formation, and expression of attitudes. In R. Deutsch, B. Gawronski, & W. Hofmann (Eds.), *Reflective and impulsive determinants of human behavior*. New York, NY: Psychology Press.

Gawronski, B., Gast, A., & De Houwer, J. (2014). Is evaluative conditioning really resistant to extinction? Evidence for changes in evaluative judgements without changes in evaluative representations. *Cognition and Emotion*, 1–15. doi:10.1080/02699931.2014.947919

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. doi:10.1214/08-AOAS191

Gilder, T. S. E., & Heerey, E. A. (2018). The Role of Experimenter Belief in Social Priming. *Psychological Science*. doi:10.1177/0956797617737128

Gorn, G. J. (1982). The Effects of Music In Advertising On Choice Behavior: A Classical Conditioning Approach. *Journal of Marketing*, *46*(1), 94–101.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*(1), 1–20. doi:10.1037/h0076157

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. doi: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*(6), 1464–1480. doi:10.1037/0022-3514.74.6.1464

Grumm, M., Nestler, S., & von Collani, G. (2009). Changing explicit and implicit attitudes: The case of self-esteem. *Journal of Experimental Social Psychology*, *45*(2), 327–335. doi:10.1016/j.jesp.2008.10.006

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of Implicit Attitudes. *Journal of Experimental Psychology. General*, *143*(3), 1369–1392. doi:10.1037/a0035028

Hammerl, M., & Grabitz, H.-J. (1993). Human evaluative conditioning: Order of stimulus presentation. *Integrative Physiological and Behavioral Science*, *28*(2), 191–194. doi:10.1007/BF02691227

Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F., & Weinberger, D. R. (2002). The Amygdala Response to Emotional Stimuli: A Comparison of Faces and Scenes. *NeuroImage*, *17*(1), 317–323. doi:10.1006/nimg.2002.1179

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). doi:10.1098/rsos.160935

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*, 1–20. doi:10.1080/02699931.2018.1429389

Hilgard, J., Engelhardt, C. R., & Rouder, J. N. (2017). Overstated evidence for short-term effects of violent games on affect and behavior: A reanalysis of Anderson et al. (2010). *Psychological Bulletin*, *143*(7), 757–774. doi:10.1037/bul0000074

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. doi: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*.

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional Versus Dual-Process Accounts of Evaluative Conditioning: I. The Effects of Co-Occurrence and Relational Information on Implicit and Explicit Evaluations. *Personality and Social Psychology Bulletin*, *43*(1), 17–32. doi:10.1177/0146167216673351

Hütter, M., & De Houwer, J. (2017). Examining the contributions of memory-dependent and memory-independent components to evaluative conditioning via instructions. *Journal of Experimental Social Psychology.*

Hütter, M., & Sweldens, S. (2013). Implicit misattribution of evaluative responses: Contingency-unaware evaluative conditioning requires simultaneous stimulus presentations. *Journal of Experimental Psychology: General*, *142*(3), 638. doi:10.1037/a0029989

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. doi:10.1037/a0026477

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. doi:10.1016/0749-596X(91)90025-F

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. doi:10.1037/a0014747

Jones, C. R., Olson, M. A., & Fazio, R. H. (2010). Evaluative Conditioning: The "How" Question. *Advances in Experimental Social Psychology*, *43*, 205–255. doi:10.1016/S0065-2601(10)43005-1

Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion*, *13*(5), 989–996. doi:10.1037/a0033164

Jraidi, I., & Frasson, C. (2010). Subliminally enhancing self-esteem: Impact on learner performance and affective state. In V. Aleven, J. Kay, & J. Mostow (Eds.), *International Conference on Intelligent Tutoring Systems* (pp. 11–20). Berlin Heidelberg: Springer.

Kant, I. (1787). *Kritik der reinen Vernunft: Zweite hin und wieder verbesserte Auflage.*

Karremans, J. C., Stroebe, W., & Claus, J. (2006). Beyond Vicary's fantasies: The impact of subliminal priming and brand choice. *Journal of Experimental Social Psychology*, *42*(6), 792–798. doi:10.1016/j.jesp.2005.12.002

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology*, *92*(6), 957–971. doi:10.1037/0022-3514.92.6.957

Kawakami, N., Miura, E., & Yoshida, F. (2015). Conscious and unconscious processes are sensitive to different types of information. *Evolution, Mind and Behaviour*, (13), 37–46.

Kerpelman, J. P., & Himmelfarb, S. (1971). Partial reinforcement effects in attitude acquisition and counterconditioning. *Journal of Personality and Social Psychology*, *19*(3), 301–305. doi:10.1037/h0031447

Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Ohio State University series on attitudes and persuasion, Vol. 4. Attitude strength: Antecedents and consequences* (pp. 1–24). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.

Krosnick, J. A., Betz, A. L., Jussim, L. J., & Lynn, A. R. (1992). Subliminal Conditioning of Attitudes. *Personality and Social Psychology Bulletin*, *18*(2), 152–162. doi:10.1177/0146167292182006

Kruglanski, A. W., Jasko, K., Chernikova, M., Milyavsky, M., Babush, M., Baldner, C., & Pierro, A. (2015). The rocky road from attitudes to behaviors: Charting the goal systemic course of actions. *Psychological Review*, *122*(4), 598–620. doi:10.1037/a0039541

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (Edition 2.). Boston: Academic Press.

Lang, D. T., & CRAN team. (2016). *RCurl: General network (http/ftp/.) client interface for r.* Retrieved from https://CRAN.R-project.org/package=RCurl

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.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, *24*(8), 1377–1388. doi:10.1080/02699930903485076

Lähteenmäki, M., Hyönä, J., Koivisto, M., & Nummenmaa, L. (2015). Affective processing requires awareness. *Journal of Experimental Psychology: General*, *144*(2), 339–365. doi:10.1037/xge0000040

Levey, A. B., & Martin, I. (1975). Classical conditioning of human "evaluative" responses. *Behaviour Research and Therapy*, *13*(4), 221–226. doi:10.1016/0005-7967(75)90026-1

Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cognition & Emotion*, *19*(2), 269–282. doi:10.1080/02699930441000319

Lipp, O. V., Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation*, *34*(3), 219–239. doi:10.1016/S0023-9690(03)00011-0

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. doi:10.1037/0097-7403.28.1.3

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy*, *1*(2), 57–101. doi:10.1016/0146-6402(78)90013-9

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442. doi:10.1006/jesp.2000.1470

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. doi:10.1177/1948550616687093

Mierop, A., Hütter, M., Stahl, C., & Corneille, O. (2018). Does attitude acquisition in evaluative conditioning without explicit CS-US memory reflect implicit misattribution of affect? *Cognition and Emotion*.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(02), 183–198. doi:10.1017/S0140525X09000855

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

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

Morey, R. D., Rouder, J. N., Pratte, M. S., & Speckman, P. L. (2011). Using MCMC chain outputs to efficiently estimate Bayes factors. *Journal of Mathematical Psychology*, *55*(5), 368–378. doi:10.1016/j.jmp.2011.06.004

Nasrallah, M., Carmel, D., & Lavie, N. (2009). Murder, she wrote: Enhanced sensitivity to negative word valence. *Emotion*, *9*(5), 609–618. doi:10.1037/a0016305

Neumann, D. L., Waters, A. M., Westbury, H. R., & Henry, J. (2008). The use of an unpleasant sound unconditional stimulus in an aversive conditioning procedure with 8- to 11-year-old children. *Biological Psychology*, *79*(3), 337–342. doi:10.1016/j.biopsycho.2008.08.005

Niedenthal, P. M. (1990). Implicit perception of affective information. *Journal of Experimental Social Psychology*, *26*(6), 505–527. doi:10.1016/0022-1031(90)90053-O

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*(5), 413–417. doi:10.1111/1467-9280.00376

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, *59*(1–2), 185–198. doi:10.1016/S1071-5819(03)00017-X

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* (Vol. 17). Oxford, England: Oxford Univ. Press.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. doi:10.1016/j.jneumeth.2006.11.017

Perruchet, P. (1985). A pitfall for the expectancy theory of human eyelid conditioning. *The Pavlovian Journal of Biological Science: Official Journal of the Pavlovian*, *20*(4), 163–170. doi:10.1007/BF03003653

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of* *social psychology* (pp. 323–390). New York, NY, USA: McGraw-Hill.

Phillips, N. (2017). *Yarrr: A companion to the e-book "yarrr!: The pirate's guide to r"*. Retrieved from https://CRAN.R-project.org/package=yarrr

Pine, A., Mendelsohn, A., & Dudai, Y. (2014). Unconscious learning of likes and dislikes is persistent, resilient, and reconsolidates. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.01051

Pleyers, G., Corneille, O., & Luminet, O. (2009). Evaluative Conditioning May Incur Attentional Costs. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(2), 279–285. doi:10.1037/a0013429

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. doi:10.1037/0278-7393.33.1.130

Pratte, M. S., & Rouder, J. N. (2009). A task-difficulty artifact in subliminal priming. *Attention, Perception, & Psychophysics*, *71*(6), 1276–1283. doi:10.3758/APP.71.6.1276

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Riketta, M., & Dauenheimer, D. (2003). Manipulating self-esteem with subliminally presented words. *European Journal of Social Psychology*, *33*(5), 679–699. doi:10.1002/ejsp.179

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308. doi:10.3758/s13423-014-0595-4

Rouder, J. N. (2016). The what, why, and how of born-open data. *Behavior Research Methods*, *48*(3), 1062–1069. doi:10.3758/s13428-015-0630-z

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default

Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. doi:10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. doi:10.3758/PBR.16.2.225

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), 995–1008. doi:10.1037/0022-3514.91.6.995

Rydell, R. J., McConnell, A. R., & Mackie, D. M. (2008). Consequences of discrepant explicit and implicit attitudes: Cognitive dissonance and increased information processing. *Journal of Experimental Social Psychology*, *44*(6), 1526–1532. doi:10.1016/j.jesp.2008.07.006

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of Two Minds Forming and Changing Valence-Inconsistent Implicit and Explicit Attitudes. *Psychological Science*, *17*(11), 954–958. doi:10.1111/j.1467-9280.2006.01811.x

Salcher, E. F. (1995). *Psychologische Marktforschung* (2., neu bearb. Aufl.). Berlin: de Gruyter.

Schemer, C., Matthes, J., Wirth, W., & Textor, S. (2008). Does “Passing the Courvoisier” always pay off? Positive and negative evaluative conditioning effects of brand placements in music videos. *Psychology and Marketing*, *25*(10), 923–943. doi:10.1002/mar.20246

Schnuerch, R., Kreitz, C., Gibbons, H., & Memmert, D. (2016). Not quite so blind: Semantic processing despite inattentional blindness. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(4), 459–463. doi:10.1037/xhp0000205

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2015). Sequential Hypothesis Testing with Bayes Factors: Efficiently Testing Mean Differences. *SSRN Electronic Journal*. doi:10.2139/ssrn.2604513

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition*, *25*(5), 638–656. doi:10.1521/soco.2007.25.5.638

Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser &

N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 436–457). Malden, MA: Blackwell.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*(3), 513–523. doi:10.1037/0022-3514.45.3.513

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. doi:10.1080/02699939008406761

Simmons, J. P., & Simonsohn, U. (2017). Power Posing: *P*- Curving the Evidence. *Psychological Science*, *28*(5), 687–693. doi:10.1177/0956797616658563

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. doi:10.1177/0956797611417632

Simonsohn, U. (2015). Small Telescopes: Detectability and the Evaluation of Replication Results. *Psychological Science*, *26*(5), 559–569. doi:10.1177/0956797614567341

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. doi:10.1037/a0033242

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a Reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General*, *144*(6), 1146–1152. doi:10.1037/xge0000104

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). *Afex: Analysis of factorial experiments*. Retrieved from https://CRAN.R-project.org/package=afex

Staats, C., & Staats, A. (1957). Meaning established by classical conditioning. *Journal of Experimental Psychology*, *54*(1), 74–80. doi:10.1037/h0047716

Stahl, C., & Heycke, T. (2016). Evaluative Conditioning with Simultaneous and Sequential Pairings Under Incidental and Intentional Learning Conditions.

Social Cognition, *34*(5), 382–412. doi: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), 286–91.

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*(9), 1107–1131. doi:10.1037/xge0000191

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. doi:10.1037/a0016196

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication Decisions Revisited: The Effect of the Outcome of Statistical Tests on the Decision to Publish and Vice Versa. *The American Statistician*, *49*(1), 108. doi:10.2307/2684823

Stroebe, W., & Strack, F. (2014). The Alleged Crisis and the Illusion of Exact Replication. *Perspectives on Psychological Science*, *9*(1), 59–71. doi:10.1177/1745691613514450

Svaldi, J., Zimmermann, S., & Naumann, E. (2012). The impact of an implicit manipulation of self-esteem on body dissatisfaction. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(1), 581–586. doi:10.1016/j.jbtep.2011.08.003

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. doi: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. doi:10.1086/653656

Todrank, J., Byrnes, D., Wrzesniewski, A., & Rozin, P. (1995). Odors can change preferences for people in photographs: A cross-modal evaluative conditioning study with olfactory USs and visual CSs. *Learning and Motivation*, *26*(2), 116–140. doi:10.1016/0023-9690(95)90001-2

Topolinski, S., & Deutsch, R. (2012). Phasic Affective Modulation of Creativity. *Experimental Psychology*, *59*(5), 302–310. doi:10.1027/1618-3169/a000159

Topolinski, S., & Deutsch, R. (2013). Phasic affective modulation of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 414–436. doi:10.1037/a0028879

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2015). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-015-0892-6

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

Van Dessel, P., De Houwer, J., Roets, A., & Gast, A. (2016). Failures to change stimulus evaluations by means of subliminal approach and avoidance training. *Journal of Personality and Social Psychology*, *110*(1), e1–e15. doi:10.1037/pspa0000039

Versluis, A., Verkuil, B., & Brosschot, J. F. (2017). Converging evidence that subliminal evaluative conditioning does not affect self-esteem or cardiovascular activity. *Stress and Health*. doi:10.1002/smi.2777

Verwijmeren, T., Karremans, J. C., Stroebe, W., & Wigboldus, D. H. (2012). Goal relevance moderates evaluative conditioning effects. *Learning and Motivation*, *43*(3), 107–115. doi:10.1016/j.lmot.2012.06.002

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. Retrieved from http://www.jstatsoft.org/v36/i03/

Vogel, T., Hütter, M., & Gebauer, J. E. (2017). Is Evaluative Conditioning Moderated by Big Five Personality Traits? *Social Psychological and Personality Science*, 194855061774019. doi:10.1177/1948550617740193

Voigt, K., Murawski, C., & Bode, S. (2017). Endogenous Formation of Preferences: Choices Systematically Change Willingness-to-Pay for Goods. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

doi:10.1037/xlm0000415

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. doi:10.1016/j.cogpsych.2009.12.001

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the Four-Picture Recognition Test. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(4), 454–459. doi:10.1037/0097-7403.32.4.454

Wickham, H. (2016). *Rvest: Easily harvest (scrape) web pages*. Retrieved from https://CRAN.R-project.org/package=rvest

Wickham, H., & Francois, R. (2016). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2017). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Retrieved from https://CRAN.R-project.org/package=tidyr

Zeelenberg, R., Wagenmakers, E.-J., & Rotteveel, M. (2006). The Impact of Emotion on Perception: Bias or Enhanced Processing? *Psychological Science*, *17*(4), 287–291. doi:10.1111/j.1467-9280.2006.01700.x

Appendix A
Literature search

Search Terms

The following search terms:
(1) evaluative conditioning
(2) evaluative learning
(3) affective conditioning
(4) affective learning
(5) attitude learning
(6) EC
were combined with these search terms:
(1) subliminal
(2) unconscious

Search approach

1. All combinations of search terms were used in an Ebsco-Host search (spring 2017) which resulted in 122 results
2. Additional articles were found in the meta-analysis by Hofmann et al., 2010
3. All abstracts were read and articles were removed that clearly did not experimentally manipulated CS or US visibility or did not use an EC paradigm
4. In total 88 articles were downloaded and 2 additional articles were found by TH
5. Studies were scanned by TH and only studies using experimental manipulation of subliminality and EC were selected (k = 19)

Appendix B

Steps data processing meta-analysis

Statistical values were selected from the publications discussed in Chapter 1 only when two means were compared by a statistical test, which was possible for 37 Experiments reported in 18 publications.

As all F tests compared two mean values, all F values were transformed to t values:

$$t = \sqrt{F}$$

For within designs d_z values were calculated and for between subject comparisons d_s values were calculated from the t values and the sample size [1]

$$d_s = 2 \times \frac{t}{\sqrt{N}}$$

$$d_z = \frac{t}{\sqrt{n}}$$

All d values were transformed to correlations [2]

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

With r being transformed back to d therefore by [3]

$$d = \pm \frac{2r}{\sqrt{1 - r^2}}$$

To acquire values on an interval scale all correlations were Fisher Z transformed [4]

$$Z = \frac{1}{2} \times \ln(\frac{1 + r}{1 - r})$$

With Z being transformed back to r therefore by

$$r = \frac{e^{2Z} - 1}{e^{2Z} + 1}$$

When more than one measure was taken in a study, a mean score was calculated from the Z values and a standard error was calculated for each Z value [5]

$$SE = \frac{1}{\sqrt{N - 3}}$$

[1] Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4.
[2] Hussy, W., & Jain, A. (2002). Experimentelle Hypothesenprüfung in der Psychologie. Göttingen: Hogrefe, Verlag für Psychologie.
[3] Berkessel, N., Berkessel, N., & Berkessel, J. (2018). Personal Communication.
[4] Bortz, J., & Schuster, C. (2010). Statistik für Human- und Sozialwissenschaftler (7., vollständig überarbeitete und erweiterte Auflage). Berlin Heidelberg: Springer.
[5] Hilgard, J., Engelhardt, C. R., & Rouder, J. N. (2017). Overstated evidence for short-term effects of violent games on affect and behavior: A reanalysis of Anderson et al. (2010). Psychological Bulletin, 143(7), 757–774.

Appendix C

Descriptive Statistics of Evaluative Measures and IAT reaction times in Experiment 1

Table C1

Unstandardized explicit ratings of the target character after Block 1 and Block 2 in Experiment 1 with either positive behavior being characteristic of the target character and negative behavior uncharacteristic ('positive') or negative behavior being characteristic of the target character and positive behavior uncharacteristic ('negative').

	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5	Rating 6	Rating 7
Block 1, positive							
Mean	7.96	8.21	7.96	7.61	8.29	7.82	84.96
SD	1.29	1.20	2.08	1.81	1.70	2.26	17.41
Block 2, negative							
Mean	2.54	2.61	2.25	2.32	2.50	2.57	21.68
SD	2.15	1.93	2.03	1.98	2.22	1.99	24.68
Block 1, negative							
Mean	1.43	1.61	1.74	1.65	1.48	2.00	12.04
SD	1.47	1.50	1.81	1.64	1.47	1.73	18.63
Block 2, positive							
Mean	7.17	7.48	7.83	7.22	8.00	7.39	78.26
SD	1.99	1.93	1.80	2.00	1.78	1.88	20.60

Note. Rating 1 = 'Unlikable-Very Likable'; Rating 2 = 'Bad-Good'; Rating 3 = 'Mean-Pleasant'; Rating 4 = 'Disagreeable-Agreeable'; Rating 5 = 'Uncaring-Caring'; Rating 6 = 'Cruel-Kind' (all 9 point scales); Rating 7 = 'feeling thermometer' (0 to 100).

Table C2

Unstandardized and untransformed IAT effects (i.e., Mean RT[target character and negative] - Mean RT[target character and positive]) in Experiment 1 with higher values indicating a more positive attitude towards the target character (in ms).

	Time 1 IAT	Time 2 IAT	Time 1 IAT	Time 2 IAT
Valence Behavior	positive	negative	negative	positive
Mean	191.94	109.73	-36.39	53.44
SD	188.25	118.96	167.1	111.25

Note. The valence of the primes was always opposite to the valence of the behavior.

Appendix D

Descriptive Statistics of Evaluative Measures and IAT reaction times in Experiment 2

Table D1

Unstandardized explicit ratings of the target character after Block 1 and Block 2 in Experiment 2 with either positive behavior being characteristic of the target character and negative behavior uncharacteristic ('positive') or negative behavior being characteristic of the target character and positive behavior uncharacteristic ('negative').

	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5	Rating 6	Rating 7
Block 1, positive							
Mean	8.39	8.03	8.10	8.00	8.32	8.39	84.65
SD	1.56	1.70	1.68	1.77	1.56	1.52	23.11
Block 2, negative							
Mean	2.06	1.81	1.71	1.71	1.81	1.74	11.13
SD	2.13	1.22	1.40	1.32	1.58	1.37	15.26
Block 1, negative							
Mean	1.58	1.58	1.46	1.58	1.50	1.58	8.38
SD	1.24	0.90	0.90	0.95	0.95	0.86	10.95
Block 2, positive							
Mean	7.42	7.15	7.31	7.23	7.77	7.69	77.73
SD	1.81	1.99	2.07	2.05	2.12	2.13	22.44

Note. Rating 1 = 'Unlikable-Very Likable'; Rating 2 = 'Bad-Good'; Rating 3 = 'Mean-Pleasant'; Rating 4 = 'Disagreeable-Agreeable'; Rating 5 = 'Uncaring-Caring'; Rating 6 = 'Cruel-Kind' (all 9 point scales); Rating 7 = 'feeling thermometer' (0 to 100).

Table D2

Unstandardized and untransformed IAT effects (i.e., Mean RT[target character and negative] - Mean RT[target character and positive]) in Experiment 2 with higher values indicating a more positive attitude towards the target character (in ms).

	Time 1 IAT	Time 2 IAT	Time 1 IAT	Time 2 IAT
Valence Behavior	positive	negative	negative	positive
Mean	227.6	59.98	43.15	94.33
SD	177.63	107.24	158.89	166.33

Note. The valence of the primes was always opposite to the valence of the behavior.

Appendix E
Mean evaluative ratings

Table E1

Mean evaluative ratings (and SDs) as a function of CS type (water vs. airline), US type (neutral, fearful face, disgusted face, disgust-eliciting IAPS picture), and CS visibility (17 ms at chance, 17 ms above chance, 1000 ms).

		US type			
CS type	Visibility	Neutral	Fearful face	Disgusted face	Disgust IAPS
Airline	17 ms, at chance	103.69	109.17	109.85	108.65
		(38.92)	(42.91)	(41.18)	(41.47)
	17 ms, above chance	106.36	101.00	103.78	96.84
		(37.31)	(40.65)	(43.48)	(34.77)
	1000 ms	116.71	106.81	104.28	102.86
		(42.94)	(43.78)	(38.15)	(50.21)
Water	17 ms, at chance	113.67	104.00	119.56	106.03
		(40.66)	(45.18)	(29.30)	(32.61)
	17 ms, above chance	118.75	124.91	113.26	101.36
		(39.22)	(41.04)	(42.72)	(48.05)
	1000 ms	116.41	110.84	115.73	97.78
		(43.59)	(42.15)	(42.53)	(54.82)

Appendix F

Sets of IADS sounds used in Experiment 1: Valence Positive, Neutral, Negative

Table F1
Sound-Nr. (Bradley & Lang, 2007)

Positive	Neutral	Negative
110	109	278
172	171	279
725	206	285
809	221	296
810	270	501
811	365	624
815	367	625
816	368	711
817	375	712
820	722	719

Appendix G

Priors for the Bayesian logistic mixed effects regression models of two-alternative forced choice responses

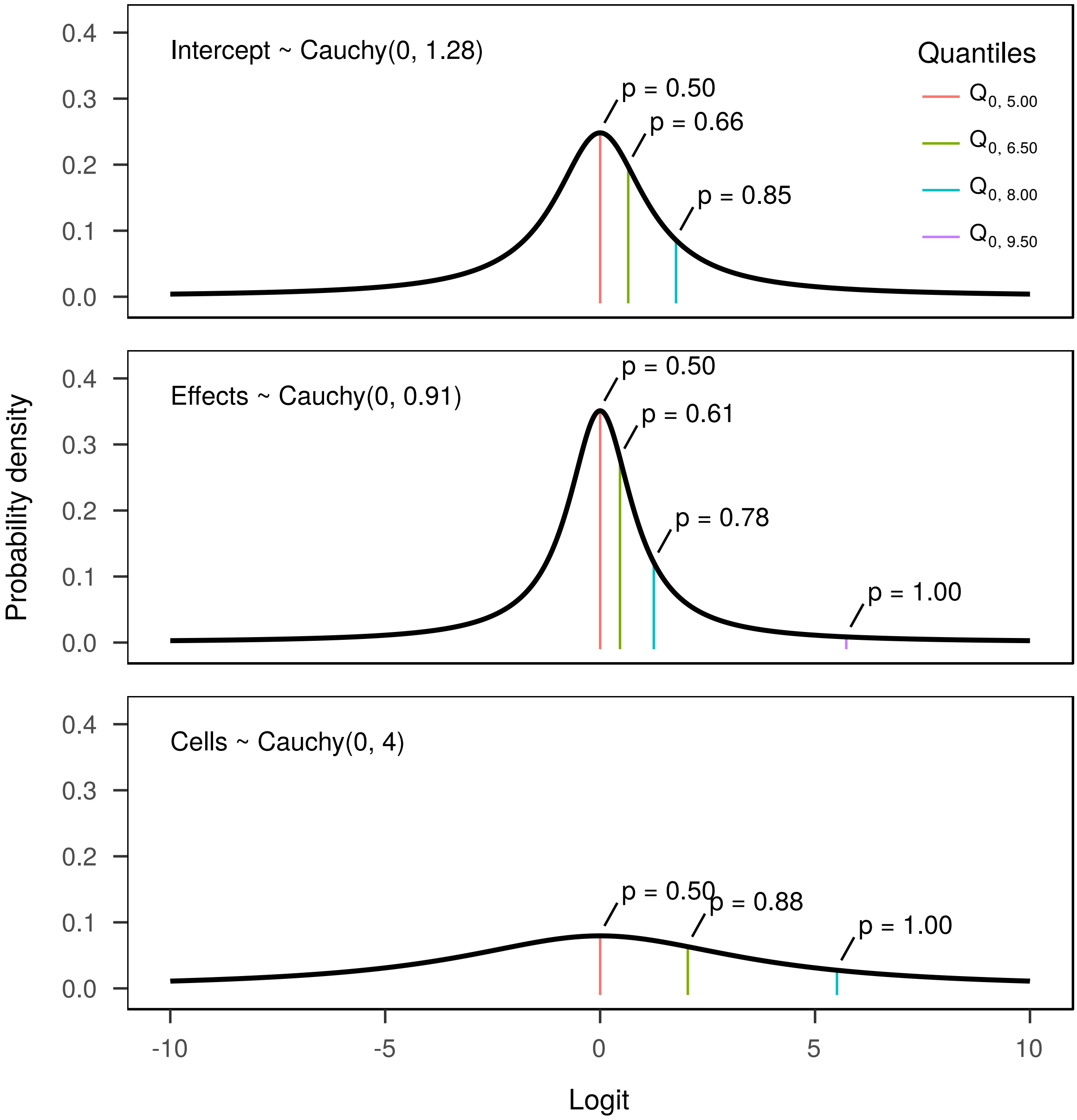

Figure G1. Priors for the Bayesian logistic mixed effects regression models of two-alternative forced choice responses. Colored lines represent distribution quantiles; annoted probabilities represent the resulting probability of choosing a positively paired CS starting from chance level (p = 0.5).

Appendix H
Mean CS visibility (Experiment 2 and Experiment 3)

Mean Visibility scores of each CS in Experiment 2 (chance level = .250, $N = 37$) and pilot of Experiment 3 (chance level = .125, $N = 7$) and the presentation time for each stimulus as used in Experiment 3.

Table H1
Mean CS visibility

CS	Visibility Study 2	Visibility Pilot	Set
03.png	.512	.400	1000 ms
08.png	.540	.329	1000 ms
14.png	.900	.657	1000 ms
22.png	.475	.400	1000 ms
04.png	.438	.200	20 ms
20.png	.400	.271	20 ms
50.png	.356	.129	20 ms
51.png	.423	.243	20 ms