UNIVERSITEIT GENT

FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN

The impact of relational information on implicit evaluation

Riccardo Zanon

Promotor: Prof. Dr. Jan De Houwer

Co-promotor: Dr. Anne Gast

Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Psychologie

2013

CONTENTS

ACKNOWLEDGEMENTS

There are many people I would like to thank, and I hope not to forget anybody. First of all, Jan De Houwer and Anne Gast, for their continuous and essential support and guidance during my doctoral training. The members of the Guidance Committee, Marcel Brass, Wim Notebaert and Adriaan Spruyt, for their important suggestions throughout the years. Then, Colin Smith, Marijke Theeuwes and Miguel Vadillo for their special help and friendship. In addition, I would like to thank the members of the Lip Lab for their collaboration and help, and also the members of the secretary office, for their courtesy and patience. Finally, I would also like to thank Bertram Gawronski, Yoav Bar-Anan, Christian Unkelbach and Dermot Barnes-Holmes for their inspiration and for the interesting discussions we had together.

GENERAL INTRODUCTION

Evaluation is considered as an important determinant of behavior and has therefore been studied extensively (e.g., Eagly & Chaiken, 1993). We define evaluation as the effect of stimuli on evaluative responses (De Houwer, Gawronski, & Barnes-Holmes, 2012). This functional definition implies that evaluation is conceptualized as an effect rather than as a mental process that mediates the impact of environment on behavior (De Houwer, 2009a; De Houwer et al., 2012; Hughes, Barnes-Holmes, & De Houwer, 2011).

Traditionally, evaluation was captured most often by self-reported evaluative judgments, that is, using evaluative responses that are given in a deliberate and controlled manner. Effects of stimuli on this type of evaluative responses are typically named explicit evaluations (De Houwer, 2009b; Gawronski & Bodenhausen, 2006). For instance, an answer to the question "On a scale from 1 to 5, how much do you like Bob?", given without time pressure, on a Likert scale would be considered an instance of explicit evaluation. During the last decade, researchers have also started to examine automatic or implicit evaluations (De Houwer, 2009b; Gawronski & Bodenhausen, 2006). Implicit evaluations differ from explicit evaluations as they occur in an automatic (i.e., unintentional, uncontrolled, unconscious, or fast; see Moors & De Houwer, 2006, for a thorough review of features of automaticity and De Houwer and Moors, 2012, for how those features can be defined functionally). Implicit evaluation have been studied in many different areas of research, from addiction (e.g., Wiers & Stacy, 2006), to clinical psychology (see Teachman, Cody, & Clerkin, 2010), close relationships (e.g., Dewitte, De Houwer, & Buysse, 2008), consumer behavior (see Perkins & Forehand, 2010), forensics (see Snowden & Gray, 2010), psychopathology (e.g., Roefs et al., 2011), politics (see Nosek, Graham, & Hawkins, 2010), and social interactions (e.g., Fazio & Olson, 2003).

In the present dissertation, we focus on the learning of implicit evaluation, that is, on how the learning history of an individual moderates the evaluative responses that are elicited by a stimulus (see De Houwer et al., 2012). One way in which implicit evaluation can be acquired is through pairings of stimuli. The change in liking of a stimulus (named conditioned stimulus or CS) due to its pairing with another stimulus (named unconditioned stimulus or US; De Houwer, 2007) is called evaluative conditioning (EC; for reviews see De Houwer, 2007; De Houwer, Thomas, & Baeyens, 2001; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). In other words, we will focus on EC of implicit evaluation, that is, a subset of EC effect that regards changes in implicit evaluation.

Associative accounts of implicit evaluation

As an effect, implicit evaluation can in principle be due to a variety of processes. Virtually all mental accounts of implicit evaluation, however, attribute these evaluations to the activation of associations in memory (for a review, see Hughes et al., 2011). Associations are defined as simple links between mental representations that are formed after the experience of repeated pairings of physical stimuli (Shanks, 2007). For instance, if a stimulus X is repeatedly paired with a picture of a spider, a mental association between the representations of “X” and of “spider” will be formed automatically. Later, if the stimulus X is encountered again, it would trigger automatically the mental representation associated to it, that is, the concept “spider”. Consequently, not only the semantic content of the mental representation will be automatically activated, but also its associated valence, which in this case is negative. In that way, the repeated pairing of the stimulus X with a picture of a spider would produce a negative implicit evaluation towards the stimulus X, and this effect would be mediated by the excitation of mental associations between the mental representations of the physical stimuli involved. This associative assumption has been fundamental to research on implicit evaluation for many years without being called into question, and has led to the development of a number of process models of evaluation. First, a number of purely associative models (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Fazio, 2007; Olson & Fazio, 2009) have been proposed. These accounts postulate that both implicit and explicit evaluations are based on the formation and activation of mental associations, formed after repeated CS-US pairings. These accounts also postulate that the strength of the object-valence association determines the evaluation's accessibility: if the representations are strongly linked, the associated representation of valence is assumed to be more accessible and can therefore be activated automatically.

In addition to those purely associative accounts, a number of dual process models (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006) have been proposed. The main characteristic of these models is that they describe two different processes that can be responsible for either implicit or explicit evaluation. According to dual process models, implicit evaluation is based on a network of mental associations, in the same fashion as postulated by the purely associative models. On the other hand, explicit evaluation is formed via the action of a propositional system of reasoning, responsible for the confirmation or disconfirmation of the validity of the automatically activated mental associations. Propositional reasoning operates on mental representations called propositions. Propositions are defined as qualified links between mental representations, which hold a truth value (De Houwer, 2009c). Hence, unlike associations, propositions specify how concepts are related (e.g., A is B, A is opposite to B, A causes B, etc.; see Lagnado, Waldmann, Hagmayer, & Sloman, 2007). In addition, propositions can be formed not only via the actual experience of pairings of stimuli, but also via instructions, inferences and intervention (see De Houwer, 2009c).

Alternative process models of implicit evaluation: the Propositional accounts

An alternative[1] to associative and dual process accounts are purely propositional, single-process models (De Houwer, 2009c; Mitchell, De Houwer, & Lovibond, 2009). Unlike dual process models, that postulate a strict division between an associative system (responsible for implicit evaluation acquisition) and a propositional system (responsible for explicit evaluation acquisition), propositional accounts postulate that every instance of evaluation and EC is mediated by propositions. As we pointed out above, propositions are qualified links between mental representations, and can hold information describing the relation between physical objects (e.g., equivalence, opposition, etc.). Dual process models already recognize the role of propositional processes in the explanation of explicit evaluation. The only necessary precondition for a purely propositional model to allow for an explanation of implicit evaluation is that propositions can be activated automatically from memory. For instance, once the proposition “the substance X heals cancer” has been formed and stored in memory, this proposition can be automatically retrieved from memory upon the next encounter with substance X, which could lead to positive evaluative responses to substance X. Interestingly, because propositions contain information about the way in which concepts are relation, single process propositional models raise the possibility that implicit evaluation might be sensitive to relational information, that is, information about the way in which stimuli are related. We will return to this issue in the next section of this introduction. In addition, purely propositional accounts of implicit evaluation postulate that the sources of information that influence implicit evaluation are not limited to stimulus pairings, but also include instructions, inference, and intervention (De Houwer 2009c). The assumptions propositional models make clearly allow for interesting new predictions in implicit evaluation research.

The present dissertation

Our research was inspired by the non-associative approach to (the formation of) implicit evaluations. Our basic assumption is that, besides co-occurrences of stimuli, also the way in which stimuli are related (i.e., relational information) is a factor that can impact implicit evaluation. If we could demonstrate an impact of relational information on implicit evaluation, this would support propositional accounts of implicit evaluation but would be difficult to account for by accounts that attribute implicit evaluation merely to associative processes. Given their nature of simple, excitatory links between mental representations, associations are structurally unable to represent relations. For example, imagine that a certain unknown substance X, present in the blood, is detected to be co-occurring with cancer. At a very superficial level, the mere contingency substance X-cancer may tell us already something, and probably would lead to a negative implicit and explicit evaluation of that substance, due to the formation and activation of the association “substance X-cancer”. However, the unknown substance X and the disease may be related in several ways. For example, the substance X could be a cause of the disease, or a harmless byproduct of it, or even an effect that signals that the disease is spontaneously healing (see Lagnado et al., 2007). Clearly, the way in which the unknown substance X is related to the disease is bound to have an important impact on how much people explicitly like the substance. But could this relational information also influence implicit evaluations of the substance? A purely associative account would exclude this hypothesis, because the association between the substance X and cancer would be the same regardless of how X and cancer are related. Therefore associative accounts of implicit evaluation would assume that, once the association substance X-cancer is mentally formed, encountering the physical substance X would trigger the excitatory link implied by the mental association, and therefore result in its negative implicit evaluation (i.e., in line with the valence of the associated representation, that is, cancer). Conversely, we argue that relational information may also play a role in (the acquisitions of) implicit evaluations. This would be evidenced by the observation that substance X evokes more negative implicit evaluations when it is said to cause cancer than when it is said to be a byproduct of cancer or an indication that the cancer is healing. In sum, the main aim of our research was to examine the impact of relational information on the acquisition of implicit evaluations, more specifically, EC of implicit evaluations.

The study of the relational determinants of implicit evaluation

Researchers only recently started to take into account the role of relational information on EC effects (Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2012; Moran & Bar-Anan, 2012; Peters & Gawronski, 2011; Zanon, De Houwer & Gast, 2012). In a particularly intriguing set of experiments, Peters and Gawronski (2011) have examined the influence of validity (i.e., true, false) information on EC effects of explicit and implicit evaluation. In their key

Experiment 3, participants were told that they had to imagine they were about to start a new job. Then, they were told that they would see the pictures of four of their future co-workers (CSs). Each picture was accompanied by either a positive behavioral description (USpos), or by a negative behavioral description (USneg), that were indicated to be representative of what the paired co-worker does or is. Participants were told that some of these behavioral statements were true and some were false. However, at this stage of the experiment, they were told to assume that every behavioral description was true, and further information about their validity would be told only afterwards. So for example, if the picture of a man named CHRIS was paired with the statement "often insults the secretary", participants had to assume that their future co-worker Chris does often insult the secretary. Two CSs were consistently paired with positive USs (resulting in the pairings Apos and Bpos), whereas two were consistently paired with negative USs (resulting in the pairings Cneg and Dneg). Importantly, in addition to the presentation of the CS-US pairings, the validity of the CS-US co-occurrences was also manipulated.

Whereas the pairings of the CSs A and C were described as true (and therefore their implications were valid), the pairings of stimuli B and D were described as false (and therefore their implications had to be mentally reversed). Alternatively, another possible way to name the two validity relations would be employing the terms "relation of equivalence" (for the true validity relation) and "relation of opposition" (for the false validity relation). To get back to the previous example, if the picture of CHRIS was paired with the statement "often insults the secretary", but this co-occurrence was stated to be false, participants had to assume that their future co-worker Chris does often *praise* the secretary. With this manipulation, Peters and Gawronski obtained four CSs that differed in terms of co-occurrence and validity information: A (positive-true), B (positive-false), C (negative-true), and D (negative-false). Importantly, they manipulated the time at which validity information was disclosed. In their *short-delay* condition, validity information was presented after each CS-US pairing. Conversely, in their *long-delay* condition, validity information was presented only twice at the end of the series of the twenty CS-US pairings. After the acquisition phase, implicit and explicit evaluations of the four CSs were assessed.

Peters and Gawronski's hypotheses were in line with the theorizations of dual process models. Hence, for both conditions, they expected the implicit evaluation of the CSs to be influenced by the mental associations formed on the basis of the CS-US co-occurrences. They did not expect validity information, mentally represented as propositions, to have an influence on implicit evaluation. In other words, they hypothesized for the positively paired CSs (A and B) to be implicitly liked more than the negatively paired CSs (C and D). Conversely, in terms of explicit evaluation, they expected the CSs to be evaluated on the basis of CS-US associations corrected by the validity propositions, and therefore that the positive-true and negative-false CSs (A and D) would be liked more than the positive-false and negative-true CSs (B and C).

Figures 1 and 2 summarize the results of their crucial study (Peters & Gawronski, 2011, Experiment 3).

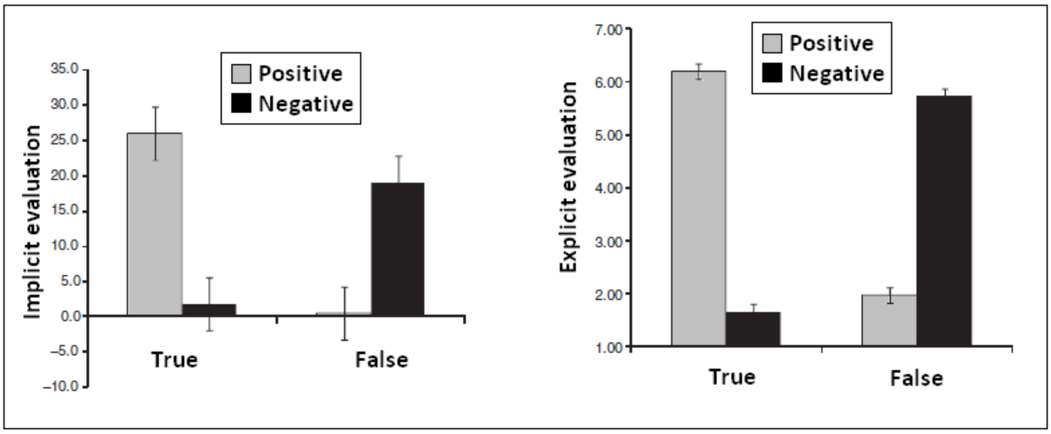

Figure 1: *Implicit and Explicit evaluation in the short-delay condition as a function of US valence (positive vs. negative) and validity information (true vs. false), (Peters & Gawronski, 2011, Experiment 3).*

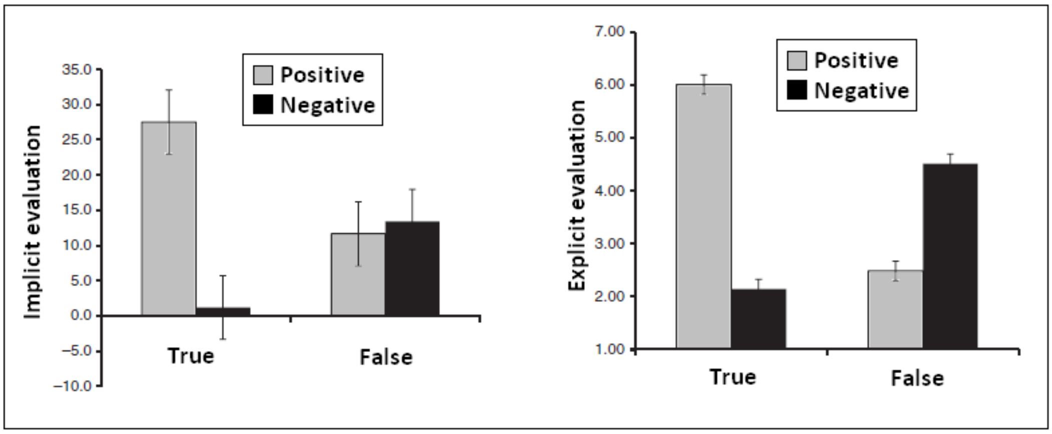

Figure 2: *Implicit and Explicit evaluation in the long-delay condition as a function of US valence (positive vs. negative) and validity information (true vs. false), (Peters & Gawronski, 2011, Experiment 3).*

However, only their explicit evaluation results confirmed their hypotheses, showing a moderation of validity information on explicit evaluation of the four CSs in both the short- and long-delay conditions. The positively paired CS was preferred to the negatively paired CSs in the true validity condition, whereas, in the false validity condition, the negatively paired CS was preferred over the positively paired CS. However, the results of their implicit measure of evaluation were somewhat unexpected. For the true validity condition, the positively paired CS was always implicitly liked more than the negatively paired CS, in both the short- and long-delay conditions. For the false validity condition, participants showed an implicit preference for the CSneg over the CSpos in the short-delay condition. In the long-delay condition, they showed only an attenuation of the EC effect, and thus no implicit preference for either of the CSs.

Peters and Gawronski's (2011; Experiment 3) implicit evaluation results are extremely intriguing. They showed that, when relational information is available at the time of the CS-US encoding (short-delay condition), it can influence implicit evaluation of the CS to a degree that it reverses the EC effect (i.e., less implicit liking for a CS paired with a positive US than for a CS paired with a negative US). Conversely, when this information is made available long after the time at which the CS-US pairs were encoded (long-delay condition), its impact is less strong, in that it does not produce a reverse EC effect but in an absence of the EC effect.

Peters and Gawronski's (2011) paper represents indeed one of the first examples available in the literature of investigation of the impact of relational information on implicit evaluation. Although the authors were inspired by a dual process view, and therefore by an associative explanation of implicit evaluation, their results are contradicting this simplistic view, by showing hints for processes of implicit evaluation acquisition that go beyond the mere association formation.

Overview of the empirical chapters

This dissertation consists of four empirical chapters, each of which aims at shedding light on whether and how relational information moderates the acquisition of implicit evaluation. In all experiments, we focused on a specific way in which implicit evaluation is acquired, that is, via pairings of stimuli (i.e., EC of implicit evaluations). In order to be able to control this process, we most often employed novel stimuli (e.g., nonwords), that were unlikely to be pre-experimentally associated with any affectively relevant concept.

In **Chapter 1**, we examined the impact of a relational context on implicit evaluation. We report three experiments, which share the same design and differ only in the implicit measure that was administered. Two compounds consisting of two nonwords each (XF and YH) were presented, paired respectively with a positive and with a negative US (resulting in the target pairings XFpos and YHneg). These target pairings were our CS-US pairings. In addition, in each condition a number of unrelated, context stimulus pairings were presented. In condition Opposite, the target pairings XFpos and YHneg were accompanied by a series of context pairings, which implied a certain rule. The pairings presented were Apos, Bpos, ABneg, Cneg, Dneg, and CDpos. The rule implied by this series of pairings was that a single cue (e.g., A), when presented alone, was followed by a certain outcome (e.g., Apos). However, when the same cue was presented in compound with another cue (e.g., AB), it was followed by the opposite outcome (e.g., ABneg). Conversely, in condition Same the series of pairings presented along with the target pairings consisted of Apos, Bpos, ABpos, Cneg, Dneg, and CDneg. In this condition, the rule implied by the context pairings stated that a single cue (e.g., A), when presented alone, was followed by the same outcome (e.g., Apos) of when it was presented in compound with another cue (e.g., ABpos). Finally, after the context pairings, in both conditions, Opposite and Same, the target CS-US pairings XFpos and YHneg were repeatedly presented. Importantly, the cues X and Y were never presented on their own, but always in compound with the cues F and H, respectively. The only difference between the two conditions consisted in the content of the rule (Opposite vs. Same), that regulated the context pairings. Therefore, applying the rule to the target compound pairings, one could infer the outcome of the single target cues X and Y, even if they were never physically presented on their own. After all (context and target) pairings were presented, implicit evaluation of the single cues X and Y was assessed, employing several implicit measures and explicit valence ratings.

This design is interesting as it allows for a comparison between the effect of mere CS-US co-occurrences and of context rules on implicit evaluation. For both conditions Same and Opposite, a purely associative view would predict that the cue X (the CSpos) would be evaluated more positively than the cue Y (the CSneg). This prediction would be based on the assumption that participants, after the experience of several XFpos and YHneg pairings, would form associations between the mental representations of the co-occurring stimuli. Therefore, the valence of the co-occurring USs would be transferred to the paired CSs, resulting in a standard EC effect for the cues X and Y, thus an implicit preference in line with the valence implied by the mere CS-US co-occurrences. However, an alternative (i.e., propositional) account could assume that not only the CS-US co-occurrences, but also the relational information implied by the two different contexts may play a role in the implicit evaluation acquisition, and therefore lead to divergent results between conditions. In condition Same, both CS-US co-occurrences and context rule have the same implications for framing the single CSs X and Y. They are only presented in compounds (XF and YH), and consistently followed by a positive and by a negative outcome, respectively. At the same time, the context rule implies that a cue that is presented in a compound would be followed by the same outcome, also when presented alone. Therefore, participants could infer that the single cue X and Y, if presented alone, would be followed by a positive and a negative outcome, respectively. Conversely, in condition Opposite, the implications deriving from the CS-US co-occurrences and from the context rule diverge. The cues X and Y are still consistently experienced in compounds (XF and YH), co-occurring with a positive US and with a negative US, respectively. However, the context rule implies that a certain cue (e.g., A), when presented alone, would be followed by an outcome (e.g., Aneg) that is the opposite of the one that follows it when it is presented in compound with another cue (e.g., ABpos). If participants learned these rules from the context pairings, and applied them to the test pairings, one would expect different implicit evaluation of the CSs X and Y across the two conditions Same and Opposite. Namely, the CSpos would be implicitly liked more than the CSneg in condition Same, whereas the CSneg would be implicitly liked more than the CSpos in condition Opposite. This hypothesis assumes that relational information influences implicit evaluation acquisition, and it could not be derived from purely associative account of implicit evaluation. The aim of Chapter 1 is to test this hypothesis, that is, whether a relational context rule can moderate the acquisition of implicit evaluation.

Chapter 2 focuses on another potential moderator of the effect of relational information on implicit evaluation, namely the implicit measures that are employed in order to assess implicit evaluation. Implicit measures create situations in which participants have to respond to a number of stimuli (see Gawronski & De Houwer, in press, for an extensive review). These situations may differ between implicit measures in terms of their procedural features and of the responses that participants are required to produce, as well as by the different conditions of automaticity that they imply. Implicit evaluation is traditionally derived from a measurement outcome (i.e., a score) that each implicit measure produces. The measurement score can be based on the combination of several factors (e.g., reaction times, errors). Therefore, it is plausible to assume that the different sub-sets of conditions implied by each implicit measure may represent potential moderators of implicit evaluation, for instance facilitating or impeding the impact of relational information.

The aim of Chapter 2 was to test the sensitivity of three widely used implicit measures to relational information: the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), the Affective Misattribution Procedure (AMP; Payne, Cheng, Govorun & Stewart, 2005) and the Evaluative Priming with a picture-picture naming task (Spruyt, Hermans, De Houwer, Vandekerckhove, & Eelen, 2006). In Experiments 1-3, we adopted a simplified version of the design employed in the experiments reported in Chapter 1. Two nonwords X and Y served as CSs. They were presented always in compound with a cue R and consistently followed by a positive (USpos) and by a negative (USneg) outcome, respectively, resulting in the target CS-US pairings XRpos and YRneg. At the beginning of the experiment, participants were told that the cue R had a special function, namely it reversed the usual outcome of the cue it was paired with. In order to exemplify this rule, a series of context pairings were presented along with the target CS-US pairings. These pairings were Apos, ARneg, Bneg, and BRpos. These pairings follow the rule that was described in the instructions: a certain cue presented alone (e.g., A) gives the opposite outcome when it is presented in compound with the reverse-cue R (e.g., Apos and ARneg). After the acquisition phase, implicit and explicit evaluation of the target cues X and Y were measured. It is important to note that the cues X and Y were always presented in compound with the cue R during the learning phase, and were presented on their own only in the measurement phase. In Experiment 1, as implicit measure of evaluation, we have employed an IAT. Conversely, in Experiments 2 and 3 we have employed an AMP. If relational information has an effect on implicit evaluation for the cues X and Y, this would lead to a reverse EC effect, that is, an implicit preference for the negatively paired cue over the positively paired cue. In addition, different results between implicit measures could tell us which implicit measures are more or less sensitive to the impact of relational information.

Experiment 4 adopted a different design, similar to that of Experiment 3 by Peters and Gawronski (2011), already described above. Briefly, four CSs were paired with either positive or negative USs, and the CS-US relation was described to be either true or false. This way, four different combinations of valence and validity were obtained, resulting in a positive-true CS, a positive-false CS, a negative-true CS, and a negative-false CS. After the acquisition phase, implicit and explicit evaluations of the four CSs were measured. Whereas Peters and Gawronski employed an evaluative priming task with an evaluative decision (EPT, Fazio, Jackson, Dunton & Williams, 1995), in our replication we employed an evaluative priming with a naming task (cf. Fazio, Sanbonmatsu, Powell & Kardes, 1986; Spruyt et al., 2007). To sum up, Chapter 2 explores whether the most widely used implicit measure of evaluation are, in different extent, sensitive to relational information. This may lead to interesting findings, and extend our knowledge about how to interpret these measurement outcomes.

The aim of **Chapter 3** was to investigate the conditions under which information that qualifies the relation between CSs and USs is effective in moderating implicit evaluation of the CSs. In particular, we were interested in how the timing in which relational information is experienced (e.g., before or after the experience of stimulus pairings) impacts implicit evaluation. In Experiment 1, a first nonword (CS1) was repeatedly paired with a word of a positive valence (USpos), and a second nonword (CS2) was repeatedly paired with a word of negative valence (USneg). Either before (condition Instructions Before) or after (condition Instructions After) the pairings, verbal instructions were presented. These instructions stated that each CS was actually the antonym of its paired US, and therefore the two were opposite to each other. For example, if the nonword LOKANTA (CS) was presented with the word HAPPY (USpos), according to the pre- or post-pairings instructions, its real meaning would be “sad” (the opposite of “happy”). After the acquisition phase, implicit and explicit evaluations of the two CSs were assessed. The aim of this experiment was to test whether the timing in which (opposite) relational instructions are presented can moderate EC effects of implicit evaluation, thus replicating the findings of Peters and Gawronski (2011).

With Experiment 2, we tested the assumption that the mere CS-US co-occurrence represents a cue that indicates that the CS and the US are equivalent, and therefore share the same valence. In particular, we aimed at changing this spontaneous framing of equivalence into a framing of opposition (i.e., the CS is framed as being opposite to the US). To this end, we presented participants with three training series of CS-US pairings. After each series, verbal instructions stated that the CSs were actually the opposite of the paired USs (in the same fashion as the instructions in Experiment 1). Aim of this manipulation was to create an environment in which the co-occurrence of stimuli becomes a cue that signals that they are oppositely (and not equivalently) related. After the three training series of pairings, a fourth, target CS-US series of pairings was presented that was followed either by instructions that stated that the CS and the US were equivalent (condition Instructions Equivalent), by instructions that stated that the CS and the US were opposite (condition Instructions Opposite) or by no instructions (condition No Instructions). After the acquisition phase, implicit and explicit evaluation of the two target CSs was measured. If the training CS-US pairings and instructions successfully changed the spontaneous framing of equivalence of the target CS-US pairings, we would expect a reduced impact of the implications of the mere CS-US co-occurrences on implicit evaluation. In addition, the manipulation of the post-pairings instructions allowed us to test the moderating influence of different relational (i.e., equivalence, opposition) instructions that one disclosed only after the CS-US pairings have been encoded.

Finally, with Experiment 3 we aimed at testing the assumption that when two propositions, based on CS-US relations, are formed one after the other, the first will have a stronger impact on implicit than on explicit evaluation. To this end, we presented participants with CS-US pairings, and with both pre- and post-pairings instructions qualifying the relation between the CSs and the USs. These instructions could be either of equivalence or of opposition. Therefore, in addition to the CS-US pairings, each participant experienced one of four possible combinations of pre- and post-pairings instructions (equivalence-equivalence, equivalence-opposition, opposition-equivalence, or opposition-opposition). In case of incongruent combinations of instructions, the second one was described as being the valid one. After the acquisition phase, implicit and explicit evaluations were assessed for the two CSs, in counterbalanced order. Our hypotheses were that, if the relative impact of the first proposition was bigger on implicit evaluations than on explicit evaluations, we would observe a bigger difference in EC between consistent conditions (i.e., equivalence-equivalence; opposition-opposition) and inconsistent conditions (i.e., equivalence-opposition; opposition-equivalence) on the explicit evaluation ratings rather than on the implicit evaluation measure.

To conclude, **Chapter 4** is a replication of a recent study by Moran and Bar-Anan (2012), who reported findings that were to some extent controversial. Although recent research has shown that relational information does, in various degrees, moderate implicit evaluation (Peters & Gawronski, 2011; Zanon et al., 2012), Moran and Bar-Anan reported two experiments in which relational information impacted only on explicit evaluation, while being totally ineffective in influencing implicit evaluation. In their original study, Moran and Bar-Anan presented four CSs, each of which started or ended the presentation of either a pleasant sound (USpos) or of an aversive sound (USneg). Hence, their design consisted in a starting-positive CS, a stopping-positive CS, a starting-negative CS, and a stopping-negative CS. Their assumption was that a CS that starts an US acquires a valence in line with the US. Conversely, their second assumption was that a CS that ends a certain US would acquire a valence opposite to the US. After the acquisition phase, they measured implicit and explicit evaluations of the four CSs. Interestingly, their results on the explicit evaluations showed a moderation of the relational information: the CS that started a positive sound was evaluated more positively than the CS that stopped it. In addition, the CS that stopped a negative sound was evaluated more positively than the CS that stopped a positive sound. However, their implicit evaluation measures showed no moderation of the relational information: the CSs that were paired with the USpos were always evaluated more positively than the CSs that co-occurred with the USneg, regardless of whether they were presented at their beginning or at their end. These results support a dual process view of evaluation, in which explicit evaluation is influenced by higher-order propositional processes (capable to encode relations), and implicit evaluation is based on the mere content of the CS-US co-occurrences, encoded in form of associations.

However, in Chapter 4 we replicated this experiment, adding two, in our opinion, crucial procedural modifications (regarding the instructions and the implicit measure) to the original Moran and Bar-Anan's design. Our aim was to test whether, with the changed design, we could find an impact of relational information not only on explicit, but also on implicit evaluation.

References

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

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230-241. http://www.ucm.es/info/psi/docs/journal/

De Houwer, J. (2009a). Comparing measures of attitudes at the procedural and functional level. In R. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures*. Erlbaum.

De Houwer, J. (2009b). How do people evaluate objects? A brief review. *Social and Personality Psychology Compass, 3*, 36-48. doi: 10.1111/j.1751-9004.2008.00162.x

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

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

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2012). *A functional-cognitive framework for attitude research*. Manuscript submitted for publication.

De Houwer, J., & Moors, A. (2012). How to define and examine implicit processes? In R. Proctor & J. Capaldi (Eds.), *Implicit and explicit processes in the psychology of science*. New York, NY, USA: Oxford University Press.

Dewitte, M., De Houwer, J., & Buysse, A. (2008). On the role of the implicit self-concept in adult attachment. *European Journal of Psychological Assessment*, *24*, 282-289. doi: 10.1027/1015-5759.24.4.282

Eagly and Chaiken, (1993). *The Psychology of Attitudes*, Fort Worth, TX: Harcourt Brace Jovanovich.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strengths. Social Cognition, 25, 603–37. doi: 10.1521/soco.2007.25.5.603

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

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

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297-327. doi: 10.1146/annurev.psych.54.101601.145225

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

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

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

Gawronski, B., & De Houwer, J. (*in press*). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), Handbook of research methods in social and personality psychology (2nd edition). New York: Cambridge University Press.

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

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.

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

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorising in implicit attitude research: Propositional and behavioral alternatives. *Psychological Record, 61*, 465–496. http://thepsychologicalrecord.siuc.edu/

Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to Our Functional Roots When Exploring New Intellectual Islands: A Voyage through Implicit Cognition. Journal of Contextual Behavioural Science,1, 17-38. doi: 10.1016/j.jcbs.2012.09.003

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y. & Sloman, S. A. (2007) Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*, (pp. 154–72). Oxford University Press.

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

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326. doi: 10.1037/0033-2909.132.2.297

Moran, T. & Bar-Anan, Y. (2012). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion,* 1-10. doi: 10.1080/02699931.2012.732040

Nosek, B. A., Graham, J., & Hawkins, C. B. (2010). Implicit political cognition. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition* (pp. 548-564). New York, NY: Guilford.

Olson, M. A., & Fazio, R. H. (2009). Implicit and explicit measures of attitudes: The perspective of the MODE model. In R. E. Petty, R. H. Fazio, & P. Briñol

(Eds.), Attitudes: Insights from the new implicit measures (pp. 19-63). New York, NY: Psychology Press.

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

Perkins, A., & Forehand, M. (2010). Implicit social cognition and indirect measures in consumer behavior. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition* (pp. 535-547). New York, NY: Guilford.

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

Roefs, A., Huijding, J., Smulders, F. T. Y., MacLeod, C. M., De Jong, P. J., Wiers, R. W., & Jansen, A. T. M. (2011). Implicit measures of association in psychopathology research. *Psychological Bulletin, 137*, 149-193. doi: 10.1037/a0021729

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

Shanks, D. R., (2007). Associationism and cognition: human contingency learning at 25. *Quarterly Journal of Experimental Psychology, 60, 291-309*. doi: 10.1080/17470210601000581

Snowden, R. J., & Gray, N. S. (2010). Implicit social cognition in forensic settings. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition* (pp. 522-534). New York, NY: Guilford.

Spruyt, A., Hermans, D., De Houwer, J., Vandekerckhove, J., & Eelen, P. (2007). On the predictive validity of indirect attitude measures: Prediction of consumer choice behavior on the basis of affective priming in the picture-picture naming task. *Journal of Experimental Social Psychology, 43,* 599-610. doi: 10.1016/j.jesp.2006.06.009

Teachman, B. A., Cody, M. W., & Clerkin, E. M. (2010). Clinical applications of implicit social cognition theories and methods. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition* (pp. 489-521). New York, NY: Guilford.

Wiers, R. W., & Stacey, A. W. (2006). Implicit cognition and addiction. *Current Directions in Psychological Science, 15*, 292-296. doi: 10.1111/j.1467-8721.2006.00455.x

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

Footnotes

[1] As pointed out by Hughes et al. (2011), there are also functional alternatives to associative models of implicit evaluations, alternatives which stem from Relational Frame Theory (RFT, Hayes, Barnes-Holmes & Roche, 2001; see Hughes, Barnes-Holmes & Vahey, 2012, for a detailed discussion). However, in the present dissertation, we decided to focus on cognitive models of evaluation and will therefore not discuss the non-cognitive, functional RFT approach. Nevertheless, we are convinced that the functional RFT approach can contribute in important ways to future developments in research on implicit evaluation.

CHAPTER 1

CONTEXT EFFECTS ON IMPLICIT EVALUATION [1]

[1] Based on Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation, 43*, 155-165. doi: 10.1016/j.lmot.2012.02.003

Introduction

Because preferences drive behavior, cognition, and emotion, it is important to understand how those likes and dislikes are formed and how they can be influenced. Research has shown that people do not necessarily evaluate stimuli in a conscious, controlled and intentional manner. As Zajonc (1980) argued in his seminal paper, evaluations can arise also in a spontaneous, uncontrolled, unconscious, efficient and fast manner. We refer to this type of automatic preferences as implicit evaluations, whereas we refer to non-automatic preferences as explicit evaluations.[1] Implicit evaluations have been shown to play a crucial role in many important psychological phenomena including psychopathology (see Roefs et al., 2011), addiction (Wiers & Stacy, 2006), and social interactions (Fazio & Olson, 2003).

Implicit evaluations often arise as the result of repeated experiences. For example, the implicit evaluation of a neutral stimulus (conditioned stimulus, CS) might be changed by repeatedly pairing it with an affectively relevant (positive or negative) stimulus (unconditioned stimulus, US). A change in liking that occurs as the result of pairings of stimuli is typically referred to as evaluative conditioning (EC; De Houwer, 2007; De Houwer, Thomas, & Baeyens, 2001; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). EC can involve changes in both explicit and implicit evaluations. Hence, implicit evaluations that result from the repeated pairing of stimuli are a subset of EC effects, namely those EC effects that involve changes in automatic rather than non-automatic evaluations.

At a mental process level, it is often assumed that EC of both implicit and explicit evaluations is due to the slow and gradual formation of associations in mind, that is, of unqualified links between mental representations. This view is endorsed by several associative (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992) and dual process models (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006). Dual process models of evaluations postulate the existence of two separate but interacting mental systems of processing: an associative, impulsive system which is responsible for the formation of implicit evaluations and a propositional, reflective system which is responsible for the formation of explicit evaluations (e.g., Rydell & McConnell, 2006). The associative system consists of a network of mental representations that are linked via unqualified associations. Associations are typically assumed to result from the direct experience of stimulus pairings. Once an association has been formed between the CS and US representations as the result of CS-US pairings, presentation of the CS not only results in the activation of the CS representation but also, via automatic spreading of activation, in the automatic activation of the US representation. As a result, the originally neutral CS will evoke implicit evaluations that are in line with the valence of the US. Under certain conditions, the automatic activation of the US representation can bias explicit CS evaluations as well and thus lead to EC of explicit evaluations (see Gawronski & Bodenhausen, 2006, for details). In sum, dual process models typically locate the source of EC effects, particularly those involving implicit evaluations, within an associative system.

Recently, single process models have been proposed that reject the existence of an associative system. More specifically, propositional models attribute all instances of EC (and other types of associative learning[2]) to the non-automatic formation of propositions about stimulus relations (e.g., De Houwer, 2007, 2009c; Mitchell, De Houwer & Lovibond, 2009). Propositions are statements about the world that can be true or untrue. They can specify not only that two stimuli are related, but also the way in which the stimuli are related (e.g., that the stimuli simply co-occur or that one stimulus causes another stimulus). Because the impact of stimulus pairings on evaluative and other responses is, according to propositional models, mediated by the formation of propositions, EC can occur only after a proposition about the stimulus pairings has been formed and will depend on the content of the proposition that is formed.

Propositional models also incorporate several other assumptions that generate a number of interesting predictions. First, the formation of propositions is assumed to be a non-automatic process that requires awareness and cognitive resources. Hence, propositional models postulate that EC and other types of associative learning should depend on awareness of the stimulus pairings. Although there have been reports of unaware EC, questions have been raised about the validity of these findings (e.g., Lovibond & Shanks, 2002; Mitchell et al., 2009), and many studies have failed to find EC effects in the absence of contingency awareness (e.g., Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Stahl, Unkelbach, & Corneille, 2009). Moreover, a recent meta-analysis showed that awareness of the stimulus pairings was by far the biggest moderator of the size of EC effects (Hofmann et al., 2010). Second, it is assumed that propositions can be formed not only on the basis of direct experience but also on the basis of instructions or inferences (De Houwer, 2009c). In line with this assumption, De Houwer (2006) showed that merely informing participants about stimulus pairings without actually presenting these pairings is enough to induce EC effects (see also Gast & De Houwer, 2012). Third, it has been proposed that after the non-automatic formation of a proposition, the proposition can be stored in memory from which it can be retrieved automatically (e.g., Bar-Anan, De Houwer, & Nosek, 2010; Hughes, Barnes-Holmes, & De Houwer, 2011). This latter assumption allows propositional models to explain EC of implicit evaluations.

In the present paper, we report studies that were inspired by the central idea of propositional models, namely that the effect of CS-US pairings on liking should depend on the content of the proposition about the CS-US relation rather than on the pairings themselves. Importantly, the same objective CS-US pairings could lead to propositions about the CS-US relation with different content depending on the context in which those pairings are presented. For instance, if a neutral stimulus co-occurs with a positive stimulus (e.g., a neutral nonword and a positive word) in a context in which this implies that the neutral stimulus actually has negative properties (e.g., when the nonwords are said to be antonyms of the existing words), such co-occurrences should lead to the formation of negative propositions about the initially neutral stimulus (e.g., the nonword has a negative meaning). As a result, the originally neutral stimulus should become negative even though it was paired with positive stimuli. Recent studies support this prediction. For instance, Förderer and Unkelbach (2011; see also Fiedler & Unkelbach, 2010) showed that neutral faces paired with positive pictures were rated *less* positive than neutral faces paired with negative pictures when participants were told that the depicted people loathed the pictures they were paired with. At a functional level, studies such as these are important because they show that the effect of stimulus pairings on liking depends on the (verbal) context in which they are presented. At the level of mental process theories, they provide support for propositional models of EC.

It is a different question, however, whether this type of context effects can be found also for EC of implicit evaluations. Assuming that in particular implicit evaluations are determined by associations in memory and that associations in memory are determined by actual stimulus pairings (e.g., Rydell & McConnell, 2006), one could predict that EC of implicit evaluations will reflect stimulus pairings independent of what the context implies about the way in which the CS and US are related. Therefore, observing context effects on EC of implicit evaluations would constrain theories about the determinants of implicit evaluations more than observing context effects on EC of explicit evaluations.

A recent study suggests that contextual information about the way CS and US are related can also moderate EC of implicit evaluations. Peters and Gawronski (2011) presented pictures of four neutral faces (CSs), two paired with a number of positive descriptions (USs+), and two paired with a number of negative descriptions (USs-). After each trial, participants were explicitly told whether the pairing they had just seen was true or false. If the pairing was said to provide false information, participants were asked to “mentally reverse” the pairing (e.g., when someone was described as SAD, they should infer that this person was HAPPY). Employing an affective misattribution procedure (AMP, Payne, Cheng, Govorun & Stewart, 2005) and an affective priming task (Fazio, Jackson, Dunton, & Williams, 1995) to capture implicit evaluations of the CSs, they found that EC effects depended not only on the valence of the USs but also on the relational information (true vs. false) that was presented right after each pairing. In particular, a standard EC effect (i.e., preference for CSs paired with positive USs over CSs paired with negative USs) was found when the pairings were said to provide correct information. However, when pairings were said to provide false information, a reversed EC effect was observed (i.e., more negative rating for CSs paired with positive USs than for CSs paired with negative USs).

In the current research, we wanted to investigate further whether and when EC of implicit evaluations is moderated by the context in which pairings are presented. Unlike earlier studies (Förderer & Unkelbach, 2011; Peters & Gawronski, 2011), we manipulated the context in an indirect way, that is, without verbally instructing participants about how the CS and US are related. Moreover, our context manipulation was embedded in a traditional EC procedure that involved pairings of single stimuli as CSs and USs rather than statements or personality trait descriptions as used by Peters and Gawronski (2011). More specifically, we manipulated relational information by presenting the target CS-US pairings in the context of other pairings that followed a certain rule. A first CS X was always presented in compound with cue F and was always paired with a positive US (i.e., winning a game; XF+). A second CS Y always occurred in compound with cue G and was always paired with a negative US (i.e., losing a game; YG-). In addition to these target compounds, we also presented context cues and manipulated the rule that determined when a context cue was paired with a win or a loss. In condition Same, a context cue was always followed by the same US regardless of whether it was presented alone or in compound with another cue (i.e., A+, B+, AB+, C-, D-, CD-). In condition Opposite, however, a context cue was followed by a different US when it was presented on its own than when it was presented in compound with another cue (i.e., A+, B+, AB-, C-, D-, CD+). Our dependent measure was the implicit and explicit evaluation of the two target stimuli X and Y when presented on their own (and thus in the absence of the cues F and G with which they appeared on compound trials).

Different mental process models of EC on implicit measures make different predictions regarding the outcome of our studies. Regarding the condition Same, all currently available models make the same predictions.

Although postulating different mediating processes, in functional terms they would all expect a standard EC effect, that is, a change in implicit evaluation of the CSs that reflects the valence of the USs they co-occurred with. The models do, however, make different predictions about whether this effect will also be found in condition Opposite. A purely associative account of EC of implicit evaluations implies that the change in implicit evaluation is driven simply by associations that are formed between stimuli that co-occurred. Importantly, in both conditions, the target stimulus X was paired with a positive outcome (XF+) whereas target stimulus Y was paired with a negative outcome. Because the pairings involving X and Y were identical in both conditions, the implicit evaluations of the target stimuli X and Y should not differ between the conditions Same and Opposite. In both cases, X should be evaluated more positively than Y.

A propositional account, on the other hand, would predict different EC effects in condition Same than in condition Opposite. In condition Same, the context pairings (A+, B+, AB+, C-, D-, CD-) imply that a stimulus is paired with the same US when presented on its own and when presented in compound with another stimulus. Based on this information, participants can infer from the XF+ and YG- trials that X on its own will be followed by a positive outcome and that Y on its own would be followed by a negative outcome. Hence our dependent measure (i.e., the evaluation of X and Y in the absence of their paired compound cues F and G) should reflect the valence of the paired USs. In condition Opposite, however, the context pairings (A+, B+, AB-, C-, D-, CD+) imply that a cue is paired with different outcomes when presented alone than when presented in compound. Hence, the XF+ pairings imply that X on its own will be followed by a negative outcome whereas Y would be followed by a positive outcome. Therefore in this case, the evaluation of X and Y when presented on their own, should reflect the opposite of the valence of the US with which they were paired. Provided that participants indeed use the context trials to form propositions about the X-US and Y-US relations, and transfer the rule from the context pairings to the target pairings, propositional models would thus predict that participants will prefer X over Y in condition Same but will prefer Y over X in condition Opposite.

Finally, a third possible scenario for condition Opposite would be a significant reduction of the EC effect compared to condition Same but not a reversal. This result would allow for two theoretical interpretations. A propositional account of such a result would entail that participants form two competing propositions, one based on the experienced co-occurrences and another based on the rule implied by the context pairs. Whereas these two propositions would lead to the same preferences in condition Same (e.g., “X goes with F and with win” and “X on its own goes with win”) this would not be the case for condition Opposite (e.g., “X goes with F and with win” and “X on its own goes with loss”). This might result in a preference for X over Y in condition Same but not in condition Opposite. Such a pattern of results, however, would also be in line with a hybrid account that assumes that both associative and propositional processes contribute to implicit EC effects. Whereas both processes would result in a preference for X over Y in condition Same, they would oppose each other in condition Opposite. For instance, the XF+ pairings would result in an association between X and winning (and thus a liking of X) but in the proposition that X on its own will be followed by a loss (and thus a disliking of X). Hence, in condition Opposite, X will not be liked more than Y.

In sum, if we observe that X and Y are evaluated differently in condition Same than in condition Opposite (either reversed or reduced effects in condition Opposite), this would support the idea that propositional processes (co-)determine EC of implicit evaluations. In addition, it would go beyond the study of Peters and Gawronski (2011) by showing that context effects in EC of implicit evaluations are not limited to the rather atypical EC procedure that they used in which USs were explicitly said to provide a true or false description of the personality of the CSs.

Experiments 1, 2, and 3

In three studies, we examined context effects in the following manner (see De Houwer & Vandorpe, 2010, for a similar approach in research on causal learning). Participants were informed that on each trial of a learning phase, they would see the picture of a slot machine frame on the screen. The slot machine had two displays, one at its top and one at its bottom. On each trial, the top one displayed one or two neutral nonwords, whereas the bottom one displayed an outcome that could either be a win or a loss. Participants were instructed that each nonword or compound of nonwords would be the cause of the respective outcome, which could either be a win or a loss. We refer to different nonwords with different letters (A-G, X and Y), to a win with "+", and to a loss with "-". In condition Opposite, participants experienced the pairings A+, B+, AB-, C-, D-, CD+, XF+, YH-. In condition Same, participants experienced the pairings A+, B+, AB+, C-, D-, CD-, XF+, YH-. The stimuli X and F which are in the learning phase always presented in compound with the stimuli, F and H, respectively, serve as CSs. After the repeated presentation of these pairings, we measured the implicit evaluations of the single stimuli X and Y, employing a range of implicit measures, that is, measures of implicit evaluations. More specifically, in Experiment 1, we used the implicit association test (IAT; Greenwald, McGhee, and Schwartz, 1998), in Experiment 2 the personalized IAT (Olson & Fazio, 2004), and in Experiment 3 the affective priming task (Fazio et al., 1995). We employed a variety of implicit measures to ensure that our conclusions were not specific to one particular measure.

Method

We report all three studies in the same section because in all three studies we employed the same procedure during the learning phase. The only major difference between experiments concerned the nature of the measure of implicit evaluations.

Participants. All participants were students at Ghent University. Fifty-six students participated in Experiment 1 (mean age = 18.64, *SD* = 1.77; 68% women), 46 in Experiment 2 (mean age = 19.63, *SD*= 2.86; 83% women); and 50 in Experiment 3 (mean age = 19.64, *SD*= 2.75; 76% women). All participants were native Dutch speakers. For their participation in the experiments, they were given either course credits or four Euros.

Materials.

Materials Learning phase. Each cue and outcome appeared on one of two displays of a slot-machine picture, which was presented on a computer screen in the size of 26 (height) x 17 (width) cm. The cues appeared on a display at the top, whereas the outcomes appeared on a display at the bottom. In all three experiments, eight nonsense words (“BAYRAM”, “ENANWAL”, “UDIBNON”, “KADIRGA”, “LOKANTA”, “SARICIK”, “FEVKANI” and “NIJARON”) were used as stimuli during the learning phase. The nonwords “LOKANTA” and “FEVKANI” were the target stimuli, whereas the other nonwords constituted the set of context stimuli. All context and target stimuli were written in black, upper case letters, and in the font “Arial Black”, font size 40. We used nonsense words to avoid any potential pre-experimental association between our stimuli and the outcomes. The positive outcome was the Dutch word for “win” (“winst”), presented in green color, upper case letters, and in the font “Britannic Bold”, font size 100.The negative outcome was the Dutch word for “loss” (“verlies”), presented in red color, upper case letters, and in the font “Haettenschweiler”, font size 100. The positive outcome was presented together with a pleasant sound (a soft melody) and several pictures of 2 euro coins. The negative outcome was presented together with an unpleasant sound (a loud buzzer) and two cartoon pictures of a sad face.

Materials Measurement phase. In Experiment 1, we employed an implicit association test designed to measure the evaluation of the target words (IAT, Greenwald et al., 1998). In this IAT, the attribute stimuli referring to the category positive were the Dutch words for “happy” (“gelukkig”), “honest” (“eerlijk”), "pleasant" ("prettig") and "sincere" ("oprecht"). The attribute stimuli for the category negative were the Dutch words for "mean" ("gemeen"), "rude" ("brutaal"), "aggressive" ("agressief") and "deceptive" ("bedrieglijk"). The target stimuli were the cues X and Y ("LOKANTA" and "FEVKANI"). As we were interested in the single nonwords X and Y, we deviated from the standard IAT procedure by using only one stimulus for each of the target categories. To avoid that stimuli were classified only on the basis of simple perceptual features, each target stimulus was presented in four different fonts (lower case Arial Black, upper case Arial Black, lower case Fixedsys, and upper case Fixedsys), resulting in 8 different target stimuli. De Houwer (2006) showed the reliability of this procedural modification and its suitability to capture recent learning.

In Experiment 2, we employed a personalized IAT (Olson & Fazio, 2004) which differed from the standard IAT in two important ways. First, the attribute labels were "I like" ("Heb ik graag") and "I don't like" ("Heb ik niet graag"), instead of "positive" and "negative", and second, there was no error feedback for the stimuli of the attribute categories. In addition to that, the attribute stimuli were selected in such a way that their valence could easily be judged in a personal (e.g., non-normative) manner. We used the Dutch words for "holiday" ("vakantie"), "summer" ("zomer"), "gift" ("cadeau"), "party" ("feest"). The attribute stimuli for the negative category were the Dutch words for "war" ("oorlog"), "vomit" ("braaksel"), "accident" ("ongeluk"), "divorce" ("scheiding"). As in Experiment 1, the target stimuli were different perceptual instantiations of the nonwords "lokanta" and "fevkani".

In Experiment 3, we applied an affective priming procedure (Fazio, et al., 1995). As targets we used the positive and negative attributes from Experiment 1. The nonwords X and Y were used as primes.

All tasks were presented on an Intel Core2 Duo PC with a 19'' 100 Hz monitor, screen resolution 1280 by 1024 pixels, and implemented using custom made Inquisit 2.0 programs. For each implicit measure, the responses were given by pressing the key A (left key) or the key P (right key) on a standard AZERTY keyboard.

Procedure. At the beginning of each experiment, participants were informed that they would see a series of trials in a slot machine game. They were told that each trial consisted of the presentation of one or two slot-machine words, followed by a winning or losing outcome. Their task was to identify which of the presented slot-machine words predicted a win and which predicted a loss. They were asked to detect and remember this information because it was important for another upcoming task. Finally, we took measures to ensure that participants would register the rule that determined the outcome of the context cues. From the perspective of propositional models, detection of the context rule is a precondition for finding an impact of this context element on EC. In all experiments, participants were therefore told that a hidden rule regulated the presentation of the slot-machine words and their outcome and were asked to try to discover the rule from the series of stimuli. In Experiment 2, we implemented an additional measure by actually informing the participants about the nature of the rule. That is, in condition Opposite, participants were told that words presented in compound result in the opposite outcome than when presented alone. In condition Same, they were told that words presented in compound produce the same outcome as when presented alone. Although these measures imply that we cannot draw strong conclusions about whether participants can learn these kinds of rules in a spontaneous manner, they are instrumental for determining whether the rules that operate in a certain context can moderate EC of implicit evaluations. In any case, unlike the case in previous related studies (e.g., Förderer & Unkelback, 2011; Peters & Gawronski, 2011), participants were not directly instructed about how CS-US pairings should be interpreted.

After the instruction page, participants could start the learning task by pressing a key. Each trial of the learning task represented a slot-machine game. A trial started with the presentation of the picture of a slot-machine, together with one or two nonword(s) which appeared at the top display position for 2000 ms. Then the outcome (the Dutch word for “win” or “loss”) was added on a second display at the bottom of the slot-machine, together with the pleasant or unpleasant sound. The nonword(s) and the outcome remained on the screen together for 3000 ms during which the sound was also present. The next trial started after an inter trial interval of 3000 ms during which the screen was blank and no sounds were played. The sequence of trials presented was the following. First, A+, B+, and AB+ (Condition Same) or A+, B+, and AB- (Condition Opposite) trials were presented twice each, resulting in a block of six trials (Block 1). This block was followed by C-, D-, and CD- (Condition Same) or C-, D-, and CD+ (Condition Opposite) trials, each presented twice, resulting in another block of six trials (Block 2). The series of pairings of Block 1 and Block 2 constituted the context, that varied between conditions. Finally, for both condition Same and Opposite, the target trials XF+ and YH- were presented two times each, resulting in a final block of four trials (Block 3). Block 3 constituted the series of target pairings of our design. This sequence of three blocks of trials was presented five times, adding up to a learning phase of eighty trials. Exclusively in Experiment 2, the trials of Block 1 (the pairings of the stimuli A, B and AB) and of Block 2 (the pairings of the stimuli C, D and CD) were only presented one time per block, instead of twice (like in Experiment 1 and 3). This was an attempt to shorten the learning phase and thus keep participants focused during the whole learning phase. The trial order within blocks was determined randomly for each participant and block separately, and the assignment of the cues X or Y to the positive or negative outcome was counterbalanced across participants.

After the presentation of all learning trials, participants performed an implicit measure with which we aimed to assess the implicit evaluation of the target stimuli X and Y. In Experiment 1, we applied an IAT (Greenwald et al, 1998). Our IAT consisted of seven phases: a target discrimination phase (B1), an attribute discrimination phase (B2), a combined (targets and attributes) practice phase (B3), a combined (targets and attributes) phase (B4), a reversed target discrimination phase (B5), a reversed combined practice phase (B6), and a reversed combined phase (B7). Phases (B1), (B2), and (B5) served only for practice, whereas Phases (B3), (B4), (B6) and (B7) were the combined test phases that served as measures of implicit evaluation. Before each phase, participants were informed about the assignment of the different categories to the left and right key. During the target and reversed target discrimination phase (B1 and B5), each target stimulus was presented four times, resulting in 32 trials. During the attribute discrimination phase (B2), each attribute stimulus was presented four times, also resulting in 32 trials. Finally, both combined phases consisted of two blocks (practice – B3 and B6 - and test phase – B4 and B7) of 32 trials each. In each block of these phases, each attribute and target stimulus was presented twice. During the reversed combined phase, the assignment of targets to response keys was opposite to the assignment during the other combined phase. The assignment of the target categories to the two response keys (e.g., during Phases B1, B3, and B4, press left for X and right for Y or vice versa) was counterbalanced across participants. Whether the X-positive / Y-negative task (e.g., X assigned to the same key as positive attributes and Y assigned to the same key as negative attributes) was presented before or after the X-negative / Y-positive IAT task (e.g., X assigned to the negative key and Y assigned to the positive key) was thus also counterbalanced across participants. The order of trials in each phase was determined randomly. On each trial, a word was presented in the center of the screen until a valid response (pressing key A or P) was registered. If the response was correct, the next word appeared after 400 ms. If the response was incorrect, a red cross was presented for 400 ms, also in the center of the screen. In this case, the next word was presented 400 ms after the red cross had disappeared.

In Experiment 2, we employed a personalized IAT (see Olson & Fazio, 2004). The procedure was identical to the standard IAT in Experiment 1, except for two modifications that are characteristic of the personalized IAT: Firstly, the attribute category labels were the Dutch words for “I like” instead of “Positive” and “I don’t like” instead of “Negative”. Secondly, there was no error feedback for the attribute stimuli. (see Materials section).

In Experiment 3, we applied an affective priming procedure. Each trial started with a 500- ms presentation of a fixation star in the center of the screen. Next, one of the cues was presented as prime for 200 ms, immediately followed by the target (stimulus onset asynchrony of 200 ms), both in the center of the screen. The task was to categorize the targets as quickly as possible as positive or negative by pressing one of two keys. Once a response was made, the next trial started after an intertrial interval of 1000 ms. After 16 practice trials, participants completed a block of 128 test trials in which combinations of the primes (2) and the positive (4) or negative (4) targets were realized equally often, and presented in random order.

For exploratory reasons, after performing the implicit measure, participants were also asked to rate, on a 9-point Likert scale, the pleasantness of cues A, C, X, Y, XF, and YH. The ratings for the cues A and C (presented in a random order) were given first, followed by the ratings for the cues X and Y (presented in a random order), and finally the ratings for the compounds XF and YH (also presented in a random order). Subsequently, participants were asked to judge for cues A and C first (presented in a random order), and then for cues X and Y (also presented in a random order) to what extent they were likely to cause “win” or “loss” in a potential new slot-machine game. This question served as a measure of the causal relations learning between stimuli and outcomes. After the causal ratings, participants were asked to report whether they had discovered the context rule and, if so, to write it down. Finally, but only in Experiments 2 and 3, participants received a list of all stimuli (single and compounds) they experienced in the learning phase. They were asked to indicate, for each of the presented cues (A, B, C, D) and combinations of cues (AB, CD, XF, YH), the outcome it was paired with. This series of questions served as a measure of contingency awareness. After this last questionnaire was completed, participants were thanked and debriefed.

Results

Experiment 1.

IAT. IAT scores were calculated using the D600 scoring algorithm (Greenwald, Nosek and Banaji, 2003). Trials in Blocks 3 and 6 of the IAT were entered as mixed practice blocks whereas trials in Blocks 4 and 7 were treated as mixed test blocks. A positive value for the D600 indicates better performance in the X-positive / Y-negative IAT block than in the X-negative / Y-positive IAT block. In other words, a positive value represents implicit evaluations in line with the target pairings (X paired with win and Y with loss). Although it is well known that caution is required when interpreting the absolute value of IAT scores (e.g., Blanton & Jaccard, 2006), we believe that a positive value of our IAT scores can be interpreted as reflecting a preference for X over Y. Unlike the case in most IAT studies, the target stimuli that we used were nonwords that were unknown to the participants before the experiment. Moreover, a nonword functioned equally often as cue X and cue Y. Hence, the direction of the IAT score (positive or negative) can reflect only the acquired relative preference of X compared to Y.

Most importantly, the difference between the two conditions was significant, $t(54) = 2.21$, $p = .031$, $d = 0.58$. In condition Same, the D600 measure was positive ($M = 0.23$, $SE = 0.11$) and significantly different from zero, $t(27) = 2.09$, $p = .046$, $d = 0.40$. In condition Opposite, the D600 measure was negative ($M = - 0.10$, $SE = 0.10$) but did not significantly differ from zero, $t(27) = -.98$, *ns*.

Explicit valence ratings. Our experiments do not allow for strong conclusions regarding EC of explicit evaluations because explicit evaluations were always assessed after the implicit measure. Nevertheless, for exploratory reasons, we did conduct Cue x Condition ANOVAs on the explicit valence ratings for the target cues X and Y. For this experiment and all subsequent experiments, the relevant mean explicit valence ratings can be found in Table 1. The analysis revealed a significant main effect of cue, $F(1, 54) = 5.40$, $p = .024$, but not of condition, $F(1, 54) = 1.77$, $p =.19$. Furthermore, the interaction between cue and condition was significant, $F(1, 54) = 10.96$, $p < .01$, *ns,* partial $\eta^2=.17$. A post-hoc t-test showed a significant effect of the cue in condition Same, $t(27) = 4.247$, $p < .01$, $d = 0.80$, but not in condition Opposite, $t(27) = -0.66$, $p = .51$. This pattern of results reflects that of the IAT data. The Pearson correlation between implicit and explicit ratings (for the cues X and Y) was also significant, $r = .52$, $p < .01$.

Table 1. Means (and standard errors) of the explicit valence ratings for the target cues in Experiments 1 – 3. X was always paired with a positive outcome whereas Y was always paired with a negative outcome.

Experiments	Cue	Condition Same	Condition Opposite
Experiment 1	X	6.68 (0.48)	5.07 (0.48)
	Y	3.21 (0.50)	5.68 (0.50)
Experiment 2	X	6.10 (0.51)	5.60 (0.59)
	Y	3.90 (0.49)	5.60 (0.56)
Experiment 3	X	5.86 (0.40)	4.95 (0.43)
	Y	4.00 (0.40)	4.61 (0.43)

Causal ratings. Cue x Condition ANOVAs were also conducted on the causal ratings of the target cues X and Y (see Table 2 for all relevant means for Experiments 1-3). The analysis revealed a significant main effect of the cue, $F(1, 54) = 16.50$, $p < .01$, partial $\eta^2=.23$, but no effect of the condition, $F(1, 54) < 1$. We also found an interaction, $F(1, 54) = 38.84$, $p < .01$, partial $\eta^2=.42$. A post-hoc t-test showed a significant effect of the cue in condition Same, $t(27) = 10.80$, $p < .01$, $d = 2.04$, but not in condition Opposite, $t(27) = -1.23$, $p = .23$. Hence, in condition Same, participants indicated that, in a potential new game, the cues X and Y would be followed by the outcome predicted by the context rule, whereas in condition Opposite this effect was not present.

Table 2. Means (and standard errors) of the causal ratings for the target cues in Experiment 1 – 3

A high score indicates that the cue is deemed to cause a win in a potential new game. X was always paired with a positive outcome whereas Y was always paired with a negative outcome.

Experiments	Cue	Condition Same	Condition Opposite
Experiment 1	X	7.82 (0.54)	4.64 (0.54)
	Y	1.89 (0.50)	5.89 (0.50)
Experiment 2	X	7.30 (0.60)	3.00 (0.69)
	Y	2.05 (0.54)	5.40 (0.63)
Experiment 3	X	6.86 (0.59)	3.67 (0.64)
	Y	2.33 (0.50)	3.39 (0.54)

Experiment 2.

Personalized IAT. As in Experiment 1, we used the D600 scoring algorithm (Greenwald et al., 2003). Data of the participants who, when asked at the end of the experiment, did not remember correctly the pairings of the test compounds XF and YH (n = 11), were excluded from the analyses. We did so because, according to the assumptions of propositional models, awareness of the contingencies is a crucial factor in the process of proposition formation. Furthermore, people who did not remember the target pairings may have had a distorted memory and might actually have formed a proposition different from what was implied in the learning phase. Within this sample (n = 35), the difference between the two conditions was significant, $t(33) = 2.10$, $p = .044$, d = 0.68. In condition Same, the D600 measure was positive (M = 0.23, SE = 0.08) and significantly different from zero, $t(19) = 2.66$, p =.015, d = 0.60. In condition Opposite, it was negative (M = - 0.10, SE = 0.14) but not significantly different from zero, *t*(14) = -0.71, *ns*. Note that the difference between conditions was no longer significant when all participants were included in the analysis, *t*(44) = 1.39, *ns*.

Explicit valence ratings. Also for the valence ratings and the causal ratings reported below, only the analyses for the participants who correctly retrieved the target compound co-occurrences will be reported. The analysis for the test cues X and Y did not reveal a significant effect of cue, $F(1, 33) = 2.62$, $p = .12$, partial η^2=.07, nor of condition, $F(1.33) = 3.09$, $p = .09$, partial η^2=.08. The interaction Cue x Condition was not significant, $F(1,33) = 2.62$, $p = .12$, partial η^2=.07. The Pearson correlation between implicit and explicit ratings of the cues X and Y was marginally significant, r = .33, $p = .05$.

Causal ratings. For the test cues X and Y, the ANOVA did reveal a significant main effect of the cue, $F(1, 33) = 4.24$, $p = .047$, partial η^2=.11, but not of condition, $F(1, 33) < 1$. The Cue x Condition interaction was significant, $F(1, 33) = 30.58$, $p <.01$, partial η^2=.48. A post-hoc t-test showed a significant effect of the cue in condition Same, $t(19) = 6.53$, $p < .01$, $d = 1.46$, and a marginal effect in condition Opposite, $t(14) = -2.02$, p = .06. Therefore, in condition Same, participants indicated that, in a potential new game, the cues X and Y would be followed by the outcome predicted by the context rule, whereas in condition Opposite this effect was only marginally significant.

Experiment 3.

Affective priming. Trials in which an incorrect response was given (7. 90% of all trials) were discarded. All latencies were pre-processed by discarding latencies that were outliers in an individual's reaction time distribution according to Tukey's (1977) extreme outlier criterion (e.g., latencies above the third quartile plus 3 times the individual's interquartile range). The data of three participants who made more than 50% errors and seemingly reacted to the prime rather than to the target were excluded from the analyses. In line with the analyses for Experiment 2, we excluded the data of eight participants who did not remember the test cue pairings correctly at the end of the experiment.

In this study, we defined trials in which the target had the same valence as the outcome with which the prime co-occurred in the previous learning phase as congruent trials (e.g., X followed by a positive target and Y followed by a negative target). Trials in which the target had the opposite valence as the outcome with which the prime co-occurred in the previous learning phase are incongruent trials (e.g., X followed by a negative target and Y followed by a positive target). We conducted an ANOVA with congruency (congruent vs. incongruent) as a within subjects factor and context condition as a between subjects factor. The two-way interaction between congruency and condition was marginally significant, $F(1,37) = 3.95$, $p = .05$, partial $\eta^2=.10$. Again, in Condition Same, reaction times for congruent trials ($M = 603$ ms, $SE = 24.2$) were marginally faster than reaction times for incongruent trials ($M = 615$ ms, $SE = 26.7$), $t(20) = 1.938$, $p = .07$, $d = .42$. In Condition Opposite, reaction times for congruent ($M = 559$ ms, $SE = 26.2$) and incongruent ($M = 555$ ms, $SE = 27.5$) trials did not differ $t(17) = -0.81$, *ns*. Note that the Congruence x Condition interaction was not significant when the data of all participants were included in the analyses, $F(1,45) = 1.79$, $p = .19$, partial $\eta^2=.04$.

Explicit ratings. Also for the valence ratings and the causal ratings reported below, only the analyses for the participants who showed a correct retrieval of the target co-occurrences will be reported. The analysis for the test cues X and Y revealed a significant effect of cue, $F(1, 37) = 5.80$, p = .021, but not of condition, $F < 1$. We did not find an interaction $F(1, 37) = 2.81$, $p =. 10$, partial $\eta^2=.071$. The Pearson correlation between implicit and explicit ratings of the cues X and Y was not significantly different from zero, r = .15, *ns*.

Causal ratings. For the test cues X and Y the ANOVA did reveal a main effect of the cue, $F(1, 37) = 17.49$, $p < .01$, partial $\eta^2=.32$, and a marginal effect of condition, $F(1, 37) = 3.50$, $p = .07$, partial $\eta^2=.09$. We found a significant Cue x Condition interaction, $F(1, 37) = 13.67$, $p < .01$, partial $\eta^2=.27$. A post-hoc t-test showed a significant effect of the cue in condition Same, $t(20) = 5.78$, $p < .01$, $d = 1.26$, but not in condition Opposite, $t(17) < 1$. Therefore, in condition Same, participants indicated that, in a potential new game, the cues X and Y would be followed by the outcome predicted by the context rule, whereas in condition Opposite this effect was not present.

General Discussion

The main aim of our research was to test the influence of context variables on EC of implicit evaluations. In three experiments, we compared the implicit evaluations of two target stimuli (X and Y) in two conditions that differed only with regard to the context in which the stimuli were presented. In both conditions, the test stimuli X and Y were consistently presented in compound with another neutral cue (F or H) and consistently followed by the same, positive or negative outcome (XF+, YH-). The context was manipulated by presenting the target cues intermixed with context cues that were paired with outcomes according to a certain rule. In condition Same, each of the context cues was consistently paired with the same outcome, regardless of whether it was presented alone or in compound with another cue (A+, B+, AB+; C-, D-, CD-). In condition Opposite, each of a series of other context cues was paired with a certain outcome when presented alone and with the opposite outcome when presented in a compound with another cue (A+, B+, AB-; C-, D-, CD+). Given that in both conditions the target stimulus X always co-occurred with a positive outcome and the target stimulus Y always co-occurred with a negative outcome, purely associative models of implicit attitudes that take into account only the pairings would predict no difference between the conditions. Results showed, however, that the two conditions differed significantly in terms of implicit evaluations towards the target cues. Across all three experiments, various implicit measures always reflected a preference for the positively paired cue X in condition Same. This preference was always absent in condition Opposite. The current results thus demonstrate that EC of implicit evaluations can be moderated by rules that are implied by the context.

We did not, however, find a reversed EC effect in the condition Opposite. The latter finding stands in contrast to recent results by Peters and Gawronski (2011, Experiments 1 & 2) who found reversed EC of implicit evaluation when participants were verbally instructed after CS-US pairings to mentally reverse the valence of the US. More specifically, the authors provided participants with information about the personality traits of four fictitious people. Persons A and B were paired in 75 % of the cases with positive USs whereas persons C and D were paired in 75 % of the cases with negative USs. The participants' task was, for each pairing, to guess whether the information provided about each of the four individuals was true or false. Information provided by pairings involving persons A and C was said to be true for the frequent valence and false for the infrequent valence, whereas the information provided by pairings involving persons B and D was said to be false for the frequent valence and true for the infrequent valence. Whereas A was preferred over C (i.e., standard EC effect), B was liked less than D (i.e., reversed EC effect). A possible explanation for why we did not find such a reversed EC effect might be that we did not directly instruct participants about how to interpret the CS-US pairings. Instead, our manipulation was more indirect and could have succeeded only if (a) participants noticed the context rule that governed the pairings involving the context cues and (b) they actually applied those rules to make inferences about the USs that were paired with the target cues X and Y when these cues were presented on their own. Although we took measures to ensure that participants did encode the context rule, it might be that some participants did not actually use the rule to make inferences about the target cues X and Y. From this perspective, it is interesting to note that in a related study that we conducted, but with a simplified rule and the explicit task to state conclusions about the target stimuli (please see Chapter 2, Experiment 1 for details), we did find significant reversed EC. This study was more similar to the studies of Peters and Gawronski in that it was much clearer to participants how the CS-US pairings had to be interpreted.

It is also interesting to note that Peters and Gawronski did not find a full reversal when participants were only informed about the correct interpretation of the CS-US after all the pairings had been presented, instead of after each trial. It is therefore possible that the presence of the relational information at the time of encoding represents a crucial factor in obtaining a reversed EC effect. In our studies, particularly in condition Opposite, different participants might have discovered the context rule at different moments during the learning phase. Consequently, the relational information embedded in the context may have been available at the time of the encoding of the CS-US pairings for only some of the participants. This could explain the lack of a reverse EC effect in our studies and its presence in studies in which the encoding of both valenced and relational information is ensured to happen at the same time. Again, note that we did find a reversed EC effect in a study that resembled the Peters and Gawronski studies (see Chapter 2, Experiment 1). Because a simpler rule was used in that experiment than in Experiments 1-3, participants might have become aware of the rule sooner and have already used this information during the first presentation of the target cues X and Y.

Regardless of the exact reasons for why we did not observe a reversal of the EC effect in Experiments 1-3, the fact that we did find a modulation of EC of implicit evaluations has important theoretical implications. First, as we explained in the introduction, purely associative models postulate that the experience of pairings should unequivocally lead to the formation of associations between the mental representations of those stimuli. Because implicit evaluation is assumed to be determined only by associations in memory, one should observe an EC effect that is determined by the experienced pairings in a way that is independent of what the context implies about the meaning of those pairings. The fact that EC was significantly moderated by the properties of the context in which these pairings were experienced does not fit well with these ideas.

According to propositional models, on the other hand, the effect of CS-US pairings on liking is mediated by the formation of a proposition about the CS-US relation. As the content of this proposition depends not only on the CS-US pairings but also on the context in which these pairings occur, aspects of the context should be able to moderate the effect of stimulus pairings on liking. Hence, we predicted EC effects on implicit measures that are moderated by the context rule that is implemented in the pairings of other stimuli. The most straightforward prediction from these models would have been that participants inferred a reversed relation between CS and US in condition Opposite, which according to a purely propositional model, should have led to a significantly reversed EC effect. The fact that we did not find significant reversal speaks against this interpretation. However, propositional models can account for the lack of a reversed EC effect in condition Opposite if certain additional assumptions are made. As we explained in the introduction, one could assume that participants formed two propositions that are relevant for the evaluation of the target stimuli: one determined by the context and one determined by the co-occurrences. Both propositions could cancel each other out, resulting in a null implicit evaluation. In sum, the current results do not refute a strict propositional model but additional assumptions are needed before such a model can explain the exact pattern of results.

Importantly, the current results can also be explained by a hybrid associative-propositional account of EC of implicit evaluations that assumes that both associative and propositional processes contribute to such EC effects. As we explained in the introduction, associative and propositional processes might cancel each other out in condition Opposite, which would explain the null-finding in this condition. On the other hand, there might be circumstances in which propositional processes dominate implicit evaluations, for instance, when unambiguous information is presented about how CS-US pairings should be interpreted. This would explain why Peters and Gawronski (2011) did find a reversed EC effect when participants were instructed to mentally reverse the meaning of CS-US pairings. A hybrid account also fits well with the observation of Peters and Gawronski that verbal information has a stronger effect when presented immediately after a CS-US pairing. In such cases, propositional inferences could overrule the formation of associations based on CS-US pairings and thus dominate implicit evaluations (see Peters & Gawronski, 2011, for details). Note that these ideas are in line with recent dual process models such as the Associative and Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011). Regardless of whether one favors a strictly propositional account or a hybrid account of our results, it is important to note that our results confirm the idea that propositional processes do (co-)determine EC of implicit evaluations (also see Peters & Gawronski, 2011).

To conclude, our results show that EC of implicit evaluations depends not only on the objective CS-US pairings but also on elements in the context that specify the implications of these CS-US pairings. Although we are not the first to observe such context effects (i.e., Peters & Gawronski, 2011; see Förderer & Unkelbach, 2011, for a study on EC of explicit evaluations), we are the first to show that these context effects are not restricted to explicit verbal instructions about how CS-US pairings should be interpreted. Our results suggest that context effects in EC of implicit evaluations might be widespread and confirm the idea that propositional processes do contribute to EC of implicit evaluations.

References

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

Bar-Anan, Y., De Houwer, J., & Nosek, B. A. (2010). Evaluative conditioning and conscious knowledge of contingencies: A correlational investigation with large samples. *Quarterly Journal of Experimental Psychology, 63,* 2313-2335. doi: 10.1080/17470211003802442

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. American Psychologist, 61, 27–41. doi: 10.1037/0003-066X.61.1.27

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

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230-241. http://www.ucm.es/info/psi/docs/journal/

De Houwer, J. (2009a). How do people evaluate objects? A brief review. *Social and Personality Psychology Compass, 3*, 36-48. doi: 10.1111/j.1751-9004.2008.00162.x

De Houwer, J. (2009b). Conditioning as a source of liking: There is nothing simple about it. In Wänke, M. (Ed.). *Frontiers of Social Psychology: The Social Psychology of Consumer Behavior*. New York: Psychology Press.

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

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6*, 202-209. doi:10.1177/1745691611400238

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

De Houwer, J. & Vandorpe, S. (2010). Using the Implicit Association Test as a measure of causal learning does not eliminate effects of rule learning. *Experimental Psychology*, *57*, 2010, 61-67. doi: 10.1027/1618-3169/a000008

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

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition: Their meaning and use. *Annual Review of Psychology*, 54, 297-327. doi: 10.1146/annurev.psych.54.101601.145225

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher-order encoding processes. *Cognition and Emotion*. doi: 10.1080/02699931.2010.513497

Förderer, S. & Unkelbach, C. (2011). Hating the cute kitten or loving the aggressive pit-bull: EC effects depend on CS – US relations. Cognition and Emotion. doi: 10.1177/1948550611398413

Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli. *Quarterly Journal Of Experimental Psychology*, *65*(9), 1657–1674. doi: 10.1080/17470218.2012.665061

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

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

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216. doi: 10.1037/0022-3514.85.2.197

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

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorising in implicit attitude research: Propositional and behavioral alternatives. *Psychological Record, 61*, 465–496. http://thepsychologicalrecord.siuc.edu/

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

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

Olson, M., & Fazio, R. H. (2004). Reducing the Influence of Extrapersonal Associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality & Social Psychology, 86(5),* 653-667, doi: 10.1037/0022-3514.86.5.653

Payne, B. K., Cheng, C. M., Govorun, O., Stewart, B. D. (2005). An Inkblot for Attitudes: Affect Misattribution as Implicit Measurement. *Journal of Personality and Social Psychology*, 89, 277-293. doi: 10.1037/0022-3514.89.3.277

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

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

Roefs, A., Huijding, J., Smulders, F. T. Y., MacLeod, C. M., De Jong, P. J., Wiers, R. W., & Jansen, A. T. M. (2011). Implicit measures of association in psychopathology research. *Psychological Bulletin, 137*, 149-193. doi: 10.1037/a0021729

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

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

Tukey, J. W. (1977) *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

Wiers, R. W., & Stacey, A. W. (2006). Implicit cognition and addiction. *Current Directions in Psychological Science, 15*, 292-296. doi: 10.1111/j.1467-8721.2006.00455.x

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151-175. doi: 10.1037/0003-066X.35.2.151

Footnotes

[1] Note that we use the term implicit evaluation to refer to a behavioral phenomenon (i.e., the fact that a stimulus evokes an evaluative response automatically) rather than to a type of mental representation (e.g., implicit attitudes; see De Houwer, 2009a, 2011). Although the study of implicit evaluations imposes limits on theories about the mental representations that mediate implicit evaluation, it should not rely on a priori assumptions about the nature of those representations (e.g., the assumption that implicit and explicit attitudes are separate entities).

[2] Associative learning is defined as changes in behavior that are due to relations between events (e.g., De Houwer, 2009c).

CHAPTER 2

ON THE SENSITIVITY TO RELATIONAL INFORMATION OF IMPLICIT MEASURES OF EVALUATION [1]

[1] This study was conducted in collaboration with Jan De Houwer and Anne Gast

Introduction

During the last decade, the interest of researchers in automatic (implicit) evaluation has considerably risen. In order to further knowledge on implicit evaluation, it is important to investigate how it can be measured. If evaluation is defined as an effect, more specifically as the effect of stimuli on evaluative responses (De Houwer, Gawronski, & Barnes-Holmes, 2012), implicit evaluation is defined as a subset of evaluation, namely as evaluation that occurs under certain instances of automaticity (De Houwer et al., 2012; De Houwer & Moors, 2012). The increase of interest in implicit evaluation has been made possible by the development of measurement techniques (see Gawronski & De Houwer, in press, for an extensive review). Most of those are based on the accurate recording of responses and reaction times. These measures create situations in which people respond to a number of stimuli. These situations differ across implicit measures with regard to the nature of the responses that are required, as well as with regard to the conditions of automaticity that they imply.

In order to examine how evaluations can be measured, it is a useful strategy to work with novel stimuli whose learning history can be controlled fully. In this way, one can experimentally manipulate the properties of the stimuli and thus make clear predictions about the way in which these stimuli should be evaluated (see De Houwer, Teige-Mocigemba, Spruyt & Moors, 2009, for a detailed discussion). A widely explored way in which evaluation is acquired is through stimulus pairings. This effect, called evaluative conditioning (EC; De Houwer, 2007; De Houwer, Thomas, & Baeyens, 2001; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010), refers to the change in liking of a stimulus (named conditioned stimulus, CS) that is due to its pairing with another stimulus (named unconditioned stimulus, US). Recently, more attention has been given to research on the possible moderation of the relation occurring between CS and US on EC effects (Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2011; Moran & Bar-Anan, 2012; Peters & Gawronski, 2011; Unkelbach, Förderer, & Stahl, 2012; Zanon, De Houwer & Gast, 2012). We use the term "relational information" to refer to the content of this relation. For example, a certain unknown substance X may co-occur with a certain object (for example a disease), but could be related to it in different ways. To name a few, it could cause it, prevent it or be an effect of it (see Lagnado, Waldmann, Hagmayer, & Sloman, 2007, for an insightful discussion). Obviously, the way in which the unknown substance X is related to the disease is a very important piece of information that can have important implications for how one evaluates the substance. As mentioned above, research has recently started to investigate the impact of relational information on evaluation (e.g., Peters & Gawronski, 2011). In this research, several explicit and implicit measures of evaluation have been used. However, it is still not completely clear whether and when relational information can moderate implicit instances of evaluation, and, in particular, which specific implicit measures are sensitive to relational information. In our studies, we are mostly interested in relational information that implies a relation of opposition between a CS and an US. This type of relation is particularly interesting, as different cognitive accounts of EC (i.e., association formation models and propositional models) make different predictions about its impact on evaluation (see the Introduction for a detailed review of cognitive models of evaluation).

Consequently, the present series of experiments aimed at shedding light on the question of whether relational information can moderate implicit evaluation and, more specifically, whether different instances of implicit evaluation (i.e., evaluations that are captured by different implicit measures) are differentially sensitive to relational information. To this end, we examined a series of widely used implicit measures of evaluation and tested their sensitivity to relational information. In this paper, we focus on the implicit association test (IAT; Greenwald, McGhee, and Schwartz, 1998), the affective misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005), and evaluative priming with a naming task (cf. Fazio, Sanbonmatsu, Powell & Kardes, 1986; Spruyt, Hermans, De Houwer, Vandekerckhove, Eelen, 2007). The following paragraphs provide a brief description of the nature of and knowledge about these three implicit measures.

The IAT (Greenwald et al., 1998) is definitely the most widely used implicit measure in psychological research. In a typical IAT, participants are asked to quickly categorize stimuli of different categories (e.g., pictures of White and Black persons intermixed with positive or negative words) by pressing one of two keys. In the relevant blocks, each of the two keys is assigned to two categories at the same time (e.g., right key for White and positive; left key for Black and negative). Stimuli belonging to all categories are presented one by one and have to be categorized as fast as possible by pressing one of the two keys. The crucial measurement is a difference in performance (as calculated via an algorithm based on reaction times and errors; see Greenwald, Nosek and Banaji, 2003) between blocks that differ in the way that categories are assigned to responses (e.g., press the right key for Blacks and positive and the press left key for Whites and negative versus press the right key for Whites and positive and press the left key for Blacks and negative). The idea behind this implicit measure is that participants perform better when associated categories are assigned to the same response than when related categories are assigned to different responses. Hence, the difference in performance is thought to provide an index of how the extent to which categories are associated.

The AMP (Payne et al., 2005) is another widely used implicit measure. A trial of this procedure typically consists of the presentation of a (neutral) Chinese ideograph, preceded by the brief presentation of a prime picture and followed by a mask. When used as an implicit measure of evaluation, the primes are stimuli belonging to the to-be evaluated categories. For instance, when the aim is to capture the evaluation of flowers and spiders, pictures of flowers and spiders can be used as primes. The participant has the task to evaluate each of the ideographs as being more or less pleasant than average, by pressing one of two keys. The crucial measurement here is the number of “pleasant” responses that are emitted as a function of the nature of the prime (e.g., the number of “pleasant” responses when the prime depicts a flower versus when the prime depicts a spider). According to the authors, participants who perform this task misattribute the valence of the prime to the target, producing an AMP effect. For example, Chinese ideographs that are preceded by a positive stimulus (e.g., the picture of a flower) would be, on average, judged more positively than those that are preceded by a negative stimulus (e.g., the picture of a spider). Although AMP effects are not based on reaction times, they could be interpreted as implicit evaluation, as the impact of the primes on the evaluation of the targets is considered to be automatic in the sense of uncontrolled. Payne and colleagues (2005) have shown that AMP effects emerge even when participants are instructed not to be influenced by the prime, and even when they are informed about the possible effect of the primes on their responses. However, Bar-Anan and Nosek (2012) reported a series of studies in which they found an AMP effect only for those participants who deliberately rated the primes instead of the targets. Indeed, there is still an open debate among researchers about whether these effects are independent from participant's control on their responses.

The third implicit measure we will focus on is an pictorial evaluative priming procedure with naming responses (see Spruyt et al., 2007), which is a variation of the standard evaluative priming paradigm (Fazio et al., 1986). In this task, participants are presented with a series of trials, each consisting of a brief presentation of a positive, negative, or neutral picture prime, followed by a positive or negative target picture. Participants have to name the object depicted in target picture, as quickly as possible, while ignoring the prime. A voice key is used to record the response time, that is, the time interval between the onset of the target picture and the onset of the naming response. For example, if the target is the picture of a dolphin, participants would have to react to its presentation, as quickly as possible, by saying out loud the word "dolphin". A congruent trial would consist, for example, of a picture of a flower as a prime, and the picture of a puppy as a target (congruent positive valence). Conversely, an incongruent trial could, for example, consist of the picture of a skull as a prime and the picture of a puppy as a target (incongruent valence). The evaluative priming effect is calculated by comparing reaction times between congruent and incongruent trials. If this task is used as an implicit measure of evaluation, the to-be evaluated stimuli are used as primes. The characteristic that makes this task particularly intriguing is the fact that, unlike the previously described implicit measures, it does not require an explicit evaluation of stimuli (i.e., a conscious intentional self-assessment of how much one likes or dislikes an object). In fact, the participant's task is simply to verbally name the target pictures they are presented with. There is also evidence that, unlike most other implicit measures, effects in this measure might be mediated by the spreading of activation of CS-US associations (e.g., Spruyt et al., 2007). Given that associations do not contain relational information, effects in this naming task might be particularly insensitive to the impact of relational information.

In the present set of experiments, we aimed at testing the impact of relational information on EC effects of implicit evaluation. In addition, we were interested in testing how sensitive three different implicit measures of evaluation were to relational information. In Experiments 1, 2 and 3, we adopted a learning procedure inspired by De Houwer and Vandorpe (2010), similar to the one described in Chapter 1 (Zanon et al., 2012). In the experiments presented in Chapter 1, we found an attenuation of an EC effect of implicit evaluation due to information about a relation of opposition, information that was implied by a context of stimulus pairings. However, in the crucial Opposite condition, we did not find a significant reverse EC effect, that is, an implicit preference for the CSneg over the CSpos. In their seminal article, Peters and Gawronski (2011 – Experiment 3) reported evidence that this reverse EC effect could be obtained under certain conditions. They presented a series of CS-US pairs, together with opposition and equivalence relational information, either presented after each pairing (short-delay condition) or at the end of the whole series of pairings (long-delay condition). In their short-delay condition, they obtained a reverse EC effect, whereas in their long-delay condition they obtained only an attenuation of the EC effect. In light of the implications of these results, with Experiment 1 we aimed at extending the findings of Zanon and colleagues (see Chapter 1), by presenting a similar but simpler design in which CS-US pairings were presented along with other pairings, embedding a rule that implies opposition between CSs and paired USs. Our hypotheses were that, if relational information derived from the rule implied by the context impacted implicit evaluation, our implicit measures would show a reverse EC effect (preference for the CSneg over the CSpos), thus confirming the finding of the short-delay condition of Peters and Gawronski (2011). If relational information did not impact the outcome of the implicit measures at all, we would expect a standard EC effect (preference in line with the valence of the CS-US co-occurrences). Obtaining a reverse EC effect would allow us to conclude that the implicit measure is sensitive to relational information. An attenuation of the effect, that is, no preference for either of the CSs, is also a possible result. On the one hand this could tell us that an implicit measure is sensitive to opposite relational information, because it did not show a standard EC effect. On the other hand, it could be plausible that a certain implicit measure does not capture any evaluation at all or is simply less sensitive to any type of manipulation. In the absence of a control condition in which no relational information is present or the context implies a relation of equivalence, finding a null effect in a context that implies a relation of opposition would not allow us to draw strong conclusions. Nevertheless, because both a significant standard EC effect and a significant reverse EC effect would allow us to draw strong conclusions, in our Experiments 1, 2 and 3 we only presented the crucial condition in which the context rule implied opposite relational information.

In the procedure of Experiments 1, 2 and 3, was similar to that of the experiments reported in Chapter 1. More specifically, context pairings of neutral cues and positive and negative outcome were presented. This series of pairings embedded a relational rule. The rule implied that a special cue reversed the valence of the outcome of the neutral cue it was presented with. For example, if the neutral cue A, presented on its own, was followed by a positive outcome (we will call this pairing Apos), the compound of A and a second cue R would have resulted in a negative outcome (ARneg). In the acquisition phase, Apos, ARneg, Bneg and BRpos trials were presented. These pairings form a context in which the rule is embedded. The target pairings, also presented, were XRpos and YRneg. It is important to note that the neutral cues X and Y (the CSs) were never presented with only the outcome, but always in compound with the special cue

R. In Experiment 1, we measured the evaluation of X and Y with an IAT, whereas in Experiments 2 and 3 we adopted an AMP.

Finally, Experiment 4 was a replication of the long-delay condition of Experiment 3 of Peters and Gawronski (2011) who measured evaluative priming in an evaluative categorization task. They observed that opposition information cancelled out the EC effect that was observed with equivalence information. Hence, with a long-delay, opposition information did have an effect on implicit evaluation, but this effect was not strong enough to reverse the effect. Our aim was to test whether opposition information would produce different results in an evaluative priming task with naming responses. Based on the idea that effects in the naming task are mediated by spreading of activation across associations (see Spruyt et al., 2007), we hypothesized that effects in the naming task would be insensitive to relational information and would thus reveal a standard EC effect despite the opposition instructions. In line with Peters and Gawronski (2011, Experiment 3), we presented the pictures of four individuals A, B, C, and D and paired them with either positive or negative behavioral statements (Apos, Bpos, Cneg, and Dneg). Participants were asked to assume that the person in the picture did hold those features. In one of the conditions (long-delay condition), participants were later informed that two of the statements were true, whereas the other two were false and had to be reversed (Atrue, Bfalse, Ctrue, Dfalse). Therefore, this manipulation produced a fully crossed design of contingency (positive vs. negative) and validity of information. In line with our previous experiments (see Chapter 2), we refer to this latter factor as the type of relation (equivalence vs. opposition). The crossing of these factors results in the four CS conditions: Apos-equivalent; Bpos-opposite; Cneg-equivalent and Dneg-opposite. Implicit evaluation of the four CSs was then measured, employing an evaluative priming procedure with a naming task.

Experiment 1

In Experiment 1, we employed a procedure similar to that used in condition Opposite of Experiments 1, 2, and 3 of Chapter 1. In this procedure, single cues and compounds of cues were followed by positive and negative outcomes (wins and losses in a slot machine game). The series of cue-outcome pairings in the experiments of Chapter 1 were regulated by rule, namely that the outcome of a single cue was always the opposite of the outcome of that cue when presented in compound with another cue. The first important difference with the studies reported in Chapter 1 was that we now used a simplified context rule. In Experiment 1, stimuli were either shown alone or in a compound with the cue R. Participants were told that the cue R reversed the outcome of the cue with which it was paired. This rule is less complex because the reversal cue R is the same across all compounds and because a smaller total number of cues is needed to implement this rule. We therefore hoped that the use of the reversal cue would encourage participants infer on the basis of XRwin and YRloss trials that Y(CSneg) on its own leads to win and X(CSpos) on its own leads to loss. A second important difference was that participants were asked to make causal judgments about X and Y before administrating the IAT. We implemented this change to give participants the opportunity to relate the context rule to the target cues X and Y and thus to form the propositions about the target cues that were implied by this information. A third difference to the experiments in Chapter 1 is that we only realized condition Opposite. We expected that the above mentioned changes to the design would make it more likely that people form propositions about the reversal rules and therefore expected that we would find a significantly reversed EC effect for stimuli sown in compound with a reversal cue.

Method

Participants. Eighty-one students at Ghent University participated (mean age = 19.03, *SD* = 1.68; 74% were women). They either received course credits or four Euros.

Materials. ***Learning phase.*** During the learning phase, the nonsense words “UYATA”, “SOKOG”, “QUQQ” (context stimuli), “LOKANTA”, and “FEVKANI”(target CSs) were presented in black on a white background, in four different combinations of font styles (Arial Black or Fixedsys) and font sizes (ranging from 30 point to 40 point). The positive US was the Dutch word for "a win" ("WINST"), presented in green on a black background, font style Britannic Bold, font size 78pt, surrounded by a number of color pictures of 2 euro coins, and accompanied by a soft melody. The negative US was the Dutch word for "a loss" ("VERLIES"), presented in red on a black background, font style Haettenschweiler, font size 100, surrounded by two cartoon faces expressing sadness, and a loud buzzer sound. CSs and USs were always presented in a top and bottom display of a slot machine frame, respectively. We refer to the CS that co-occurred with the positive US as CSpos and to the CS that co-occurred with the negative US as CSneg. For half of the participants, the nonword "LOKANTA" was the CSpos, and the nonword "FEVKANI" was the CSneg. This assignment was reversed for the other half of the participants.

IAT. The target stimuli used in the IAT were the CSs "LOKANTA" and "FEVKANI". Each target stimulus was presented in four different fonts (lower case Arial Black, upper case Arial Black, lower case Fixedsys, and upper case Fixedsys), resulting in eight different target stimuli, which were identical to those presented in the acquisition phase. We varied the font of the target words in order to avoid that stimuli would be classified only on the basis of simple perceptual features. De Houwer (2006) showed the reliability of this procedural modification and its suitability to capture recent learning. In addition, in order to resemble as much as possible the learning phase, each target stimulus was presented in the top display of a slot machine frame. The positive attributes were the Dutch words for "happy" ("GELUKKIG"), "honest" ("EERLIJK"), "pleasant" ("PRETTIG") and "sincere" ("OPRECHT"). The negative attributes were the Dutch words for "mean" ("GEMEEN"), "rude" ("BRUTAAL"), "aggressive" ("AGRESSIEF") and "deceptive" ("BEDRIEGLIJK"). Both positive and negative attributes were presented in the middle of the screen, in black color on a white background, font style Arial Black, font size 48 point. The error signal was a red cross, which also appeared in the middle of the screen.

Causal ratings. Participants were first asked to rate for the cues "UYATA" and "SOKOG" (presented in random order) and then for the cues "LOKANTA" and "FEVKANI" (also presented in random order) to what extent they were likely to cause either "win" or "loss" in a potential new slot-machine game, by indicating a number on a scale ranging from 1 (loss) to 9 (win).

Explicit valence ratings. Participants were asked to rate, on a 9-point Likert scale, how pleasant they found the cues "UYATA", "SOKOG", "LOKANTA", "FEVKANI" and the compounds "LOKANTA + QUQQ" and "FEVKANI + QUQQ". Participants could respond by selecting a value ranging from 1 (very unpleasant) to 9 (very pleasant).

Procedure. At the beginning of the experiment, participants were explicitly told that they were about to see a series of slot machine games. Each game would consist either of a single cue followed by an outcome, or of two cues followed by an outcome. The outcome could be a win or a loss at the slot machine. The initial instructions also explained that when two cues were shown, one of them was always a particular reverse-cue that had the function of reversing the outcome of the other cue that was presented at the same time on the screen. Participants were asked to pay attention during the presentation of the slot machine games, as questions about them were going to be asked afterwards.

After the instruction page, participants could start the task by pressing a key. At the beginning of each trial, the neutral cue(s) appeared on a slot machine frame. After 2000 ms, the outcome appeared on a second display of the slot machine frame and both cue(s) and outcomes remained on the screen for 3000 ms. Afterwards, the screen turned blank for a 3000 ms intertrial interval. The sequence of trials during the learning task was the following. First, Apos (single cue A followed by a positive outcome) and ARneg (nonword A and reverse cue R, followed by a negative outcome) trials were presented twice each (Block 1). This block of four trials was followed by Bneg and BRpos trials (Block 2), each presented twice. Finally, XRpos and YRneg were presented twice each (Block 3). Whereas A, B, and R were the context cues, X and Y were the CSpos and CSneg, respectively. This sequence of three blocks was presented five times. In total, the learning phase consisted of sixty trials. The trial order within the blocks was determined randomly for each participant and block separately.

After the presentation of all learning trials, participants were asked to judge first for the cues A and B (presented in random order) and then for the cues X and Y (also presented in random order) to what extent they were likely to cause “win” or “loss” in a potential new slot-machine game. After these causal ratings, participants performed an IAT. After performing the IAT, participants were asked to rate the pleasantness of cues A, B, X, Y, and of the compounds XR and YR on a 9-point Likert scale. The ratings for the cues A and B (random order) was asked first, followed by the ratings for the cues X and Y (random order), and the ratings for the compounds XR and YR (random order). As our focus is on implicit measures of evaluation, explicit ratings were only assessed for exploratory purposes and always after the implicit measure. After completing the experiment, participants were thanked and debriefed.

Results

IAT. Scores were calculated using the *D4* algorithm (Greenwald, Nosek and Banaji, 2003). High scores indicate an implicit preference in line with the valence of the co-occurrences (CSpos liked more than the CSneg). As predicted, the D600 measure for the single cues X and Y was negative ($M = -0.15$, $SE = 0.49$) and significantly different from zero, $t(80) = -2.98$, $p < .01$, $d = 0.33$. This indicates that the cue that had co-occurred with the negative outcome was liked more than the cue that had co-occurred with the positive outcome.

Explicit evaluation. Scores of the explicit valence ratings were coded in a way that high scores reflect a preference for the CSpos over the CSneg. The explicit score was negative, and significantly different from zero, ($M = -1.38$, $SE = 0.49$), $t(80) = -2.85$, $p < .01$, $d = 0.45$. Like in the IAT, the CSneg was evaluated more positively than the CSpos.

Causal ratings. The causal ratings were coded in a way that high scores indicated a causal judgment based only on mere co-occurrences (i.e., what co-occurred with “win” was more likely to cause “win”). The causal rating score was negative and significantly different from zero, (M = −4.68, SE = 0.55), $t(80)$ = −8.51, $p < .001$, d = 0.95, reflecting the implications of the rule implied by the context pairings. In other words, participants indicated that the CS that was in the losing (winning) compound would have been, when on its own, more likely to cause win (loss) in a potential new game.

The IAT score was positively correlated with the explicit valence ratings (r = .53, $p < .001$), but not with the causal ratings (r = .12, p = .57). Explicit valence ratings and causal ratings were not correlated either (r = .20, p = .07).

Discussion

In this experiment, two CSs were repeatedly shown together with a positive or negative USs and an additional cue. Participants were instructed that this additional cue indicated that the US that would follow a particular stimulus would be reversed if this cue was present. This rule was also implemented in the presentation of context stimuli A and B. The additional cue thus indicated that a CS that was followed by a certain outcome would be followed by the opposite outcome when presented alone. The CSs, however, where never shown alone. The results of an IAT showed that participants formed an implicit preference for the CS that co-occurred with the negative US (and the reversal cue) over the CS that co-occurred with the positive US and the reversal cue. The implicit preference as shown in the IAT, are thus in line with the context rule, that is, opposite to the valence implied by the mere CS-US co-occurrences. The same pattern was found for explicit evaluation. This result confirms the findings of Peters and Gawronski (2011) and generalizes them to different stimulus types.

Experiment 1 thus showed that the measurement outcome of an IAT, as an instance of implicit evaluation, is sensitive to relational information. In the following experiments, we tested whether other implicit valence measures show the same sensitivity to relational information.

Experiment 2

In this experiment, we tested whether the AMP (Payne et al., 2005) is sensitive to relational information that was presented in a learning phase. We employed the same induction phase as in Experiment 1 but used the AMP as implicit measure. In the AMP a prime (a positive, a negative or a neutral picture) is presented briefly and followed immediately by a neutral Chinese character and then by a mask. Participants have to indicate by pressing one of two keys whether the Chinese ideograph is more or less pleasant than the average ideograph. Payne and colleagues have shown that the evaluation of the Chinese characters is influenced by the valence of the primes. Ideographs that are preceded by positive primes are more likely to be evaluated positively than ideographs that are presented after negative primes (Payne et al., 2005). In our version of the AMP, the primes were the two CSs (i.e., nonwords X and Y). Another nonword that was not presented during the acquisition phase, served as a neutral prime. The measure of implicit preference that is provided by the AMP is based on the ratio of positive responses towards the Chinese ideograph that follow a particular prime. If the AMP is sensitive to relational information, we should find a higher ratio of positive responses to the ideographs preceded by the negatively paired CSs than to the ideographs preceded by the positively paired CSs.

Method

Participants. Sixty-two Ghent University students participated in the experiment, in exchange for course credit or four euro (mean age = 20.27, *SD* = 4.64; 85% were women).

Materials. ***Learning phase.*** The materials used in the learning phase were identical to those used in Experiment 1.

AMP. The stimuli used as primes in the AMP were the CSs, the nonwords “LOKANTA”, “FEVKANI”, and the nonword “ENANWAL” (neutral prime), which were presented in black on a white background, font style Arial Black, font size

48. The target stimuli were 36 Chinese ideographs, which were drawn in black on a white background.

Causal and explicit valence ratings. The materials employed for the causal ratings and explicit valence ratings were identical to those used in Experiment 1.

Procedure. The procedure of the acquisition phase was identical to that of Experiment 1. After the acquisition phase, participants performed the causal ratings as in Experiment 1. Afterwards, participants could proceed to the AMP instructions by pressing a key. Before the main block was presented, participants performed ten practice trials, whose structure was identical to that of the test trials. Then, participants performed the main block, in which the cues X (CSpos), Y (CSneg), and Z (neutral prime) were employed as primes. Each of these primes were used 12 times, adding up to 36 trials, which were presented in random order. Within each trial, the prime was presented for 75ms, followed by a Chinese character for 200ms and a mask. Participants had to indicate, by pressing one of two keys, whether the Chinese ideograph was more or less pleasant than average. They were asked to respond only on the basis of their liking of the Chinese character and thus to ignore the primes. After the AMP, participants performed the explicit valence ratings with the same modalities as in Experiment 1.

Results

AMP. We calculated the percentage of positive responses for each of the three primes (positive, negative, and neutral). High scores thus indicate a preference for the specific prime (Payne et al., 2005). Data of three participants were discarded due to an excessive rate of responses with the same key (>95%). Means and standard errors of the percentages of pleasant responses for each prime can be found in Table 1. Results were analyzed using a one-way ANOVA with three conditions of the factor prime. The main effect of prime was not significant, $F(2,58) = 0.42$, $p = .58$, partial $\eta^2 = 0.01$. In addition, a t-test did not reveal a difference between positive and negative primes, $t(59) = -0.46$, $p = .65$, *d* = 0.06. In sum, our AMP did not show a significant implicit preference for any of the CSs.

Explicit evaluation. As in Experiment 1, **s**cores of the explicit valence ratings were coded in a way that high scores reflect a preference for the CSpos over the CSneg. The explicit score was negative, and marginally different from zero, ($M = -0.97$, $SE = 0.48$), $t(58) = -2.00$, $p = .50$, $d = 0.26$. This result means that the CSneg was evaluated marginally more positively than the CSpos.

Casual ratings. The causal ratings were coded in a way that high scores indicated a causal judgment based only on mere co-occurrences (i.e., what co-occurred with "win" was more likely to cause "win"). The causal rating score was negative and significantly different from zero, ($M = -5.32$, $SE = 0.64$), $t(59) = -8.33$, $p < .001$, $d = 1.08$, meaning that, on average, participants indicated that the CS that was in the losing (winning) compound was deemed, when on its own, more likely to cause win (loss) in a potential new game.

The AMP score was not correlated with the explicit valence ratings ($r = .07$, $p = .61$), neither with the causal ratings ($r = .12$, $p = .36$). In addition, explicit valence ratings and causal ratings were correlated ($r = .40$, $p < .01$).

Table 1. *Means (and standard errors) of the percentage of "pleasant" responses on the AMP in Experiments 2 and 3.*

Prime	Experiment 2	Experiment 3
Positive	53.4 (2.42)	53.5 (2.73)
Negative	55.2 (2.48)	52.3 (3.33)
Neutral	52.1 (2.40)	54.6 (2.85)

Discussion

Experiment 2 was a replication of Experiment 1 with a different type of implicit measure, the AMP. In this experiment, the context rule effect that was found with the IAT could not be replicated with the AMP. The results of the causal ratings suggest that participants understood the context rule, but that the AMP did not pick up its effect. Please note that we also did not find an effect in line with the co-occurrences, which might suggest that the impact of the co-occurrences and the impact of the opposite relational information cancelled each other out. Because we do not have a control condition without opposite relational information that showed a significant effect on the AMP, it is also possible that the AMP was just not sensitive enough for any evaluative information. A possible reason why the AMP produced a null result in this experiment is that we used nonwords as primes, whereas the AMP usually employs pictures as prime stimuli. Therefore, with Experiment 3 we replicated Experiment 2, employing picture stimuli instead of nonwords.

Experiment 3

Experiment 3 was a replication of Experiment 2 with the difference that pictures instead of nonwords were used as CSs and as primes.

Method

Participants. Twenty-nine Ghent University students participated in the experiment, in exchange for course credit or four euro (mean age = 18.55, *SD* = 0.63; 90% were women).

Materials. ***Learning phase.*** The materials used in the learning phase were the cartoon characters (Pokemons) *Diglett, Kakureon, Jigglypuff* and *Ruriri*. They were chosen because they had already successfully been used in previous EC research (Olson & Fazio, 2001). The two characters ("*Jigglypuff*"and" *Ruriri*") which were evaluated most similarly and neutrally in a previous normative procedure (Olson & Fazio, 2001) served as CSs. As the special reverse-cue (R), we used a simple symbol, composed of two half-circular arrows.

The CSpos was the CS consistently paired with the positive US, whereas the CSneg was the CS consistently paired with the negative US. For half of the participants, the picture of "*Jigglypuff*" was the CSpos, and the picture of "*Ruriri*" was the CSneg. This assignment was reversed for the other half of the participants.

AMP. The stimuli used as primes in the AMP were the pictures of the Pokemons "*Jigglypuff*" and "*Ruriri*" (the CSs) and of a grey rectangle (neutral prime). The target stimuli were 36 Chinese ideographs, which were drawn in black on a white background

Causal ratings and explicit evaluative ratings. The materials employed for the causal ratings and explicit evaluative ratings were identical to those adopted in Experiment 1 and 2, except that the CSs were Pokemon pictures instead of nonwords.

Procedure. Except for the use of pictures as CSs, the procedure of the whole experiment was identical to that of Experiment 2.

Results

AMP. We calculated the proportion of positive responses for each of the three primes (positive, negative, and neutral) in the same way as in Experiment 2 (high scores indicate a preference for the specific prime). Results were analyzed using a one-way ANOVA with three conditions of the factor prime. The main effect of prime was not significant, $F(2,28) = 0.20$, $p = .82$, partial $\eta^2 = 0.015$. In addition, a t-test did not reveal a difference between positive and negative primes, $t(29) = 0.27$, $p = .79$, $d = 0.06$. Therefore, like in Experiment 2, our AMP did not show a preference for any of the CSs. Means and standard errors of the proportions of "pleasant" responses for each prime can be found in Table 1.

Explicit evaluation. As in Experiment 1 and 2, **s**cores of the explicit valence ratings were coded in a way that high scores reflect a preference for the CSpos over the CSneg. The explicit score was positive, but not different from zero, ($M = 0.14$, $SE = 0.47$), $t(28) = 0.29$, $p = .77$, $d = 0.05$. This result means that participants did not show an explicit preference for either of the CSs.

Causal ratings. As in Experiment 1 and 2, the causal ratings were coded in a way that high scores indicated a causal judgment based only on mere co-occurrences (i.e., what co-occurred with "win" was more likely to cause "win").

The causal rating score was negative and significantly different from zero, (M = −4.31, SE = 1.08), $t(29) = -3.99$, $p < .001$, d = 0.74. On average, participants indicated that the CS that was in the losing (winning) compound was deemed, when on its own, more likely to cause win (loss) in a potential new game.

The AMP score was not correlated with the explicit valence ratings (r = .16, p = .41), neither with the causal ratings (r = - .06, p = .76). Moreover, explicit valence ratings and causal ratings were not correlated either (r = .12, p = .55).

Discussion

As was the case in Experiment 2, the effect of rule learning on EC was not obtained even though we now used and AMP with pictures (Pokemons). This null effect occurred even though, as in Experiments 1 and 2, the results of the causal ratings showed a general understanding of the rule regulating the CS-US presentations. To conclude, also the AMP with picture stimuli did not seem to be sensitive to our manipulation, that is, both the CS-US co-occurrences and the opposite relational information implied by the context pairings. Again, we cannot be sure whether the observed null effect is due to a problem with our version of the AMP or to a combined influenced of co-occurrences and relational information.

Experiment 4

After testing the effect of opposite relational information on implicit preference as measured with the IAT and with the AMP, with Experiment 4 we aimed at testing the impact of relational information on evaluative priming in the picture-picture naming task. Previous research (Spruyt et al., 2007) suggested that effects in this task is mediated by the spreading of activation across associations, and therefore may be less sensitive to the impact of relational information. The design of the current experiment was inspired by Experiment 3 of Peters & Gawronski (2011). This means that we also realized conditions in which the context did not suggest an opposite relationship but a relation of equivalence. In case of a null finding in the condition with opposite relational information, this design allows us to draw conclusions about whether the implicit measure was just not sensitive to any evaluative information or whether the effect of co-occurrences and relational information might have cancelled each other out. Furthermore replicating the design of Peters and Gawronski with a different type of implicit measure (they used an evaluative priming procedure with an evaluative decision task while we used an evaluative priming procedure with a naming task), allows us to draw conclusions about whether these types of evaluative priming procedures differ in their sensitivity to evaluative information.

Method

Participants. Twenty-seven Ghent University students participated in the experiment, in exchange for course credit or four euro (mean age = 20.55, *SD* = 1.67; 85% were women).

Materials. ***Learning phase.*** All materials were adapted from Peters and Gawronski (2011). In the learning phase, four neutral color pictures depicting the faces and the name of four young men and ten positive and ten negative behavioral descriptions were employed. The four pictures were used as CSs, whereas the positive behavioral descriptions were the USpos, and the negative behavioral descriptions were the USneg.

We refer to the CS that co-occurred with the positive US as CSpos and to the CS that co-occurred with the negative US as CSneg. The assignment of pictures and the valence of the behavioral statements was counterbalanced across participants.

Naming task. The stimuli used as primes in the naming task were the four faces employed in the acquisition phase. In addition, two neutral pictures were also employed as primes. The target stimuli were ten IAPS pictures, five of which were positive (bride, teddy bear, kitten, dolphin, baby) and five were negative (gun, skulls, corpse, worms, explosion).

Explicit valence ratings. Following the procedure of Peters and Gawronski (2011), we assessed explicit evaluation of the four CSs by asking participants to rate them in random order on likeability, friendliness, and trustworthiness. The responses were given on a 7-point Likert scale, whose anchors ranged from 1 (not at all) to 7 (very much).

Procedure. The procedure mimicked that of Peters and Gawronski (2011). Initial instructions introduced the experiment as a study on impression formation. Participants were instructed that they had to imagine to have recently found a new job, and they were about to see pictures of four of their new co-workers, together with second-hand comments that other colleagues had previously made about them. Hence each picture was accompanied with a behavioral description that was instructed to be descriptive of the person in the picture. The instructions also stated that some of the behavioral descriptions may have been false. However, the initial task was to assume that all the descriptions were true, and in a later stage their validity would be communicated. Once they had read the instructions, participants could move to the pairing phase by pressing a key. The pairing phase consisted of twenty trials, each of which started with the presentation of a CS in the center of the screen for 2000 ms. Afterwards, a behavioral description appeared below the CS. Both stimuli remained together on the screen for 6000 ms. Then, before the next trial, the screen remained blank for an intertrial interval of 2000 ms. The sequence of trials consisted of the (random) presentation of trials Apos (face A followed by one of five positive behavioral descriptions), Bpos, Cneg, Dneg, each of which presented five times. After all twenty trials were presented, participants were instructed that the behavioral statements about two of the persons (A and C) were actually true, and that the statements about the other two people (B and D) were false, and thus had to be mentally reversed. This information was repeated twice.

Once participants had read these instructions, they performed the evaluative priming task. It started with an instruction page that explained the procedural features of the task. The first block was a practice phase consisting of ten trials, in which each target picture was presented together with its name. The participant's task was to name each picture out loud into a microphone by using the word displayed below the picture. This phase served to familiarize the participant with the task and to tune their voice tone to an optimal volume (i.e., to name the stimuli sufficiently loudly), to be easily picked up by the voice key device. After this short practice phase, the experimental block started. It consisted of 120 trials, in which each of the primes (four CSs and two neutral pictures) were presented on the screen for 250 ms. After the prime disappeared, one of the ten target pictures appeared, and stayed on the screen until the participant had responded. The next trial appeared after a 1000 ms intertrial interval. Afterwards, participants completed the explicit valence ratings of the four CSs, before being thanked and debriefed. Unlike Peters and Gawronski (2011), we preferred to not counterbalance the implicit-explicit measurements order, and kept the implicit measure always first. We did so because we were particularly interested in studying the effect of our manipulation on our implicit measure (i.e., evaluative priming in the picture-picture naming task).

Results

Evaluative Priming. The data preparation was based on Spruyt et al. (2007). Neutral priming trials were considered filler trials and were not included in the analyses. In addition, the data from naming trials on which the voice key was not appropriately activated or trials on which an incorrect response was given were excluded from the analysis. Finally, for each of the four crucial priming conditions, response latencies that deviated more than 2.5 standard deviations from a participant's mean latency were also discarded (4%). Implicit evaluation scores were obtained by subtracting, for each prime, the reaction time of trials in which positive targets were presented from that in which negative targets were presented. This way, higher scores reflect quicker responses to positive targets and, therefore, could serve as a measure of implicit preference for the correspondent prime.

We conducted a 2 (validity, true vs. false) x 2 (valence, positive vs. negative) ANOVA on the implicit evaluation scores. Both main effects and the interaction were not significant, all *F*s < 1. Means and standard errors of the reaction times for each prime can be found in Table 2.

Explicit evaluation. The scores for each of the three items of the explicit evaluation questionnaire (likeability, friendliness, and trustworthiness) were averaged, and coded so that high values indicate a preference for the relative CS (high likeability, high friendliness and high trustworthiness). Then, a 2 (validity, true vs. false) x 2 (valence, positive vs. negative) ANOVA was conducted on these scores. We found a main effect of valence, $F(1,26) = 18.97$, $p <.001$, partial $\eta^2 = 0.42$, but no significant main effect of validity, $F(1,26) = 2.95$, $p =.098$, partial $\eta^2 = 0.10$. The interaction was significant, $F(1,26) = 48.36$, $p <.001$, partial $\eta^2 = 0.65$. For the true validity conditions, the CSpos was evaluated more positively than the CSneg, $t(26) = 7.99$, $p < .001$, $d = 1.54$. In addition, for the false validity condition, the explicit evaluations of the CSpos and the CSneg did also differ significantly, $t(26) = -2.71$, $p = .012$, $d = 0.52$, but in the opposite direction (the CSneg was evaluated more positively than the CSpos). Means of all explicit evaluation scores can be found in Table 2.

Table 2. *Means (and standard errors) of the implicit and explicit evaluation scores, Experiment 4.*

US valence	Validity	Implicit evaluation	Explicit evaluation
Positive	True	61,17 (8.80)	5,44 (0.27)
Negative	True	57,10 (9.43)	1,96 (0.24)
Positive	False	60,67 (10.56)	2,75 (0.25)
Negative	False	54,33 (7.94)	3,89 (0.28)

Discussion

The implicit measure used in Experiment 4 (evaluative priming with a naming task) with which we assessed implicit valence after a learning phase adopted from Peters and Gawronski (2011) did not show an effect of the co-occurrences, nor of validity information. These results differ from those that Peters and Gawronski found with evaluative priming of evaluation responses. Given that no effect was present neither of the co-occurrences, nor of the validity information, we assume that our implicit measure may have just failed and did not successfully tap into any potential effect derived from our acquisition phase (neither of co-occurrence nor of validity). The explicit evaluation scores, even if collected for exploratory reasons only, were in line with the findings of Peters and Gawronski. That is, we observed a significant effect of valence, which was moderated by the relational information for both equivalently and oppositely-related CSs.

General Discussion

In four experiments, we tested the extent to which three different implicit measures of evaluation are sensitive to relational information. Each implicit measure represents a specific situation in which people experience stimuli while performing a task that allows one to capture the evaluative response that the presented stimuli evoke. Hence, given that evaluation is defined as the effect of stimuli on evaluative responses, different implicit measures can be conceived as different instances of implicit evaluation. In other words, each different measure creates a different subset of conditions under which implicit evaluation can be observed. Hence, from a functional perspective, studying the impact of relational information in various implicit measures corresponds to an exploration of the conditions under which relational information moderates implicit evaluation.

In order to optimize control over the evaluative responses that stimuli would evoke, we repeatedly paired neutral stimuli (CSs) with affectively relevant stimuli (USs), that is, we realized a typical EC procedure. In addition, we provided relational information that qualified the CS-US relation, via verbal instructions and/or via context. Finally, implicit evaluation of the CSs was assessed employing various implicit measures.

In Experiment 1, the CS paired with a positive outcome (i.e., Xpos) and the CS paired with a negative outcome (i.e., Yneg) were always presented together with a special cue R. This special cue had the function to reverse the outcome of the CS it was paired with. Therefore, if a cue A was followed by a positive outcome (Apos), the same cue A presented with the cue R was then followed by a negative outcome (ARneg). A series of context pairings (Apos, ARneg, Bneg, and BRpos), presented in addition to the target CS-US pairings, followed this rule. The relation implied by this rule (relation of opposition) is particularly intriguing, as its implications are opposite to those of the mere CS-US co-occurrences. After the acquisition phase, a series of causal ratings tested whether participants had learned and applied the rule. Next, implicit evaluation was assessed with an IAT. Results of this experiment showed a reverse implicit EC effect (that is, a preference for the CSneg over the CSpos), thus replicating the results of the short-delay condition of Peters and Gawronski (2011, Experiment 3). The findings of Experiment 1 imply that the IAT is sensitive to the impact of relational information that is implied by the context.

The results of Experiment 1 confirmed that the procedure that was used in this experiment can be used to demonstrate an impact of relational information on implicit measures of evaluation. Hence, in Experiments 2 and 3, we employed the same procedure to test the impact of relational information on the AMP. Unfortunately, the AMP did not replicate the results of the IAT. Instead, it showed no implicit preference for any of the CSs. One possible reason why the AMP did not show any influence of our manipulation would be that the stimuli used in Experiment 2 were not the traditionally employed AMP stimuli, as they were nonwords, whereas the AMP typically features picture stimuli as primes. For this reason, Experiment 3 was a replication of Experiment 2, but with pictures instead of nonwords as CSs and primes. Unfortunately, the results of

Experiment 3 were also inconclusive: despite a good causal learning of the context rule, participants did not show an implicit preference for neither the CSpos or the CSneg.

A number of post-hoc explanations can be constructed for why our AMP performed so poorly in these two experiments. On the one hand, the AMP could have been influenced by both co-occurrences and relational information. Given that the implications of these two pieces of information are opposite to each other (the rule implies that a stimulus presented in a compound paired with a USpos would, if presented on its own, be paired with a USneg), the sum of the two could have resulted in a null effect. On the other hand, it could just be that the AMP was simply insensitive to either co-occurrences or relational information, and therefore it did not tap in the effect of anything (neither CS-US co-occurrences nor CS-US relations). Unfortunately, not having presented a control condition (i.e., a condition in which equivalence, instead of opposition, relational information is presented) prevents us from making strong conclusions about the extent to which AMP effects are sensitive to relational information. For exploratory reasons, we also registered explicit evaluations. Whereas Experiments 1 and 2 revealed an explicit preference in line with the context rule (CSpos liked more than the CSneg), this was not the case in Experiment 3, in which no preference is found. Given that the order of the implicit and explicit measures was not counterbalanced, we refrain from interpreting these results.

In Experiment 4, a different design was adopted. The acquisition phase was inspired by Experiment 3 by Peters and Gawronski (2011) but, unlike to what was the case in their experiment, we used evaluative priming with naming responses as the task to capture implicit evaluation. In addition, only the long-delay condition was implemented. In Experiment 4, two neutral stimuli were repeatedly paired with positive behavioral statements (Apos, Bpos) whereas another two stimuli were consistently paired with negative behavioral statements (Cneg, Dneg). At the end of the series of CS-US pairings, validity information was given in form of instructions. Participants were informed that the relation between two CS-US pairs (Apos and Cneg) was true (i.e., a relation of equivalence), whereas they were informed that the relation between the other two CS-US pairs (Bpos and Dneg) was false and had to be reversed (i.e., a relation of opposition). This way, a fully crossed valence-validity designed was obtained (Apos-true; Bpos-false; Cneg-true and Dneg-false). Afterwards, implicit evaluation for the four CSs was measured, employing an evaluative priming procedure with a naming task. This measure did not show an implicit preference for any of the CSs. It is important to note that an implicit preference was not found even for those CSs that were said to be equivalent to the USs. The results of the original study by Peters and Gawronski (2011, Experiment 3), who found a standard EC effect for these stimuli, and an attenuation for the CSs said to be opposite to the USs, suggest that our picture-picture naming task simply did not tap in any effect of the induction phase. Again, we added at the end of the experiment a series of explicit valence ratings, for exploratory reasons. The explicit liking of the CSs was in line with the valence-relation combination: the CSpos was preferred to the CSneg when the CS-US pairings were equivalently related. Conversely, when the CS-US pairings were oppositely related, the CSneg was explicitly evaluated more positively than the CSpos. Although their order was not counterbalanced with the implicit measure, the results of our explicit ratings show an effect of relational information on the acquisition of explicit evaluation, thus replicating the findings of Peters and Gawronski (2011).

Unfortunately, the results of Experiments 2-4 are inconclusive. On the one hand, one could argue that AMP effects and evaluative priming of naming responses are less sensitive to relational information than the IAT. Whereas Experiment 1 revealed a significant reversed EC effect on the IAT measure, an EC effect was not observed in Experiment 2-4 that employed an AMP of evaluative priming measure. On the other hand, it is possible that our implicit measures in Experiments 2-4 simply failed to capture any evaluation at all. This would mean that based on the results of those experiments, we cannot make any definite statement about the measures' sensitivity to relational information. In Experiment 4, an EC effect was not even present for CSs that were said to be equivalent to USs. This support the idea that the implicit measure in Experiment 4 simply failed to capture the evaluation of stimuli.

Given that we cannot interpret the results of Experiments 2-4 in a conclusive manner, it is also impossible to derive theoretical conclusions based on these results. The results of Experiment 1, however, confirm that relational information can influence IAT effects. As we have discussed in Chapter 1, this finding does constrain theoretical models of implicit evaluation and EC. More research is needed, however, to determine whether different measures of implicit evaluation are differentially sensitive to the impact of relational information.

References

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality and Social Psychology Bulletin, 38,* 1193-1207. doi 10.1177/0146167212446835

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

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230-241. http://www.ucm.es/info/psi/docs/journal/

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*, 347-368. doi: 10.1037/a0014211

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

De Houwer, J. & Vandorpe, S. (2010). Using the Implicit Association Test as a measure of causal learning does not eliminate effects of rule learning. *Experimental Psychology, 57,* 2010, 61-67. doi: 10.1027/1618-3169/a000008

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2012). *A functional-cognitive framework for attitude research*. Manuscript submitted for publication.

De Houwer, J., & Moors, A. (2012). How to define and examine implicit processes? In R. Proctor & J. Capaldi (Eds.). Implicit and explicit processes in the psychology of science (pp. 183-198). NY: Oxford University Press.

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

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

Gawronski, B., & De Houwer, J. (in press). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd edition). New York: Cambridge University Press

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216. doi: 10.1037/0022-3514.85.2.197

Hermans, D., Baeyens, F., & Eelen, P. (2003). On the acquisition and activation of evaluative information in memory: The study of evaluative learning and affective priming combined (pp. 139-168). In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion*. Mahwah, NJ: Lawrence Erlbaum.

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

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y. & Sloman, S. A. (2007) Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*, (pp. 154–72). Oxford University Press.

Moran, T., & Bar-Anan, Y. (2012). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion.* doi: 10.1080/02699931.2012.732040

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

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

Spruyt, A., Hermans, D., De Houwer, J., Vandekerckhove, J., & Eelen, P. (2007). On the predictive validity of indirect attitude measures: Prediction of consumer choice behavior on the basis of affective priming in the picture-picture naming task. *Journal of Experimental Social Psychology, 43,* 599-610. doi: 10.1016/j.jesp.2006.06.009

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

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

CHAPTER 3

THE IMPACT OF RELATIONAL INFORMATION ON IMPLICIT EVALUATION [1]

[1] This study was conducted in collaboration with Jan De Houwer, Anne Gast and Colin Smith

Introduction

The way in which stimulus evaluations are acquired has been drawing the interest of researchers for a long time. We use the term evaluation to refer to the effect of stimuli on evaluative responses (De Houwer, Gawronski & Barnes-Holmes, 2012). Furthermore, the term "implicit evaluation" is meant to refer to instances of evaluation that occur under conditions of automaticity (e.g., under time pressure or without awareness of the effect) whereas "explicit evaluation" refers to all other instances of evaluation (e.g., instances that occur when there is ample time and participants are aware of the effect). In this paper, we focus on one way in which implicit and explicit evaluations can be acquired, namely as the result of stimulus pairings. This type of acquisition is often referred to as evaluative conditioning (EC). EC is defined as the change in the evaluation of a stimulus (named conditioned stimulus, CS) that is due to its pairing with another stimulus (named unconditioned stimulus, US) (De Houwer, 2007; De Houwer, Thomas, & Baeyens, 2001; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). For example, imagine that a person evaluates a stimulus A as neutral (e.g., by indicating a zero score on a Likert scale). In a second phase, A is repeatedly paired with a picture of positive valence. EC can be said to have occurred if, as a result of the stimulus pairings, the person responds to the nonsense word in a more positive way after the pairings than before the pairings (e.g., indicates a positive score on the Likert scale). Research has shown that EC can be found for both implicit and explicit evaluation (e.g., De Houwer, Hermans, & Eelen, 1998; Hermans, Baeyens & Eelen, 2003).

In the present studies on EC, we focus on the role of the information regarding the type of relation between the stimuli that are paired (Hughes, Barnes-Holmes, & De Houwer, 2011). Stimuli and events can be related in many different ways. For instance, one stimulus can be the cause of another stimulus or it can be the effect of another stimulus (e.g., Lagnado, Waldmann, Hagmayer, & Sloman, 2007). The way in which stimuli are related can have important implications for behavior. For instance, if a certain chemical substance in one's blood causes a disease, it makes perfect sense to try to remove the substance from the blood in order to cure the disease. If, however, the substance is merely a harmless byproduct of the disease, filtering it from the blood makes little sense. Until recently, however, EC research (and associative learning research in general) ignored the impact of this vital factor.

In recent years, a handful of studies demonstrated that EC can be moderated by cues that provide information about the nature of the relation between the paired stimuli (Fiedler & Unkelbach, 2011; Peters & Gawronski, 2011; Unkelbach, Förderer, & Stahl, 2012; Zanon, De Houwer & Gast, 2012). In a particularly intriguing set of studies, Peters and Gawronski (2011) tested the joint impact of three factors on both implicit and explicit CS evaluation: CS-US pairings, relational information concerning CS and US, and the point in time at which this relational information became available. Across three experiments, they presented participants with pairings of four neutral faces (CSs) and positive and negative behavioral statements (USs). In the crucial third experiment, the faces A and B were consistently paired with positive statements, whereas the faces C and D were consistently paired with negative statements, for a total of twenty CS-US pairings. Peters and Gawronski also presented verbal instructions that provided relational information. More specifically, participants were either told that the CS and US are equivalent (i.e., the person in the picture does or is what is stated in the behavioral description) or opposite (i.e., the person in the picture does or is the opposite of what is stated in the behavioral description). By doing so, they obtained a fully crossed USvalence-validity design which included a positive-true CS, a positive-false CS, a negative-true CS and a negative-false CS. Finally, they presented the relational information (i.e., are CS and US equivalent or opposite) immediately after each CS-US pairing trial (i.e., short-delay condition) or only after all twenty CS-US pairings had been shown (i.e., long-delay condition). In the subsequent measurement phase, implicit and explicit evaluations of the four faces were assessed. Results showed that, in the short-delay condition, implicit and explicit evaluations of the four faces were in accordance with the combination of valence and validity. That is, participants evaluated the positive-equivalent CS and of the negative-opposite CS more positively than the positive-opposite CS and the negative-equivalent CS. Hence, informing participants that the USs described the opposite of what was true, actually reversed both the explicit and implicit evaluations of the CSs. Conversely, in the long-delay condition, such a reversal was found only for explicit evaluation, whereas the results of the implicit evaluation measures were merely attenuated. More specifically, when the relational information was given after the entire set of pairings, implicit evaluation of the positive-equivalent CS was more positive than that of the negative–equivalent CS, but the evaluation of the positive-opposite CS and negative-opposite CS did not differ.

The results of Peters and Gawronski (2011) have important implications for mental process theories of EC. Current cognitive theories postulate that EC is mediated either by associations, by propositions, or by both associations and propositions. Propositions are statements about the world that can be valid or invalid. When they concern relations between stimuli, propositions can encode not only that stimuli are related but also how they are related. Associations, on the other hand, are merely structures that link mental representations. Associations are therefore not valid or invalid and do not contain relational information (Lagnado et al., 2007). Whereas it is typically assumed that associations are formed as the result of actual stimulus pairings, propositions can be based not only on experience but also on instructions or inferences (De Houwer, 2009; Mitchell, De Houwer, & Lovibond, 2009). Finally, it is often assumed that explicit evaluations reflect the content of propositions whereas implicit evaluations are determined only by associations in memory (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnel, 2006).

Single process association formation models have difficulties dealing with the observation of Peters and Gawronski (2011) that implicit evaluations are impacted by relational information. Because associations do not contain relational information and assuming that implicit evaluations are driven solely by associations in memory, one would indeed expect to observe that CS-US pairings lead to the same implicit evaluations regardless of how the CS and US are said to be related. In order to accommodate the results, single process association formation models could be revised in such a way that association formation is driven not by the actual CS-US pairings but by the mental encoding of those pairings (e.g., Melchers, Lachnit & Shanks, 2004). For instance, if participants see a CS that co-occurs with the US word UNFRIENDLY while being told the CS is opposite to the US, participants might mentally recode the US as FRIENDLY, thus resulting in a pairing of the CS and the concept "friendly". Such mental recoding could lead to an association between the CS representation and the representation of the concept "friendly". One could raise at least two objections against such a post-hoc association formation account of the results of Peters and Gawronski. First, such instances of mental recoding seem to rely on propositional representations in that it requires knowledge about how stimuli are related and about what relation is valid. Hence, although the mental recoding assumption could be added to single process association formation models, it would actually change these models into dual process models that require both propositions and associations. Second, adding the mental recoding assumption would still not explain the second crucial result of Peters and Gawronski, namely the differential effect of the timing of the relational information (during or after the pairings) on implicit and explicit evaluations. If all changes in evaluation are mediated by associations, no such dissociations between implicit and explicit evaluations should arise.

Dual process models of EC and evaluation (e.g., Gawronski & Bodenhausen, 2006, 2011; Rydell & McConnell, 2006) can account for dissociations between changes in implicit and explicit evaluations if it is assumed that both types of evaluations are mediated in a different way by propositions and associations. For instance, one could assume that explicit evaluations are mediated only by propositions whereas implicit evaluations reflect only associations or a mix of associations and propositions. If propositions are sensitive to relational information regardless of when this information is presented, then explicit evaluations would be in line with that relational information regardless of when it is presented. If one also adds the assumption that association formation can be blocked by propositional processes (e.g., when relational information is presented during the pairings) or that it can in part reflect the content of propositional processes (e.g., by mental recoding as the result of relational information that is available during the pairings), it can be explained that implicit evaluations are affected by relational information but more so when that information is presented during the pairings.

The results of Peters and Gawronski (2011) also challenge single process propositional models of EC. Such models postulate that all associative learning, including EC, is mediated by the formation of propositions about stimulus relations (De Houwer, 2009; Mitchell et al., 2009). If this is the case, there seems to be little reason to expect dissociations between EC of implicit and explicit evaluations. Hence, the fact that the time of giving relational information does moderate EC of implicit evaluations but not EC of explicit evaluations seems problematic for single process propositional models of EC.

Nevertheless, a post-hoc propositional account of the results of Peters and Gawronski can be constructed. It is reasonable to assume that, at the time of encoding, participants formed propositions based on both the pairings and the relational information. For instance, the pairing of a CS with the word UNFRIENDLY would result in the proposition “the CS is unfriendly” in the equivalence information condition but in the proposition “the CS is friendly” in the opposition information condition. Importantly, even if relational information was not provided during the pairings, participants might still form propositions about the CS and US relation. Because in daily life, the co-occurrence of two stimuli most often points at some type of similarity or dependence between the stimuli (i.e., a frame of co-ordination; see Hayes, Barnes-Holmes, & Roche, 2001, for a discussion), participants might assume by default that the US provided a valid description of the CS (e.g., “the CS is unfriendly” when the CS and the US word UNFRIENDLY are paired) even when they were not explicitly told that the CS and US were related in this manner. When relational information was presented after the pairings, participants may have constructed a second set of propositions (e.g., “the CS is unfriendly” in the equivalence information condition or “the CS is friendly” in the opposition information condition). Therefore, the resulting attenuation of the implicit evaluation in the invalid information condition could be explained by the conflict between two opposite sets of propositions.

In order to explain that explicit evaluations did reverse in the invalid information condition but implicit evaluations did not, one additional assumption is necessary, namely the assumption that implicit evaluations are more sensitive to initial propositions than explicit evaluations. That is, when two propositions are formed one after the other, the impact of the first proposition relative to the second proposition will be bigger for implicit evaluations than on explicit evaluations. Initial support for this latter assumption is provided by Gregg, Banaji and Seibt (2006, Experiment 3), who presented participants with a description of two fictitious social groups. Whereas group A was initially said to hold positive characteristics, group B was described as holding negative characteristics. Afterwards, participants were told that, due to a computer malfunction, they received the wrong induction procedure, and were asked to mentally switch the descriptions (i.e., that A is negative and B is positive). Hence, participants were given conflicting relational information on two occasions, which should have allowed participants to form two conflicting propositions, one after the other. Gregg et al. observed that, at the end of the experiment, the explicit evaluation of A and B was in line with the most recent information whereas implicit evaluations were in line with the initial information. Although this result suggests that implicit evaluations are influenced more by the first than by the second relational information, one could also explain Gregg et al.’s results by arguing that implicit (but not explicit) evaluations are impacted more by equivalence information than by opposition information. Such an alternative account is possible because equivalence information was provided first (i.e., A is positive, B is negative) and opposition information second (i.e., reverse the initial information). In order to exclude this alternative account, the order in which information is presented (first or second) needs to be manipulated independently from the type of information that is presented (equivalence or opposition). Also note that Gregg et al. measured implicit evaluations twice, once after the initial information and once after the opposition instructions. The fact that implicit measures had been administrated before the presentation of opposition instructions could have rendered them less susceptible to the opposition instructions.

The aim of our research was two-fold. First, we wanted to replicate and extend the results of Peters and Gawronski (2011). Although their results do not refute in a definite manner any of the three classes of models of EC (i.e., association formation, propositional, dual process), they do impose serious constraints on all of these models in that they force the models to change or add core assumptions. Given the high theoretical value of their results, it is important to examine whether the main findings of Peters and Gawronski can be replicated across a range of situations. The need for replication is underscored also by the fact that Moran and Bar-Anan (2012) recently failed to find an impact of relational information on implicit evaluations. Given these mixed results, it is clear that further research is necessary. Apart from trying to replicate the main findings of Peters and Gawronski, we aimed to extend their findings in several ways. First, to test the generality of the findings, we used different stimuli that are related in slightly different ways. We also used an implicit measure different from those used by Peters and Gawronski. Second, in Experiments 2 and 3, we included baseline conditions in which no instructions were given about the way in which stimuli are related. This allowed us to compare the impact of instructions about different types of relations.

The second main aim of our research was to test the post-hoc propositional explanation of the findings of Peters and Gawronski. This could inform us not only about the validity of propositional models of EC but also shed new light on the conditions under which relational information influences implicit and explicit evaluations.

Experiment 1 provided a conceptual replication of the crucial third experiment of Peters and Gawronski (2011) in which the time of presentation of the relational information was manipulated. As in the studies of Peters and

Gawronski, we presented a series of CS-US pairings. This time, however, the CSs were nonwords which were said to have a meaning in a foreign language unknown to the participants. The USs were positive and negative English words. We also presented information that implied a relation of opposition between CSs and USs, namely that each foreign word's meaning was the antonym of the paired English word. The relational content of these instructions was thus very similar to that in the opposition (false information) condition in the study by Peters and Gawronski. Half of the participants received these relational instructions before experiencing the pairings, and the other half received these instructions after the pairings. Later on, implicit and explicit evaluation for the CSs was assessed. Different from Peters and Gawronski (2011), we kept the number of times in which we presented relational (validity) information equal between conditions. In their short-delay condition, Peters and Gawronski presented relational information after each pairing (resulting in 20 presentations of the validity instructions) whereas in their long-delay condition, they presented their validity instructions only twice at the end of the experiment. In our experiment relational information was always given only once, either before or after the pairings.

Experiments 2 and 3 further tested the generality of the effects of Peters and Gawrsonki (2011) and directly tested the two assumptions that must be made by propositional accounts in order to explain their results. The first assumption is that the mere CS-US spatio-temporal co-occurrence represents a cue that suggests that the two stimuli are equivalent. As we noted earlier, the tendency of framing co-occurring stimuli as being equivalent is probably a default response (Hayes et al., 2001). If this assumption is correct, one would expect that the results change if one undermines the spontaneous interpretation of paired stimuli as being equivalent. Therefore, in Experiment 2, we tried to create a context in which the co-occurrence of stimuli is perceived as a cue that implies, instead of equivalence, opposition between the stimuli. To this end, we added a pre-training phase in which we presented three series of stimulus pairings. At the end of each series, we provided a verbal instruction that implied a relation of opposition between the paired stimuli (i.e., we told that the US always corresponded to the antonym of the CS). With this procedure we aimed at creating an environment that encouraged participants to believe that the stimuli paired in the fourth series – the actual CS-US pairings – were opposites. The test series of CS-US pairings was then followed either by verbal instructions implying a relation of opposition between the CSs and the USs (opposition condition), verbal instructions implying a relation of equivalence between the CSs and the USs (equivalence condition) or by no instructions. If participants form propositions about the relation between a CS and a US even when information about the relation is not given explicitly and if we are successful in changing this default way of framing into a frame of opposition, then one would expect a reversed conditioning effect in the no instructions condition. In other words, CSs paired with positive USs should be liked less than CSs paired with negative USs. The reason for this reversal is that, because of the pre-training, a CS-US pairing would be seen as a cue to form the proposition “the CS is opposite to the US”. In the opposition condition, the propositions formed during the CS-US pairings (i.e., “CS is opposite to US”) will be in line with the propositions formed after providing the relational information (i.e., that the US is indeed the opposite of the CS), which, if anything, should strengthen the reversed conditioning effect. In the equivalence condition, however, no conditioning effect should arise. Assuming that during the pairings, participants encode the CS and US as being opposite, these initial propositions should counteract the effect of the propositions formed after presenting the relational information. In sum, we predicted the reverse pattern of results as in the long-delay condition of Peters and Gawrsonki: A stronger (reversed) EC effect in the opposition condition than in the equivalence condition. Moreover, we expected to observe a reversed EC effect in the no instructions condition.

Experiment 3 tested the assumption that when two conflicting propositions are formed one after the other, the first proposition will have a bigger effect on implicit evaluations than on explicit evaluations. To this end, in Experiment 3, we presented two instructions that informed participants about the CS-US relations. One of these was presented before and one after a series of CS-US pairings. Both instructions could either state that the CS and the US were equivalent or that they were opposite to each other. Hence, each participant received one of four different combinations of pre- and post-pairings instructions, which could be either congruent with each other (equivalence-equivalence or opposition-opposition) or incongruent with each other (equivalence-opposition or opposition-equivalence). In case of incongruent combinations of instructions, the second information was always described as being the valid one. We also added a control condition, in which no relational information was presented. After the learning phase, implicit and explicit evaluation of the CSs were measured. If the relative impact of the first proposition is bigger on implicit evaluations than on explicit evaluations, we would expect to observe a bigger difference in EC between consistent conditions (i.e., equivalence-equivalence; opposite-opposite) and inconsistent conditions (i.e., equivalence-opposite; opposite-equivalence) on explicit evaluations than on implicit evaluations.

Experiment 1

Participants were shown a series of trials in which a nonword (CS) was paired with a positive or negative word (US). Either prior to or after the pairings, participants were instructed to imagine that the CSs were actually the Turkish translations of the antonyms of the paired US. For example, when the nonword LOKANTA was presented together with the word HAPPY, participants were asked to imagine that LOKANTA was the Turkish word for SAD. After the learning phase, the implicit evaluations of the CSs were measured using an Implicit Association Test (IAT; Greenwald, McGhee, and Schwartz, 1998). In addition, explicit evaluation was measured, for exploratory reasons, after the implicit measure.

Method

Participants. A total of 532 English-speaking volunteers (mean age = 31.2, 59.2 % women) participated online in the experiment, which was posted on the website of Project Implicit (http://implicit.harvard.edu).

Materials. ***Learning phase.*** As CSs, we used the nonwords LOKANTA and FEVKANI. The words HAPPY and UGLY were used as USs. We refer to the CS that paired with the positive US as CSpos and to the CS paired with the negative US as CSneg. The codes USpos and USneg are used to refer to the positive US and negative US, respectively. Half of the participants were presented with the nonword "LOKANTA" as CSpos and the nonword "FEVKANI" as CSneg, and the reverse assignment was presented to the other half of the participants. Both CSs and USs were presented in white, uppercase letters in 30pt Arial font on the center-left and the center-right of a black screen, respectively. The verbal instructions were written in black, 18pt Arial font, presented on a white page.

IAT. The positive attributes employed in the IAT were the English words GLORIOUS, MARVELOUS, SUCCESS, and WONDERFUL. The negative attributes were the negative English words AGONY, EVIL, FAILURE, NASTY and UNPLEASANT. Both positive and negative attributes were presented in green, 18pt Arial font, in the middle of a black screen. The target stimuli were the nonwords LOKANTA and FEVKANI. They were presented in white, and in eight different combinations of font types (Arial Black and Fixedsys), capitalizations (uppercase and lowercase) and size (16pt and 18pt), resulting in a total of eight different stimuli. This measure was taken in order to avoid categorization based on simple perceptual features of the nonwords, and has been successfully employed in previous research (De Houwer & Vandorpe, 2010; Zanon et al., 2012). In addition, the error message was a red X that appeared in the middle of the screen. If they made an error, participants had to correct their response in order to proceed to the next trial.

Explicit valence ratings. For exploratory reasons, explicit valence ratings were also administered after the IAT. Participants rated their liking of each of the

CSs by answering to the question “How much do you like the word (CS)” by selecting an option on a dropdown menu, indicating a value on a 5-point- scale, with “Not much at all”, “A little bit”, “A moderate amount”, “A lot” and “Very much” as possible answers.

Procedure. At the beginning of the experiment, all participants were instructed that they were about to take part in a study about the meaning of words. The first part of the task was described as consisting of a series of pairings of Turkish and English words. In addition, the instructions also requested participants to pay attention during the presentation of the pairings, and to try to remember the words, because questions about the pairings would be asked in a later stage of the experiment. In condition Instructions Before, participants were also requested to imagine that the Turkish words they were about to see were antonyms of their paired English word, and therefore had the opposite meaning. This was illustrated with the example that if a Turkish word A was to be paired with the English word “fast”, the supposed real meaning of the word A would be “slow”, that is, the antonym of “fast”. In condition Instructions After, participants were simply told that the relation between English and Turkish words would be revealed only at the end of the acquisition phase. After reading the instructions, participants could move to the subsequent acquisition phase by pressing a key.

The acquisition phase consisted in a simple slideshow in which two pairs of neutral nonwords and English words were presented for six times each. The CSpos was consistently paired with the USpos, whereas the CSneg was consistently paired with the USneg. Each CS-US pair was presented for 3000ms. The intertrial interval was 1000 ms.

After the acquisition phase, participants in condition Instructions After were asked to imagine that the Turkish and English words that had just been presented together were antonyms, and therefore had the opposite meaning. The example described above was also added, in order to keep this post-pairings instruction page as similar as possible to the pre-pairings instructions of condition Instructions Before. Conversely, after the CS-US pairings, participants in condition Instructions Before did not receive supplemental instructions and proceeded directly to the subsequent phase.

After instructions and pairings were presented, participants performed a so-called inference check. It consisted of two questions, in which participants were asked to indicate the first letter of the CSs' real meaning. For example, imagine that the Turkish word LOKANTA had been repeatedly paired with the English word HAPPY, and further instructions had indicated to imagine that Turkish and English words are antonyms. A correct answer to the request "Please indicate the first letter of the real English meaning of the Turkish word LOKANTA, would be "s", that is, the initial letter of the word SAD, the antonym of "happy". The aim of the inference test was to check whether participants had understood our instructions and formed the appropriate propositions on the basis of the stimulus pairings and the instructions. The order of the inference test in the sequence was counterbalanced: half of the participants received it before the IAT, the other half received it after the IAT.

After or before the inference check (depending on the counterbalancing condition), the IAT was administered. At the beginning of the task, participants were instructed about the assignment of each category of stimuli (CSpos, CSneg, positive and negative attributes) to the right or to the left key. The first phase (Block 1, B1) was a target discrimination phase, consisting of 20 trials in which the targets (the CSs) were presented 10 times each. The second phase (B2) was an attribute discrimination phase that consisted of 20 trials in which five positive and five negative attributes were presented, twice each. For example, in this phase participants had to press the left key for negative attributes, and the right key for positive attributes. The following two phases, B3 (20 trials) and B4 (40 trials), constituted the first combined phases, in which both attributes and targets were presented, and each of the two keys was assigned to two types of stimuli. For example, participants had to press the left key to categorize the CSneg and the negative stimuli, and the right key to categorize the CSpos and the positive stimuli. These two blocks were followed by B5 (40 trials), that is, an inverted target discrimination phase, in which the target words were presented as in B1, but with inverted labels. In the example, participants in this phase would have to categorize the CSpos using the left key, and the CSneg using the right key. The last phase was the second set of the combined phase, that consisted of B6 (20 trials) and B7 (40 trials), in which both targets and attributes were presented. The difference from the first combined phase (B3 and B4) is that the CSs had to be categorized according to reversed response assignments.

After completing the IAT, participants were asked to rate the valence of the CSs, as a measure of explicit evaluation. Finally, participants were asked a last memory question, namely to select, from a dropdown list, the US that was paired with each CS.

Results

Implicit evaluation. IAT data from 23 of the 532 participants (4.3%) were excluded from analysis because of too high error rates (larger than 40% in a single block or larger than 30% overall), resulting in valid IAT scores for 509 participants. The IAT scores were calculated using the *D2*-algorithm (Greenwald, Nosek, & Banaji, 2003). This algorithm impie that latencies for all trials are calculated from the onset of the stimulus until the correct response is made, also for trials in which initially a wrong response was given. Trials with reaction times shorter than 400 milliseconds are discarded. Positive scores indicate an implicit preference for the CSpos over the CSneg, that is, a preference for the conditioned stimuli in line with the valence of the unconditioned stimuli they co-occurred with.

Overall, the IAT score was not significantly different from zero, $M = -0.01$, $SD = 0.50$, $t(509) = -0.63$, $p = .53$, $d = 0.03$. We conducted a 2 (Instructions Order: Before the pairings vs. After the pairings) x 2 (Inference Check Order: Pre-IAT vs. Post-IAT) ANOVA on the IAT scores. Results did not show an interaction between both variables, $F < 1$. The main effect of the Inference Check Order was not significant, $F < 1$, suggesting that the timing of the inference check in the procedure had no significant effect on the outcome of the implicit measure. The main effect of Instructions Order, however, was significant, $F(1, 505) = 9.52$, $p <$

.01, partial $\eta^2 = 0.02$. In condition Instructions Before, the IAT score was negative ($M = -0.08$, $SD = 0.50$) and significantly different from zero, $t(250) = -2.60$, $p = .011$, $d = 0.16$, whereas in condition Instructions After, the IAT score was positive ($M = 0.05$, $SD = 0.48$) but not significantly different from zero, $t(257) = 1.74$, $p = .082$, $d = 0.10$. Thus, opposition instructions presented before the CS-US pairings led to implicit evaluation of the CSs opposite to the valence of their paired USs. On the other hand, when opposition instructions were presented after the CS-US pairings, participants showed no significant preference for one CS over the other. In an additional analysis based on only those participants who correctly answered the questions in the inference check ($n = 236$) and who therefore were likely to have processed all instructions correctly, the main effect of Instructions Order was still significant, $F(1, 232) = 11.80$, $p < .001$, partial $\eta^2 = 0.05$. In particular, within this sample, the IAT score in condition Instructions Before was still negative ($M = -0.16$, $SD = 0.46$) and significantly different from zero, $t(134) = -4.12$, $p < .001$, $d = 0.35$. In condition Instructions After, the IAT score was still positive ($M = 0.07$, $SD = 0.48$), but not significantly different from zero, $t(100) = 1.41$, $p = .16$, $d = 0.15$.

Explicit evaluation. Explicit ratings were included for exploratory reasons only and should be interpreted with caution because they were always collected after the IAT. During the IAT, forgetting or additional learning might have taken place which could have biased the subsequent evaluative ratings. To make the results more comparable to those on the implicit measure, we computed difference scores based on the two CS ratings (CSpos - CSneg) that reflect the preference for the CSpos over the CSneg. A high explicit evaluation score thus represents a preference for the positively paired CS over the negatively paired CS. Data from eight participants who did not answer one or both of the explicit valence rating questions were excluded from the analyses. Overall, the explicit evaluation score was positive ($M = 0.65$, $SD = 2.19$), but not significantly different from zero, $t(523) < 1$, *ns.,* $d = 0.30$. We conducted an ANOVA that did not reveal an interaction between Instruction Order and Inference Check Order, $F < 1$, nor a main effect of Inference Check Order, $F < 1$. The main effect of Instruction Order, however, was significant, $F(1, 520) = 10.63$, $p < .01$, partial $\eta^2 = 0.02$. The explicit evaluation score for participants in condition Instructions Before was negative ($M = -0.25$, $SD = 2.23$), but not significantly different from zero, $t(257) = -1.71$, $p = .09$, $d = 0.11$. In condition Instructions After, the explicit evaluation score was positive ($M = 0.37$, $SD = 2.00$), and significantly different from zero, $t(265) = 3.00$, $p = .003$, $d = 0.19$. When opposition instructions were presented only after the experience of the pairings, participants thus showed a standard EC effect, that is, a preference for the CSpos over the CSneg. When participants are instructed before the pairings about their opposite relations, the EC effect is descriptively (but not significantly) reversed.

Also for the ratings, we performed an additional analysis based only on the data of participants who performed the inference test correctly ($n = 243$). Overall, participants showed a negative explicit evaluation score ($M = -0.22$, $SD = 2.14$), but not significantly different from zero, $t(242) = -1.59$, $p = .11$, $d = 0.10$. The ANOVA did not reveal a significant Instruction Order x Inference Check Order interaction, $F < 1$. We found, however, a main effect of Instruction Order, $F(1, 239) = 4.83$, $p = .03$, partial $\eta^2 = 0.20$. In the reduced sample, participants in condition Instructions Before showed a negative explicit evaluation score ($M = -0.53$, $SD = 2.19$) which was different from zero, $t(136) = -2.81$, $p < .01$, $d = 0.24$. Participants in condition Instructions After showed a positive score ($M = 0.18$, $SD = 2.02$), but their preference did not differ from zero, $t(105) = -0.91$, *ns,* $d = 0.09$. This analysis shows that, among participants who probably made the correct inference about the real nature of the CSs, explicit evaluation was in line with the valence implied by the inference but only in condition Instructions Before. We also found a main effect of Inference Check Order, $F(1, 239) = 9.98$, $p = .002$, partial $\eta^2 = 0.40$. Participants who performed the inference check before the IAT, showed a positive explicit evaluation score ($M = 0.14$, $SD = 2.05$), but this preference did not differ from zero, $t(146) = 0.85$, *ns,* $d = 0.07$. Furthermore, participants who received the inference test after the IAT, and therefore right before the explicit evaluation measure, showed a negative explicit evaluation score ($M = -0.77$, $SD = 2.16$), which differed significantly from zero, $t(95) = -3.49$, *p* = .001, *d* = 0.36. This result suggests that, for participants who made correct inferences about the nature of the CSs, the proximity of the inference check to the moment in which the explicit evaluation took place led to a preference for the CSneg over the CSpos, that is, a preference in line with the valence implied by the inference and not by the mere co-occurrences. Implicit and explicit evaluation scores were significantly correlated, *r* = .40, *p* = .01.

Discussion

Experiment 1 provided a conceptual replication of Experiment 3 of Peters and Gawronski (2011). We presented a series of CS-US pairings, as well as relational information that implied a relation of opposition between CSs and USs. We also manipulated the time at which this relational information was available, that is, before or after the CS-US pairings. Afterwards, implicit and explicit evaluation were measured. Our results confirmed those of Peters and Gawronski, showing that opposition instructions that were presented before the experience of CS-US pairings led to an implicit preference for the CS paired with a negative US over a CS paired with a positive US. When instructions implying opposition were presented after all the CS-US pairings were shown, there was no difference between the implicit evaluation of the CSpos and the CSneg. We cannot draw strong conclusions about explicit evaluation, since our explicit valence ratings were presented always after the implicit measure. However, our results did not reveal a difference between the evaluation of the CSpos and CSneg in condition Instructions Before but explicit evaluations were in line with the valence of the CS-US pairings in condition Instructions After.

Experiment 2

With Experiment 2, we aimed at testing the assumption that the mere co-occurrence of a CS and US is perceived as a cue that suggests equivalence of those stimuli and therefore leads to an evaluation of the CS that is in line with the evaluation of the US. We therefore tried to create a context in which the mere co-occurrence of two stimuli could be spontaneously interpreted as a cue that the two stimuli are opposite to each other. To this end, we presented three series of stimulus pairings before the target CS-US series was shown. Each of the three training series was followed by verbal instructions that implied a relation of opposition between the stimuli being shown. We hoped that, as a result of this pre-training phase, participants would spontaneously frame the CS-US pairings of the fourth series as being opposite in valence. In addition, after the acquisition phase, we presented instructions implying a relation of opposition between the CSs and the USs, instructions implying a relation of equivalence between the CSs and the USs or no instructions. Finally, implicit and explicit evaluation of the CSs were measured. In order to measure implicit evaluations, some participants completed the IAT whereas others completed a personalized version of the IAT. Although both measures are procedurally similar, there is evidence suggesting that they can be dissociated in important ways (e.g., Han, Czellar, Olson & Fazio, 2010). For the present purposes, however, it is only important to note that by using different measures, we could test the generality of our findings.

Method

Participants. A total of 1098 English-speaking volunteers (mean age = 29; 69.1 % women) participated online in the experiment, which was posted on the website of Project Implicit (http://implicit.harvard.edu). The computer program randomly assigned participants to one of three groups (Conditions Equivalence instructions, Opposition instructions, or No instructions).

Materials. ***Learning phase.*** In the learning phase, pairings of neutral and affectively relevant words, as well as verbal instruction pages were presented. The nonwords were “SARICIK”, “ENANWAL”, “SOKOG”, “UYATA”, “HILEJ”, “AFUBO”, “LAAPIAN”, and “NIFFIAN”, whereas the affectively relevant words were the English words “LIFE”, “SAD”, “PLEASANT”, “SICKNESS”, “LOVE”, “FAILURE”, “HAPPY” and “UGLY”. The CSs were the nonwords “LAAPIAN” and “NIFFIAN”, and the USs were the words “HAPPY” (USpos) and “UGLY” (USneg).

Half of the participants were presented with the nonword “LAAPIAN” as CSpos and the nonword “NIFFIAN” as CSneg, and the reverse assignment was presented to the other half of the participants. Both nonwords and affectively relevant were presented in white, capital letters on, respectively, the center-left and the center-right of a black screen, in 30pt Arial font. The verbal instructions were written in black, 18pt Arial font, presented on a white page.

IAT and Personalized IAT. The stimuli employed in the IAT and in the Personalized IAT (Olson & Fazio, 2004) were identical to those of the IAT of Experiment 1, except for the target stimuli used (the nonwords “LAAPIAN” and “NIFFIAN”).

Explicit valence ratings. Participants rated their liking of each of the CSs with the same scale as used in the previous experiments.

Procedure. Like in Experiment 1, at the beginning of the experiment participants were instructed that they were about to take part in a study about the meaning of words, and that they would see pairings of Turkish and English words. In this experiment, the initial instructions stated that there were in total four series of pairings. Participants were encouraged to pay attention to the pairs, because questions were going to be asked about this later on in the experiment. In addition, participants were informed that the relationship between the presented words was to be communicated only after the series of pairings. After reading the instructions, participants could move to the first acquisition phase by pressing a key. The first pairing series consisted of the Apos and Bneg pairs (nonword A paired with a positive word and nonword B paired with a negative word), paired six times each. After all twelve pairs were shown, an instruction page was presented that asked participants to imagine that the Turkish and English words presented together were antonyms of each other, that is, that they had the opposite meaning in a similar fashion to Experiment 1. We will again refer to this type of instructions as “opposition instructions”. In order to facilitate the understanding of the relation, also an example (e.g., if you had seen the pair “KADIRGA – LIGHT”, the real meaning of the word “kadirga” would be “dark”, that is, the opposite of “light”) was given. At this stage, participants were also asked to take a moment to think about the real meaning of the allegedly Turkish words, in light of the instructions just received. By pressing a key, participants could move on to the next phase, which was a manipulation check in which they had to indicate the English meaning of the nonwords A and B. With this manipulation check we could assess whether participants had combined the information derived from the pairings with that derived from the post pairings instructions. After the first series' manipulation check, another two series of pairings (Cpos, Dneg and Epos, Fneg), opposition instructions and manipulation checks followed, with the same modalities as the first series. After the third series, the CS-US pairings were presented in the same way as the A-B, C-D, and E-F pairings.

What followed after the CS-US pairings varied across conditions. In condition Equivalence instructions, participants were asked to imagine that the CSs were actually synonyms of the USs they were paired with, and that they hence had the same meaning. An example was also given to make sure that the relation was clearly described. In condition Opposition instructions, participants were asked to imagine that the CSs were actually antonyms of the USs they were paired with, and that they therefore had the opposite meaning. Also in this case, an example of this relation was provided. In condition No instructions, no further instructions were presented. In conditions Opposition instructions and Equivalence instructions, participants were asked to take a little time to think about the real meaning of the CSs, that is, to integrate the information implied by the CS-US pairings with the information implied by the post-pairings instructions. After this last instruction page, followed the implicit measures, designed to measure the implicit evaluation of the CSs. Half of the participants were administered an IAT, and the other half on a personalized IAT. The personalized IAT differed from the standard IAT in two important ways. First, the attribute labels "Positive" and "Negative" were replaced by the labels "I like" and "I don't like". Second, there was no error feedback for the attributes. In our version of the personalized IAT, we decided, however, not to implement the second feature. This decision was taken due to the fact that the experiment was conducted online and that it was therefore likely that some participants would without error feedback give random responses. This version of the personalized IAT had already been successfully employed by Smith, De Houwer and Nosek (in press). In addition, in order to ensure the non-normative nature of the categorization task, participants were asked to indicate whether they liked or disliked each of the attributes employed in the last phase of the experiment. This procedure allowed us to select only the participants that genuinely liked the positive attributes and disliked the negative attributes.

After the implicit measure, but only in no instructions condition, participants were informed that the meaning of the CSs was actually the opposite of that of the English words they were paired with (similar to the instructions of condition Opposition instructions). The subsequent phase consisted of a series of explicit ratings of the pleasantness of the CSs that served as a measure of explicit evaluation. After the explicit valence ratings were given, a series of manipulation checks was presented, in order to test participants' memory of the stimuli presented. Finally, the group who performed the personalized IAT also rated whether they liked or disliked the attributes presented in the implicit measure

Results

Implicit evaluation. For both the standard and the personalized IAT, scores were calculated using the *D2*-algorithm (Greenwald et al., 2003). We excluded IAT data from 47 participants (4.3%) because of too high error rates (greater than 40% in a single block or greater than 30% overall) and data of 185 participants (16.8%) who performed the personalized IAT but did not indicate to like all likeable attributes and dislike all dislikeable ones. This restriction did not affect the results. The resulting sample consisted of 899 participants who were randomly assigned to one of three groups, resulting in 304 participants assigned to condition Equivalence instructions, 298 participants assigned to condition Opposition instructions and 297 participants assigned to condition No instructions. As in Experiment 1, positive IAT scores indicate an implicit preference for the CSpos over the CSneg, that is, a preference for the conditioned stimuli in line with the valence of the unconditioned stimuli they co-occurred with.

Overall, the IAT score was positive (M = 0.10, SD = 0.53), and significantly different from zero, $t(865) = 5.72$, $p < .001$, d = 0.19. This result shows an implicit evaluation of the CSs in line with the valence of the US with which they co-occurred. We conducted a 3 (Instructions: Equivalence, Opposition, No instructions) x 2 (Implicit measure: Standard vs. Personalized IAT) ANOVA. The interaction was not significant, $F(5, 860) = 2.21$, p = .11, but there was a significant main effect of Instructions, $F(2, 860) = 4.47$, p = .012, partial $\eta^2 = 0.01$. Participants in condition Equivalence instructions showed a positive IAT score (M = 0.14, SD = 0.55), that was significantly different from zero, $t(290) = 4.24$, $p < .001$, d = 0.25. Also in condition No instructions the IAT score was positive (M = 0.14, SD = 0.52), and significantly different from zero, $t(289) = 4.42$, $p < .001$, d = 0.27. Therefore in these two conditions, participants showed an implicit evaluation in line with the valence of the US with which they co-occurred. Moreover, in condition Opposition instructions, the IAT score was also positive (M = 0.04, SD = 0.51), but it did not differ from zero, $t(284) = 1.17$, p = .24. A contrast test showed that the IAT score in condition Equivalence instructions did not differ from that in condition No instructions, $t(863) = 0.02$, p = .99, whereas the IAT score in condition Opposition instructions significantly differed from condition No instructions, $t(863)$ = -2,28, p = .02. The contrast test between condition Equivalence instructions and condition Opposition instructions also showed a significant difference, $t(863)$ = -2,30, p = .02. These results show that the instructions in condition Opposition instructions led to an attenuation of the EC effect but the instructions in the equivalence condition did not strengthen the IAT effect. Finally, there was also a main effect of implicit measure, $F(1, 860)$ = 4.64, p = .031, partial $\eta^2 = 0.005$. The personalized IAT showed a stronger preference for the CSpos over the CSneg (M = 0.15, SD = 0.53), $t(364) = 5.22$, $p < .001$, d = 0.28, than the standard IAT, (M = 0.07, SD = 0.53), $t(500) = 3.07$, p = .002, d = 0.13.

Explicit evaluation. As in Experiment 1, explicit evaluation measures were administered for exploratory reasons only and always administrated after the IAT. Like with the implicit measure, scores for each of the two explicit valence rating questions were combined into a single score that reflects the preference for the CSpos over the CSneg. In other words, a high explicit evaluation score represents a preference for the positively paired CS over the negatively paired CS. Data of 26 participants who did not answer to one or both of the explicit valence rating questions were excluded from the analyses. Overall, the explicit evaluation score was positive (M = 0.41, SD = 2.70), and significantly different from zero, $t(872)$ = 4.47, p < .001, d = 0.15. This result shows an explicit preference for the CSpos over the CSneg. A 3 (Instructions: Equivalence, Opposition, No instructions) x 2 (Implicit measure: Standard vs. Personalized IAT) ANOVA was also conducted. The interaction was not significant, $F(5, 867)$ = 1.99, p = .14, nor was the main effect of the type of Implicit measure, $F(1, 860)$ = 1.65, p = .20. However, the main effect of Instructions was significant, $F(2, 860)$ = 26.99, p < .001, partial η^2 = 0.06. Participants in condition Equivalence instructions showed a positive explicit evaluation score (M = 1.28, SD = 2.55), that was significantly different from zero, $t(294)$ = 8.64, p < .001, d = 0.50. This result represents an explicit preference for the CSpos over the CSneg. In condition Opposition instructions the explicit evaluation score was negative, (M = -0.45, SD = 2.66), but was not different from zero, $t(289)$ = -0.29, p = .77. Also in condition No instructions the explicit evaluation score was negative (M = -0.31, SD = 2.68), and not different from zero, $t(287)$ = -0.20, p = .84. A contrast test showed that the manipulation in condition Opposition instructions did not differ from condition No instructions, $t(870)$ = -0.06, p = .95, whereas the manipulation in condition Equivalence instructions significantly differed from condition No instructions, $t(870)$ = 6.03, p < .001. The contrast test between condition Equivalence instructions and condition Opposition instructions showed a significant difference, $t(870)$ = 6.10, p < .001. Thus, in conditions Opposition instructions and No instructions, explicit evaluation scores show an attenuation of the EC effect. It is important to note that, in condition No instructions, relational information implying a relation of opposition was given after the IAT. Therefore, whereas implicit evaluation could be influenced only by the acquisition phase, explicit evaluation could have been influenced also by these additional instructions. Overall, implicit and explicit evaluation scores were significantly correlated, $r = .28$, $p < .001$. This was the case also for each of the three conditions analyzed separately (condition Equivalence instructions, $r = .30$, $p < .001$; condition Opposition instructions, $r = .36$, $p < .001$; condition No instructions, $r = .18$, $p < .01$)

Discussion

Results showed that the training series of opposite CS-US pairings did not lead to a reverse EC effect in any of the conditions. In the conditions Equivalence instructions and No instructions participants showed an implicit preference in line with the valence of the co-occurrences, that is, a standard EC effect. In condition Opposition instructions, however, participants showed an attenuation of the EC effect, therefore confirming that post-pairings opposite instructions can lead to a significant reduction of the standard EC effect (as shown also in Experiment 1). There are at least two interpretations of this result in terms of the mental processes that mediate EC. First, it might indicate that, unlike to what is assumed on the basis of propositional models, participants do not spontaneously form propositions about CS-US pairings. For instance, it is possible that in the absence of explicit relational information, CS-US pairings result only in the formation of associations in memory. These associations result in a transfer of valence from the US to the CS, an effect that can only be counteracted but not undone by subsequent information that the CS and US are actually opposites. Second, it is possible that participants still framed the CS and US as equivalent despite the pre-training that encouraged a framing in terms of opposition. In other words, one could argue that the pre-training was ineffective in changing the default assumption that CS-US pairings indicate equivalence and therefore did not provide an adequate test of propositional models of EC.

Although we cannot distinguish between these two options, our results do provide important new information about the effects of relational information on EC of implicit evaluations. Interestingly, the EC effect was as big in the No instruction condition as in the Equivalence instruction condition. This means that after CS-US pairings in the absence of explicit relational information, equivalence instructions seem to add little. Instead, difference between providing post-acquisition equivalence instructions and post-acquisition opposition instructions seems to be driven entirely by the opposition instructions. This results further constraints models of EC. Models which assume that CS-US pairings in the absence of explicit relational information result in the formation of associations must either assume that implicit evaluations are driven only by the associations (which contradicts the fact that post-acquisition opposition instructions do attenuate implicit evaluations) or that the effect of associations and propositions that imply equivalence is non-additive. Propositional models must assume that the co-occurrence of CS-US associations is a strong cue for the formation of propositions about the equivalence of the CS and US. Once these propositions have been formed during acquisition, providing equivalence instructions does not change or strengthen the effect of the initial propositions.

Experiment 3

The aim of Experiment 3 was to test assumption that the impact of the initial relational information is bigger on implicit evaluation than on explicit evaluation. Five groups of participants experienced a series of pairings of CSs and USs. Two out of five groups received congruent pre- and post-pairings instructions (i.e., Equivalence-Equivalence, Opposition-Opposition), whereas other two groups received incongruent instructions (i.e., Equivalence-Opposition, Opposition-Equivalence). A fifth control group received no relational information. After instructions and pairings were presented, implicit and explicit evaluation of the CSs were measured.

Method

Participants. A total of 524 English-speaking volunteers (mean age = 32.45, 65.5 % women) participated online in the experiment that was posted on the website of Project Implicit (http://implicit.harvard.edu).

Materials. ***Learning phase.*** As in the first two studies, the learning phase consisted of pairings of stimuli and of verbal instructions. The stimuli used as CSs were the nonwords “LAAPIAN” and “NIFFIAN”, whereas the stimuli used as USs were the English words “HAPPY” (USpos) and “UGLY” (USneg). The CSpos was the CS paired with the positive US, and the CSneg was the CS paired with the negative US. Half of the participants were presented with the nonword “LAAPIAN” as CSpos and the nonword “NIFFIAN” as CSneg, and the reverse assignment was presented to the other half of the participants. Both CSs and USs were presented in white, capital letters on, respectively, the center-left and the center-right of a black screen, in 30pt Arial font. The verbal instructions were written in black, 18pt Arial font, presented on a white page.

Implicit association test. The IAT employed in this experiment was identical to that used in the previous experiments.

Explicit valence ratings. Participants rated their liking of each of the CSs with the same scale as used in the previous experiments.

Procedure. The cover story presented to the participants was identical to that of Experiments 1 and 2. Equivalence instructions implied that the Turkish and the English words being paired were synonyms, hence participants had to assume that they had the same meaning. Conversely, opposition instructions implied that the Turkish and English words being paired were antonyms, and therefore participants had to assume that they had the opposite meaning. After reading the initial instructions, participants could move to the following stage by pressing a key. The second phase consisted of six pairings of the CSpos and the USpos and six pairings of the CSneg and the USneg. In addition, the point in time at which specific relational instructions were presented, that is, before or after the pairings, was manipulated. In our manipulation, we fully crossed the type of the first relational information (Equivalence vs. Opposition) and the type of the second relational information (Equivalence vs. Opposition), resulting in four different conditions (Equivalence-Equivalence, Equivalence-Opposition, Opposition-Equivalence, Opposition-Opposition) between participants. For example, a participant in condition Equivalence-Opposition learned (1) that the CSs have the same meaning as the USs they are paired with, (2) experience the CS-US pairings, and (3) learn that the CSs were actually opposite in meaning to the USs they were paired with, and that hence the initial instructions were incorrect. In addition, a condition No instructions, in which no relational information was shown either before or after the pairings adding up to a total of 5 conditions. After the post pairings instructions were presented, participants performed implicit and explicit measures of evaluation for the CSs, of which, unlike in Experiments 1 and 2, the order was counterbalanced across participants. The counterbalancing of the order of the implicit-explicit evaluation measures was crucial to test our research question, as it allowed us to directly compare implicit and explicit evaluation. The implicit evaluation was measured with an IAT. Explicit evaluation was measured by a question.

Results

Implicit-explicit evaluations comparison. The main aim of Experiment 3 was to compare the relative impact of the first vs. second proposition on implicit and explicit evaluations. In order to allow for the crucial comparison of the two types of evaluation, scores for implicit and explicit evaluation were z-transformed. Then, a 2 (type of evaluation: implicit vs. explicit) x 2 (pre-instructions: equivalence vs. opposition) x 2 (post-instructions: equivalence vs. opposition) ANOVA was administered. For the IAT, the raw scores were calculated using the *D2*-algorithm (Greenwald, Nosek, & Banaji, 2003). As in Experiments 1 and 2, positive scores indicate an implicit or explicit preference for the CSpos over the CSneg, that is, a preference for the conditioned stimuli in line with the valence of the unconditioned stimuli they co-occurred with.

The main effect of type of evaluation was not significant, $F(1, 391) = .04$, *ns*, whereas the interaction between type of evaluation and pre-instruction was significant, $F(1, 391) = 8.03$, $p < .01$, partial $\eta^2 = 0.02$, as well as the interaction between type of instructions and post-instructions, $F(1, 391) = 12.07$, $p = .001$, partial $\eta^2 = 0.03$. However, the three-way interaction was not significant, $F(1, 391) = 0.20$, $p = ns$.

In order to unravel the crucial two-way interactions we conducted a series of t-tests. The analysis showed that, when averaged across post-instruction condition, implicit evaluation was significantly influenced by the relation implied in the pre-instructions (Equivalence, $M = 0.12$, $SD = 1.00$; Opposition, $M = -0.33$, $SD = 0.94$), $t(401) = 4.54$, $p < .001$, $d = 0.45$. In line with our predictions, this effect was not present for explicit evaluation (Equivalence, $M = -0.04$, $SD = 0.98$; Opposition, $M = -0.14$, $SD = 1.05$), $t(410) = 1.05$, $p = .29$, $d = 0.10$. We used a similar test also to check the effect of the post-instructions. Implicit evaluation was significantly influenced by the content of post-instructions, (Equivalence, $M = 0.03$, $SD = 1.03$; Opposition, $M = -0.23$, $SD = 0.95$), $t(401) = 2.59$, $p = .01$, $d = 0.26$. The effect of post-instructions on explicit evaluation had the same direction but was more outspoken (Equivalence, $M = 0.24$, $SD = 0.88$; Opposition, $M = -0.42$, $SD = 1.04$), $t(410) = 6.91$, $p < .001$, $d = 0.69$.

Implicit evaluation. IAT data from 25 participants were excluded (4.8%) because of too high error rates (greater than 40% in a single block or greater than 30% overall). The resulting sample consisted of 499 participants. The counterbalancing of the order in which implicit and explicit measures were presented did not appear to have an effect on implicit evaluation, $F(1, 402) = 1.17$, $p = .28$. Overall, the IAT score was positive ($M = 0.18$, $SD = 0.56$), and significantly different from zero, $t(498) = 7.04$, $p < .001$, $d = 0.32$.This result shows an overall implicit evaluation of the CSs in line with the valence of their co-occurrences. In order to unravel the effect of the pre- and post- pairings instructions and their different relational content (equivalence vs. opposition relations), we conducted a 2 (pre-instructions: equivalence vs. opposition) X 2 (post-instructions: equivalence vs. opposition) ANOVA. The main effect of pre-instructions was significant, $F(1, 402) = 22.27$, $p < .001$, partial $\eta^2 = 0.04$. Participants who received equivalence instructions prior to the pairings showed an implicit liking ($M = 0.24$, $SD = 0.57$) in line with the co-occurrences, and this value was significantly different from zero, $t(208) = 6.17$, $p < .001$, $d = 0.42$. Conversely, participants who received opposition pre-instructions showed no implicit preference ($M = -0.01$, $SD = 0.53$) for any of the CSs, $t(193) = -0.17$, $p = .86$, $d = 0.02$. The main effect of post-instructions was also significant, $F(1, 402) = 8.38$, $p < .01$, partial $\eta^2 = 0.02$. Participants who received post-pairings equivalence instructions showed a liking in line with the valence of the co-occurrences ($M = 0.19$, $SD = 0.58$), and this value was significantly different from zero, $t(207) = 4.77$, $p < .001$, $d = 0.33$. On the other hand, those who received opposition post-pairings instructions did not show an implicit preference for any of the CSs ($M = 0.05$, $SD = 0.54$), $t(194) = 1.25$, $p = .21$, $d = 0.09$. The interaction between pre- and post-pairings instructions was not significant, $F(3, 402) = 0.21$, $p = .65$.

Participants in condition No instructions showed a positive IAT score ($M = 0.41$, $SD = 0.50$), that reflects an implicit preference in line with the valence implied by the co-occurrences, $t(95) = 7.98$, $p < .001$, $d = 0.82$. Contrast tests were used to compare the IAT score in condition No instructions with those of the other four experimental conditions. Except for the Equivalence-Equivalence condition, all other conditions significantly differed from the No instructions condition in terms of the IAT score. Means of the IAT scores and results of the contrast test for all five conditions can be found in Table 1.

Explicit evaluation. Data of 13 participants (2.5%), who did rate one or both of the CSs, were excluded from the analyses, which left us with a sample of 511 participants. The counterbalancing of the order in which implicit and explicit measures were presented had also no effect on explicit evaluation, $F(1, 511) = 1.96$, $p = .16$. Overall, the explicit evaluation score was positive ($M = 0.40$, $SD = 2.39$), and significantly different from zero, $t(510) = 3.80$, $p < .001$, $d = 0.17$. Participants explicitly evaluated the CSpos more positively than the CSneg. To test the effect of the type of relational information and of the timing of its availability, we conducted a 2 (pre-instructions: equivalence vs. opposition) X 2 (post-instructions: equivalence vs. opposition) ANOVA. The main effect of pre-instructions was not significant, $F(1, 408) = 2.22$, $p = .14$. Analyzing the conditions separately, however, showed that participants who received equivalence pre-instructions evaluated the CSpos significantly more positively than the CSneg ($M = 0.32$, $SD = 2.35$), $t(216) = 1.99$, $p < .05$, $d = 0.14$. This difference was not significant for participants who received opposition pre-instructions ($M = 0.07$, $SD = 2.51$), $t(194) = 0.37$, $p = .71$. The main effect of post-instructions was significant, $F(1, 408) = 48.73$, $p < .001$, partial $\eta^2 = 0.11$. Participants who received equivalence post-instructions evaluated the CSpos more positively than the CSneg ($M = 0.97$, $SD = 2.10$), $t(209) = 6.68$, $p < .001$, $d = 0.46$, whereas those who received opposition post-instructions evaluated the CSneg to be more positive than the CSpos ($M = -0.62$, $SD = 2.49$), $t(201) = -3.42$, $p < .01$, $d = 0.21$. The interaction between pre-instructions and post-instructions was not significant, $F(1, 408) < 1$, *ns.* Finally, participants in condition No instructions showed an explicit evaluation of the CSs in line with the valence of the CS-US co-occurrences ($M = 1.24$, $SD = 2.03$), $t(98) = 6.10$, $p < .001$, $d = 0.61$. A contrast test compared the explicit evaluation score in condition No instructions with those of the other four experimental conditions. Except for the Equivalence-Equivalence condition and of the Opposition-Equivalence condition, scores for all other conditions significantly differed from the No instructions condition. Means of the explicit evaluation scores and results of the contrast test for all five conditions can be found in Table 2. Implicit and explicit evaluation scores were significantly correlated, $r = .30$, $p = .01$.

Table 1. *Means (and standard errors) of the implicit and explicit evaluation scores, and the results of contrasts test with the condition No Instructions, Experiment 3.*

Instructions	Implicit evaluation	Explicit evaluation	Contrast: Implicit evaluation	Contrast: Explicit evaluation
Equivalence-Equivalence	0.31 (0.59)**	1,12 (1,92)**	$t(494) = -1.29$, $p = .20$	$t(506) = -0.38$, $p = .71$
Equivalence-Opposition	0.18 (0.54)**	-0.44 (2,47)	$t(494) = -3.05$, $p < .01$	$t(506) = -5.42$, $p < .001$
Opposition-Equivalence	0.08 (0.55)	0.81 (2,26)**	$t(494) = -4.35$, $p < .001$	$t(506) = -1.38$, $p = .17$
Opposition-Opposition	-0.11 (0.50)*	-0.80 (2,51)**	$t(494) = -6.49$, $p < .001$	$t(506) = -6.24$, $p < .001$
- (No instructions)	0.41 (0.50)**	1,24 (2,03)**	-	-

* significantly different from zero at the .05 level

** significantly different from zero at the .01 level

Discussion

Aim of Experiment 3 was to examine the assumption that when two conflicting proposition are formed one after the other, the first proposition would more strongly influence implicit evaluation than explicit evaluation and, vice versa, that the second proposition would have a bigger impact on explicit than on implicit evaluation. In line with this idea, our results showed that the type of relation implied in the pre-pairings instructions influenced implicit evaluation but not explicit evaluation whereas the effect of post-pairings instructions was significant for both implicit and explicit evaluation but bigger for explicit than for implicit evaluations. Another striking result was that implicit evaluations in the no instruction condition paralleled those in the equivalence-

equivalence condition but differed from those in the other conditions. This finding confirms the conclusion of Experiment 2 that equivalence instructions seem to add little to the effect of actual CS-US pairings.

General Discussion

Across three experiments, we examined the impact of relational information on EC of implicit evaluations. Experiment 1 provided a conceptual replication of the key study of Peters and Gawronski (2011, Study 3). Participants were presented with two nonwords (CSs), each paired six times with either a positive word (USpos) or a negative word (USneg). Either before or after the CS-US pairing, opposition instructions were presented that indicated that the nonword had a meaning that was the opposite of the word it was paired with. An IAT and rating measures were administered after the learning phase to capture implicit and explicit evaluations, respectively. Most importantly, results showed that the time at which opposition instructions were presented, influenced the implicit evaluation of the CSs. In particular, participants showed an implicit preference for the negatively paired stimuli (thus opposite to the valence implied by the mere co-occurrences) when opposition instructions were presented before the pairings, replicating Peters and Gawronski's findings in their short-delay condition. When opposition instructions were presented after the pairings, no implicit preference was found for one over the other CS. Therefore, confirming the conclusions of Peters and Gawronski's, (1) relational information influences implicit evaluations (2) but more so when the relational information is presented before rather than after the CS-US pairings.

Our results do not only confirm the robustness and generality of the findings of Peters and Gawronski, they also show for the first time that the observed effects of relational information are driven primarily by opposition instructions rather than equivalence instructions. In both Experiments 2 and 3, implicit evaluations in a baseline condition without equivalence or opposition instructions paralleled those in a condition with only equivalence instructions but differed from those in conditions with opposition instructions. Regardless of the mental representations that mediate the effect of relational information on implicit evaluations, our findings suggest that equivalence instructions add little to the effect of merely observing CS-US pairings. Opposition instructions, on the other hand, can overrule the effect of CS-US pairings when presented before or during the pairings or reduce the effect of CS-US pairings when presented after the CS-US pairings.

A second aim of this series of experiments was to test a post-hoc propositional explanation of the findings of Peters and Gawronski (2011). With Experiments 2 and 3 we tested the validity of two assumptions that must be made by propositional models in order to explain their results.

Experiment 2 tested the first assumption, being that the mere co-occurrence of a CS and a US is perceived as a cue that indicates the equivalence of the two stimuli. We aimed at changing this spontaneous framing by presenting participants with a pre-training phase that contained three training series of nonword-words pairs. Each of the three training series was followed by opposition instructions, in order to recreate an environment in which the co-occurrence of two stimuli would be interpreted as a cue that actually indicated that the two stimuli had an opposite meaning and, consequently, an opposite valence. Later, a test series of the actual CS-US pairs was presented, followed either by opposition instructions (condition Opposition instructions), equivalence instructions (condition Equivalence instructions) or no instructions (condition No instructions). Contrary to our predictions, the results of Experiment 2 showed a standard EC effect in conditions Equivalence instructions and No instructions, in spite of the relational framing implied by the context represented through the three training series. As in Experiment 1, Opposition instructions merely attenuated but did not significantly reverse the implicit preference of the CSpos over the CSneg. The results of Experiment 2 therefore do not provide support for a post-hoc propositional account of the results of Peters and Gawronski (2011). One could argue, however, that we did not succeed in our attempt to change the default interpretation of co-occurrence as a cue for equivalence. Perhaps the pre-training phase that aimed to change the default interpretation was not long enough or participants did not see the pre-training phase as relevant for the events during the actual CS-US pairings. Unfortunately, we did not have an independent way to verify whether our pre-training phase had the desired effect. Despite the fact that the results of Experiment 2 were inconclusive with regard to the validity of the post-hoc propositional account, the results of Experiment 2 again replicated the most important findings of Peters and Gawronksi (2011) and showed for the first time that equivalence instructions add little to the effect of CS-US pairings.

In Experiment 3, we aimed at testing a second assumption of the post-hoc propositional account, namely that in a context in which two conflicting propositions are generated, the initial proposition has a stronger impact on implicit evaluation than on explicit evaluation. To this end, we presented participants with CS-US pairings that were preceded and followed by instructions describing the relation occurring between each CS-US pair. The relation could either be of equivalence (i.e., the CS is a synonym of the US) or opposition (i.e., the CS is an antonym of the US), thus creating two conditions with consistent instructions (equivalence-equivalence and opposition-opposition) and two conditions with inconsistent instructions (equivalence-opposition and opposition-equivalence). In addition, we implemented a control condition in which no instructions were shown. After the acquisition phase, implicit and explicit evaluations were measured. The order of the measures was counterbalanced across participants, to allow for a comparison between implicit and explicit evaluation. In line with our predictions, initial instructions had a bigger impact on implicit than on explicit evaluations. Post-hoc tests showed that implicit evaluations were in line with the initial relational information whereas explicit evaluations were not in line with this information. Assuming that the two pieces of relational information (before and after pairings) led to two conflicting propositions, our results thus support the idea that, compared to explicit evaluations, implicit evaluations are influenced more by the initial proposition than by the proposition that is formed afterwards. This result is in line with the previous findings of Gregg et al. (2006), who found implicit evaluation to be affected only by the first information received, and, unlike explicit evaluation, to be unaffected by a conflicting second piece of information. However, our findings go beyond those of Gregg et al. in several ways. Most importantly, whereas Gregg always presented equivalence information as the first information, we also had conditions in which opposition instructions were presented first. Our results show that the first relational information has a bigger impact on implicit than on explicit information irrespective of whether the first information implied equivalence or opposition. Unlike Gregg et al., we can therefore conclude that the crucial variable is the order in which relational information is presented, rather than the type of relational information (equivalence or opposition) that is presented.

Our findings impose important constraints on mental process models of (changes in) evaluation. In line with the findings of Peters and Gawronski (2011), we found that (1) implicit evaluations are sensitive to relational information (2) but more so when the relational information is provided before the CS-US pairings than after the CS-US pairings. As we explained in the introduction, these findings force all existing mental process models into adopting certain core assumptions. Our findings impose at least two additional constraints on these models: (1) Equivalence information adds little to the effect of CS-US pairings whereas opposition instructions counteract the effect of pairings; (2) When two conflicting pieces of relational information are encountered one after the other, the initial piece of information will have more impact on implicit than on explicit evaluations, regardless of whether the initial information implies equivalence or opposition. We will now address how association formation models, propositional models, and dual process models can deal with the latter two findings.

Single process association formation models could accommodate the first conclusion if it is assumed that equivalence information allows for association formation whereas opposition information blocks association formation. However, such an explanation would not be able to explain the reverse EC effect after opposition instructions as, for instance, found in Experiment 1 (condition Before). This reversed effect could not simply be explained by an absence of association formation as such. It could, however, be explained if it is assumed that associations are formed in line with the mental encoding of events. Simple pairings or equivalence instructions would then lead to mental encoding in line with equivalence and thus to associations between the CS representations and US representation. Opposition pairings would lead to different a mental encoding and thus association between CS representation and the representation of a US which means the opposite of the presented US. However, such a process would also imply the existence of a propositional system that guides the mental encoding. With regard to the second conclusion (i.e., first information has more impact on implicit than explicit evaluation), one could argue that associations that are formed first have a bigger impact on implicit than on explicit evaluation. Regardless of the merits of these post-hoc assumptions, it is clear that our findings do not fit well with and severely constrain single process association formation models.

Dual process models can account for the first assumption in the same way as the single model-associative models. In fact, because dual process models allow for the existence of a propositional system, they fit well with the idea that CS-US associations are formed according to the valence implied by the mental (i.e., propositional) encoding of the events (see Gawronski & Bodenhausen, 2011). Regarding the second assumption, a dual process account could argue that the first relational information leads to a certain encoding of the pairings that results in association formation. In addition, one must assume that the second relational information presented has little or no effect at the level of associations. Thus, the initially formed associations are the basis of implicit evaluation, and implicit measures show effects in line with them. Conversely, the impact of the initially formed associations is smaller on explicit evaluation because it is based mainly on propositional processes that can change (and even reverse) in content and thus cancel or reverse the effect of the first relational information.

Finally, single process propositional models postulate that changes in evaluation are based only on the formation of mental propositions. These models can explain the fact that equivalence information adds little to the effects of CS-US co-occurrences whereas opposition information can counteract those effects. The crucial assumption in this context is that stimulus co-occurrences are by default (i.e., in the absence of conflicting relational information) interpreted as a cue that indicates similarity between the co-occurring stimuli. Hence, in the case of equivalence instructions, instructions and CS-US pairings lead to the similar proposition (e.g., CS is similar to the US). Equivalence instructions therefore add little to the effect of CS-US pairings. Opposition instructions, on the other hand, do lead to a different proposition (i.e., CS is opposite to US) than CS-US pairings (i.e., CS is similar to US). Depending on which proposition is formed first, opposition instructions can therefore counteract the effect of CS-US pairings. Regarding the second assumption, it is possible that the firstly formed proposition has a stronger influence on implicit than on explicit evaluation. This effect could be due to the fact that only the firstly formed proposition can be retrieved automatically (e.g., because it has been practiced more often), whereas propositions that are formed afterwards are less likely to be retrieved in an automatic manner. This hypothesis can be tested by manipulating the probability of automatic retrieval of a proposition independently from the time at which the proposition was formed.

To conclude, our results replicate and extend the intriguing findings of Peters and Gawronski (2011). In doing so, they impose important constraints on current and future models of EC and implicit evaluation.

References

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230-241. http://www.ucm.es/info/psi/docs/journal/

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

De Houwer, J., Hermans, D., & Eelen, P. (1998). Affective and identity priming with episodically associated stimuli, Cognition and emotion, 12, 145-169. doi: 10.1080/026999398379691

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

De Houwer, J. & Vandorpe, S. (2010). Using the Implicit Association Test as a measure of causal learning does not eliminate effects of rule learning. *Experimental Psychology, 57,* 2010, 61-67. doi: 10.1027/1618-3169/a000008

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2012). *A functional-cognitive framework for attitude research*. Manuscript submitted for publication.

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

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

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

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216. doi: 10.1037/0022-3514.85.2.197

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90*, 1-20. doi: 10.1037/0022-3514.90.1.1

Han, H. A., Czellar, S., Olson, M. A., & Fazio, R. H. (2010). Malleability of attitudes or malleability of the IAT? *Journal of Experimental Social Psychology*, *46*, 286-298. doi: 10.1016/j.jesp.2009.11.011

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.

Hermans, D., Baeyens, F., & Eelen, P. (2003). On the acquisition and activation of evaluative information in memory: The study of evaluative learning and affective priming combined (pp. 139-168). In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion*. Mahwah, NJ: Lawrence Erlbaum.

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

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorising in implicit attitude research: Propositional and behavioral alternatives. *Psychological Record, 61*, 465–496. http://thepsychologicalrecord.siuc.edu/

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y. & Sloman, S. A. (2007) Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*, (pp. 154–72). Oxford University Press.

Melchers, K. G., Lachnit, H., & Shanks, D. R. (2004). Within-compound associations in retrospective revaluation and in direct learning: A challenge for comparator theory. *Quarterly Journal of Experimental Psychology, 57B*, 25-53. doi: 10.1080/02724990344000042

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

Moran, T. & Bar-Anan, Y. (2012). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion,* 1-10. doi: 10.1080/02699931.2012.732040

Olson, M., & Fazio, R. H. (2004). Reducing the Influence of Extrapersonal Associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality & Social Psychology, 86(5),* 653-667, doi: 10.1037/0022-3514.86.5.653

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

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

Smith, C. T., De Houwer, J., & Nosek, B. A. (in press). Consider the source: Persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin*.

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

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

CHAPTER 4

THE IMPACT OF START AND STOP RELATIONS ON EVALUATIVE CONDITIONING EFFECTS OF IMPLICIT EVALUATION [1]

[1] This study was conducted in collaboration with Jan De Houwer and Anne Gast

Introduction

In the previous chapters, we have investigated the impact of relational information on evaluative conditioning effects (EC). EC is defined as the change in liking of a stimulus (named conditioned stimulus or CS) that is due to its pairing with another stimulus (named unconditioned stimulus or US; De Houwer, 2007; for reviews see De Houwer, 2007; De Houwer, Thomas, & Baeyens, 2001; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010; Levey & Martin, 1975). Relational information is contained by any type of cue that indicates how events in the world are related. In our work, we have focused on two types of relations: a relation of equivalence, which implies that two events share certain features, and a relation of opposition which implies that two events have opposing values on certain dimensions . A number of experiments reported in the previous chapters, as well as in the literature (Fiedler & Unkelbach, 2011; Förderer & Unkelbach, 2012; Peters & Gawronski, 2011; Zanon, De Houwer, & Gast, 2012) demonstrated that relational information can moderate EC, even when changes in implicit evaluation are registered.

Moran and Bar-Anan (2012), however, recently reported a result that diverges from the majority of the previous research. In two experiments, they showed that opposite relational information influenced explicit evaluation but not implicit evaluation. Their manipulation consisted of the presentation of four CSs (CS1, CS2, CS3, and CS4). Each CS was presented either at the beginning or at the end of the reproduction of a sound that could be pleasant (USpos) or aversive (USneg). Two of the four CSs co-occurred with a USpos (CSpos) and two co-occurred with a USneg (CSneg). More specifically, in their design, CS1 was always presented at the beginning of the USpos and CS2 at its end. In addition, CS3 was always presented at the beginning of the USneg and CS4 at its end. Importantly, the CSs that were presented before the sounds were said to cause the start of the sounds whereas the CSs that were presented after the sounds were said to stop the sounds. Hence, the instructions provided information about how the CSs and USs were related.

According to Moran and Bar-Anan, the implications of such a mix of co-occurrences and relational information for the transfer of valence between CSs and USs could be twofold depending on the impact of relational information. If relational information matters, then the changes in valence for the CSs that precede the USs would be opposite to the changes in valence for the CSs that followed the USs. More specifically, whereas the CS that start good sounds (CS1) should be liked more than the CS that start bad sounds (CS3), the CS that stops good sounds (CS2) should be liked less than the CS that stops bad sounds (CS4). If, however, relational information does not matter, both CSs that co-occur with the good sound (i.e., CS1 and CS3) should be liked more than the CSs that co-occur with the bad sound (i.e., CS2 and CS4) regardless of whether the CSs start or stop the sound. Importantly, both explicit and implicit evaluations of the four CSs were assessed. Implicit evaluations were capture either by a standard Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) or by a Sorting Paired Features task (SPF; Bar-Anan, Nosek; & Vianello, 2009).

The results of their explicit evaluation ratings showed that the CS that started a USpos (CS1) was liked more than the CS that started a USneg (CS3).In addition, the CS that ended a USneg (CS4) was liked more than the CS that ended a USpos (CS2). This result indeed reflects an impact of relational information on explicit evaluation, because it shows that, depending on the implied CS-US relation (start or stop), the CS-US pairings produced different explicit evaluation effects. The implicit measures, however, showed always a relative preference for the CSs that co-occurred with the USpos, regardless of whether it was presented at the start or the end of the US. Importantly, this standard EC effect of implicit evaluations (CSpos liked more than the CSneg) was found even for the stopping CSs. If relational information would have had an impact on implicit evaluation, one should have observed a moderation of the effect of the CS-US pairings on implicit evaluation by the type of relational information, that is, a different EC effect for the starting and stopping CSs.

Moran and Bar-Anan (2012) interpreted their results as strong support for the Associative-Propositional Evaluation model (APE, Gawronski & Bodenhausen, 2006). The APE model is a dual process model that postulates that implicit evaluation is based on associations in memory whereas explicit evaluation is based on a combination of associations and propositions. Associations are conceived of as simple links between mental representations that are formed after repeated pairings of stimuli whereas propositions are qualified links that hold a truth value and can also represent relational information. Because only propositions contain relational information and propositions influence only explicit evaluations, the APE model indeed predicts that relational information will impact only on explicit evaluations but not implicit evaluations, as was observed by Moran and Bar-Anan. In particular, according to the APE model, CS-US co-occurrences lead to the formation of mental associations regardless of whether participants were told that a CS started or stopped the US. These mental associations then determine implicit evaluations. Explicit evaluation, on the other hand, are assumed to be determined not only by these associations but also by propositions that do contain relational information.

Moran and Bar-Anan, however, did not completely rule out a purely propositional explanation of their findings. Propositional models (De Houwer, 2009; Mitchell, De Houwer & Lovibond, 2009) postulate that every source of information about the relation between stimuli can be encoded in propositions, including information about mere co-occurrences of stimuli. For instance, for CS2, participants might have formed both the proposition “CS2 co-occurred with the good sound” and the proposition “CS2 ended the good sound”. Likewise, for CS4, participants might have learned that “CS4 co-occurred with the bad sound” and “CS4 ended the bad sound”. If, for some reason, only the propositions about co-occurrence are retrieved during a measure, even a propositional model could explain that CS2 would be liked more than CS4. Therefore Moran and Bar-Anan’s results could be due to the effect of co-occurrence information that has been mentally represented in form of propositions, and not necessarily in form of associations. However, in order to take serious such a post-hoc propositional account, this account needs to make novel predictions that can be verified empirically. One such prediction is that interventions that increase the probability that participants (a) do from propositions that contain more than co-occurrence information (e.g., “CS2 stops a good sound”) and (b) do retrieve these propositions during the (implicit) measure, should increase the probability that (implicit) evaluations are determined by the information which specifies the type of relation between the stimuli (i.e., starting or stopping).

In the present chapter, we report a study that was inspired by this prediction. We examined whether an impact of stop/start relational information on implicit evaluation could be obtained with the paradigm used by Moran and Bar-Anan when their design is modified in two, in our opinion crucial, ways. First, it is possible that in the original study of Moran and Bar-Anan, participants were not optimally encouraged to encode relational information. In their pre-pairings instructions (see Appendix 1), Moran and Bar-Anan instructed the participants that they would hear a series of human and musical sounds. Furthermore, they added that, in order to clearly separate the reproduction of two consecutive sounds, also a ticking noise would be presented during the break in between the sounds. In addition, they instructed participants that also four families of creatures would be presented, parallel to the sounds. The families, which corresponded to the CSs, were described as differing from each other, by their color and head's shape. Each family was presented as having a specific role in stopping and starting the sounds, but in the final part of the instructions, the families were again described as simply appearing at the beginning or at the end of the sound fragments. Although these instructions do provide specific relational information about starting and stopping, they also emphasizes co-occurrence and might thus not convey the start/stop relational information in an optimal way. This conclusion is in line with the fact that even for the explicit evaluations, the reversed EC effect for CS2 and CS4 (i.e., the CSs that stopped the sounds) was much smaller than the standard EC effect for CS1 and CS3 (i.e., the CSs that started the sound). Although the presence of a reversed EC effect for CS2 and CS4 demonstrates that participants did encode the start/stop relational information to some extent, it is possible that the level or degree of encoding of this information was not sufficient to also influence implicit evaluations. For this reason, in our replication we employed slightly different instructions (see Appendix 2), in which the role and the responsibility of each CS in the starting and stopping of the USs were particularly and consistently emphasized. Unlike

Moran and Bar-Anan's, our instructions never referred to the CSs as co-occurring with the USs. Conversely, our instructions consistently stated that each family of creatures (CSs) determined the start and the end of each sound. This information was repeated several times, and emphasized using capital letters. This way, we aimed at clearly conveying the importance of the CS-US relation in the encoding of the CSs.

The second important modification of Moran and Bar-Anan's original design was the use of a personalized Implicit Association Test (personalized IAT; Olson & Fazio, 2004) as implicit measure of evaluation. In their two experiments, Moran and Bar-Anan employed a standard IAT (Greenwald et al., 1998) and a SPF (Bar-Anan et al., 2009) and found consistent results across the two implicit measures. The personalized IAT differs from the standard IAT (and also from the SPF) in that it requires the participants to give non-normative responses to the attributes presented. Positive and negative attributes have to be categorized using the labels I LIKE and I DO NOT LIKE, and no error feedback for the categorization of the attributes is presented. Hence, the personalized IAT encourages participants to respond according to their personal evaluations. Conversely, in a standard IAT and in an SPF attributes have to be categorized according to the normative labels POSITIVE and NEGATIVE, and an error signal appears in case of wrong categorization responses. Hence, the standard IAT requires responding in terms of what people in general like or dislike.

Han, Czellar, Olson, and Fazio (2009) have argued that the personalized IAT, compared to the standard IAT, is less sensitive to extrapersonal associations, that is, associations based on information that may be inconsistent with people's personal attitudes. Across three experiments, they found that the results of a standard IAT were susceptible to momentary contextual influences (induced by a previous, unrelated task) and that this influence disappeared when the labels were changed from normative (POSITIVE vs. NEGATIVE) to personalized (I LIKE vs. I DO NOT LIKE). In their Experiment 2, participants had been presented with a short scenario describing that, in a post nuclear world, flowers had actually become venomous (and therefore unpleasant) and insects had become the only edible items available (and therefore pleasant). In an IAT that followed this manipulation they found a preference for insects over flowers (in line with the post-nuclear scenario), whereas in an personalized IAT they found the typical preference for flowers over insects. Han and colleagues concluded that the results of the standard IAT can be particularly malleable, and thus can be easily distorted by unrelated information (e.g., a previous task). They also concluded that the personalization of the labels (i.e., the use of a personalized IAT) led to results that can be considered more in line with people’s personal attitudes.

Although Han and colleagues’ conclusions regard extrapersonal associations, we argue that it is also plausible to assume that in Moran and Bar-Anan’s paradigm the standard IAT may have simply picked up the effect of the information derived from the mere co-occurrences of stimuli. In other words, the IAT might have detected that participants knew that some CSs went together with good sounds whereas other CSs went together with bad sounds. If that is the case, then the IAT might not have detected the personal attitudes that participants had towards the CSs. For this reason, in our study, we used two personalized IATs (one for the two stopping CSs, and one for the two starting CSs), hoping that it would be more effective in also tapping in information deriving from the CS-US relations, and therefore providing an index of implicit evaluation that takes into account all learning instances.

A third difference with the original Moran and Bar-Anan’s design regards the order of the implicit and explicit measures. One of their goals was to examine potential dissociations in implicit and explicit evaluation, and therefore they presented their implicit and explicit measures in an order that was counterbalanced across participants, to allow for they comparison. In our study, our main focus was on whether relational information could reverse EC of implicit evaluation because such a result would demonstrate that relational information does influence implicit evaluation. To this end, we always presented participants first with the implicit measure (the personalized IAT), and then with the explicit measure (the valence ratings). In particular, the personalized IAT designed to test implicit evaluation for the stopping CSs was always presented first, followed by the personalized IAT for the starting CSs. Greenwald, Nosek, and Banaji (2003) argued that, when more than one IAT are presented in a row, the measurement effect in the first one of the series is larger, whereas those of the following IAT are smaller in magnitude. For this reason, we implemented this order to ensure an optimal measurement of the dependent variable in which we were most interested in, that is, the implicit evaluation of the oppositely related (stopping) CSs.

In sum, this experiment aimed at testing the moderation of relational information on implicit evaluation with a paradigm similar to the one used by Moran and Bar-Anan (2012). In our experiment, we adopted instructions that emphasize the salience of the CS-US relations (i.e., stop or start), and used another implicit measure, that is, the personalized IAT. We hypothesized that the start-CSpos would be implicitly liked more than the start-CSneg, but that implicit evaluation for the stop-CSs would be reversed (stop-CSneg liked more than the stop-CSpos).

Method

Participants

Forty-eight students at Ghent University participated. They either received course credits or four Euros (mean age = 20.70, *SD* = 3.38; 79% were women).

Materials

Acquisition phase. During the acquisition phase, sixteen color pictures of cartoon creatures (size 400 x 250 pixels) were employed as CSs, and presented on a white background. The creatures were said to belong to four families that differed in color (red, green, yellow and purple) and in the shape of their head (squared, rectangular, trapezoidal and circular). Each family was composed of four elements. In addition, the sound of a soft melody (presented in the instructions as a melody sound fragment) was employed as USpos, and the sound of a loud scream (presented in the instructions as a human sound fragment) was employed as USneg. Finally, as a cue that signaled the end of each trial, a ticking noise, was employed. All CSs, USs, and the ticking noises were identical to those used by Moran and Bar-Anan (2012). All sounds were briefly described in the initial instructions, and actually played during the acquisition phase. The assignment of the CSs to the USs was counterbalanced across participants, as well as the assignment of the CSs to their relation (start or stop). The result of the counterbalancing was that green and red creatures always had the same relational role, either stop or start, and differed only in terms of the US (positive or negative) that co-occurred with each of them. The same counterbalancing principle applied to the purple and yellow creatures (same relational information, different US contingency).

Personalized IAT. Two personalized IATs were presented, one that assessed the relative implicit valence of the stopping CSs, and one that assessed the relative implicit valence of the starting CSs. In the personalized IAT, the same color pictures employed as CSs in the acquisition phase were presented, and had to be categorized according to their color (green or red; yellow or purple). The target categories to be compared were those that held the same relational function. Therefore in one personalized IAT, purple and yellow creatures were compared, and in a second personalized IAT, green and red creatures were compared. For both personalized IATs, the likeable attributes were the Dutch words for “holiday” (VAKANTIE), “party” (FEEST), “present” (CADEAU) and “summer” (ZOMER). The dislikable attributes were the Dutch words for “war” (OORLOG), “vomit” (BRAAKSEL), “accident” (ONGELUK) and “divorce” (SCHEIDING). The target category labels were the Dutch words for “red creatures” (RODE WEZENS), “green creatures” (GROENE WEZENS), “yellow creatures”(GELE WEZENS), and “purple creatures” (PAARSE WEZENS). The labels for the attributes were the Dutch words for “I like” (HEB IK GRAAG) and “I do not like” (HEB IK NIET GRAAG). All attributes were presented in capital letters, black color, font Arial Black 48, on a white background.

Explicit valence ratings. Participants were asked to rate on a 9-point Likert scale how pleasant they found one random exemplar for each of the four families of creatures. Participants could respond by selecting a value ranging from 1 (very unpleasant) to 9 (very pleasant). This measure was identical to that used by Moran and Bar-Anan (2012).

Manipulation check. Two manipulation checks were added at the end of the experimental procedure. First, participants had to indicate, for an exemplar member for each of the four families, presented in random order, whether it started or stopped a sound. Second, participants had to indicate, for an exemplar member for each of the four families, presented in random order, whether they had co-occurred with a human sound fragment (i.e., the negative US) or with a melody sound fragment (i.e., the positive US). Again this was identical to the manipulation check used by Moran and Bar-Anan (2012).

Procedure

Participants were presented with an instruction page that explained their task. Unlike Moran and Bar-Anan's instructions (see Appendix 1), the instructions we employed (see Appendix 2) emphasized consistently the role of each of the families in starting and stopping the sounds. The creatures were described as responsible for starting or stopping the reproduction of a specific sound, and not simply as co-occurring with the beginning and end of the sound. This way, we wanted to convey the message that the creatures were actually actors that influenced the experience of the pleasant and of the aversive sounds. The sounds were described as being either a human sound fragment (USneg) or a musical sound fragment (USpos). In order to ensure attention in the acquisition phase, the instructions also warned the participants that their memory about the events of the acquisition phase was going to be tested later in the experiment and that, in case of an insufficient performance, the acquisition phase would have to be repeated.

After participants had read the instructions, they could proceed to the pairing phase by pressing a key. The pairing phase was identical to that used by Moran and Bar-Anan (2012) and consisted of a sequence of twenty trials. In each trial, a first creature (CS1 or CS3) appeared on the screen. After 500ms of silence, a sound (USpos or USneg) was played. The sound length could vary within 10000, 15000, 20000, 25000 or 30000 ms. Each positive and negative sound was presented twice in every length, in random order, making a total of ten positive and ten negative sound reproductions (i.e., trials).The first creature stayed on the screen for the first 2000 ms of the sound, after which it disappeared, leaving the screen blank. 2000 ms before the end of the sound, a second creature (CS2 or CS4) appeared on the screen, and remained on the screen for a total of 2500 ms, thus co-occurred with the sound for 2000ms and with silence for 500 ms. Between each trial, a ticking sound on a blank screen signaled the intertrial interval. The length of each intertrial interval was randomly determined, and could either last 5000, 10000, or 15000 ms. A first family of creatures would always start the positive sound (CS1), a second would always stop the positive sound (CS2), whereas a third would always start the negative sound (CS3) and a fourth would always stop the negative sound (CS4).

After all twenty pairings were presented, a short instructions page reminded the participants that in the previous phase their task was to learn and remember which families started or stopped the sounds. Then participants performed the first personalized IAT, in which always the stopping families of creatures were presented. After the first personalized IAT, participants performed a second personalized IAT, in which the starting families of creatures were presented. We decided to present the personalized IAT for the stopping creatures always first because the effect of the relation "CS stops US" was crucial in determining the impact of relational information on implicit evaluation, given that it implies an effect in the opposite direction than the mere co-occurrences. The personalized IAT consisted of seven blocks. The first block consisted in an attribute discrimination phase, in which only likeable and dislikeable attributes were presented and had to be categorized. The second block consisted in a target discrimination phase, in which members of the stopping (starting) families of creatures were presented and had to be categorized according to their color. The third and fourth phases were two combined blocks, in which both attributes and targets were presented and had to be categorized. The fifth block was a reversed target discrimination phase, in which only the target creatures were presented, but their assignment to the left and right key was reversed. Finally, in the sixth and seventh blocks attributes and targets were presented again, but targets had reversed key assignments, in line with those of Block 5. The order of the combined and reversed combined blocks was counterbalanced across participants.

After participants had completed the two personalized IATs, they performed the explicit valence ratings of four creatures, one per family, presented in random order. At the end of the experiment, participants answered two questions, which served as manipulation checks. They were presented with the picture of a randomly selected member of each family, and asked to indicate for each one whether it was starting or stopping a sound in the pairings phase. Later, they were also asked to indicate whether a randomly selected member for each of the four families of creatures had co-occurred with a human sound (the negative US) or with a musical sound (the positive US). Finally, participants were thanked and debriefed.

Results

Manipulation check

Ninety-two percent of the participants answered all eight questions of the manipulation check correctly. Results of both implicit and explicit evaluation did not change when excluding from the analyses the participants who failed the manipulation check.

Personalized IAT

Scores were calculated according to the *D4* scoring algorithm (Greenwald et al., 2003). High scores represented an implicit preference in line with the valence of the co-occurrences (CSpos liked more than the CSneg). The D score in the personalized IAT for the stop-CSs was not significantly different from zero (M = −0.01, SE = 0.63), $t(47) = -0.21$, $p = .83$, $d = 0.03$. This result shows that the personalized IAT did not detect an implicit preference for either of the stopping CSs over the other stopping CSs. The score in the personalized IAT for the start-CSs was positive, (M = 0.16, SE = 0.63) and significantly different from zero, $t(47) = 2.57$, $p = .013$, $d = 0.37$. This result indicates that the CSpos that started a positive sound was evaluated more positively than the CSneg that started a negative sound. The fact that the order of the two personalized IATs was not counterbalanced renders it problematic to make a statistical comparison between the two.

Explicit evaluation

A 2 (relation, start vs. stop) x 2 (valence, positive vs. negative) repeated measures ANOVA was conducted on the explicit valence ratings for the four CSs. We did not find a main effect of relation, $F(1,47) = 2.46$, $p = .12$, partial $\eta^2 = 0.05$. The main effect of valence, however, was significant, $F(1,47) = 41.61$, $p < .001$, partial $\eta^2 = 0.47$, and so was the interaction, $F(1,47) = 22.96$, $p < .001$, partial $\eta^2 = 0.33$. The CS that started a positive sound was evaluated more positively than the CS that started a negative sound, $t(47) = 6.53$, $p < .001$, $d = 0.94$. However, the CSs that stopped a positive or a negative sound did not differ significantly in explicit evaluation, $t(47) = -1.11$, $p = .27$, $d = 0.16$. Means of the explicit valence ratings can be found in Table 1.

Finally, the implicit and the explicit measures did not correlate (stopping CSs: $r = -.26$, $p = .07$; starting CSs: $r = .05$, $p = .75$). The scores of the two personalized IAT did not correlate with each other either ($r = -.03$, $p = .82$), but the explicit evaluations of starting and stopping CSs were negatively correlated ($r = -.53$, $p < .001$).

Table 1. *Means (and standard errors) of the explicit valence ratings for the four CSs.*

Relation	US valence	Explicit valence rating
Start	Positive	6.77 (0.32)
Start	Negative	3.21 (0.33)
Stop	Positive	5.15 (0.25)
Stop	Negative	5.60 (0.31)

General Discussion

In the present experiment, we investigated the effect of relational information on an implicit measure of evaluation in an experiment similar to the one reported by Moran and Bar-Anan (2012). In their original experiment, they presented four CSs, either at the beginning or at the end of the reproduction of a pleasant sound (USpos) or of an unpleasant sound (USneg). A CS1 was always presented at the start of the USpos, a CS2 was always presented at the end of the USpos, a CS3 was always presented at the start of the USneg, and a CS4 was always presented at the end of the USneg. Moran and Bar-Anan found no impact of relational information (i.e., start or stop) on implicit evaluation. In other words, the CSspos (the CSs that always co-occurred with the positive sound) were on an implicit measure always liked more than the CSsneg (the CSs that co-occurred with the negative sound), regardless of whether they started or stopped it (i.e., regardless of relational information). Conversely, explicit evaluation was moderated both by the information deriving from CS-US co-occurrences and by relational information. In other words, their explicit evaluation results showed that the CS that started a USpos (CS1) was liked more than the Cs that started a USneg (CS2). In addition, the CS that ended a USneg (CS3) was liked more than the CS that ended a USpos (CS4). In terms of the mental processes involved, Moran and Bar-Anan explained their results as a strong support for the APE model (Gawronski & Bodenhausen, 2006).

Moran and Bar-Anan's controversial result contradicts a number of other findings presented in this thesis (see Chapters 1 and 3), as well as from other literature (Peters & Gawronski, 2011), that have reported an influence, moderated by several factors (e.g., content, timing, modalities of presentation), of relational information on implicit evaluation. We argue that, if the relational information was optimally implemented, one would find a moderation on implicit evaluation even when employing Moran and Bar-Anan's (2012) procedure. In particular, we emphasize the importance of employing pre-pairings instructions that unequivocally describe the relations occurring between CSs and

USs (also see Chapter 3). In addition, we argue that the features of the implicit measure of evaluation employed could be another possible moderator of the effect.

For this reason, our experiment modified the one by Moran and Bar-Anan in three crucial respects. First, the importance of the CS-US relations was made more salient in our instructions. Second, in an attempt to tap into personal attitudes rather than into knowledge relating to the mere co-occurrence of stimuli, we used a personalized IAT instead of a standard IAT as an implicit measure (see Han et al., 2009). Finally, it is worth noting that, in order to distinguish between the effect of CS-US co-occurrences and relational information, the crucial measurement outcome would be that of the implicit measure comparing the stopping CSs. For this reason, in our procedure the personalized IAT for the stopping CSs was always presented before the one for the starting CSs (for a review on effects of IAT experience see Greenwald et al., 2003). Moreover, given that our main focus was on implicit evaluation, in our procedure we always presented the implicit measures before the explicit ratings. We expected to find both explicit and implicit evaluation to be moderated by relational information. In particular, with regard to implicit evaluation, we expected a preference in line with the co-occurrences (i.e., CSpos liked more than the CSneg) for the starting CSs and a reverse impact on implicit preferences (i.e., CSneg liked more than the CSpos) for the stopping CSs.

Our results showed that, for the start-CSs, the findings of Moran and Bar-Anan were replicated exactly, for both implicit and explicit evaluation: the CSpos was preferred over the CSneg. For the stop-CSs, however, no significant preference for either of the CSs was found, neither on implicit nor on explicit evaluation. Hence, the results for the critical condition (the implicit evaluation of the stopping CSs) did not show a significant preference for any of the CSs, making their interpretation more difficult. The fact that the personalized IAT showed a standard EC effect for the starting CSs may at least rule out the possibility that the implicit measure was not suitable to pick up any acquired preferences. Hence, the absence of the EC effect for the stopping CSs is probably due to the (opposite) relational information implied by the instructions and by the way they co-occurred with the CS (i.e., that the CSs stopped the USs). However, we cannot be certain whether the difference to the results by Moran and Bar-Anan was due to either the different relational instructions or to the use of the personalized IAT. On the one hand, it is plausible that our instructions may have conveyed the relation of CS and US more clearly and hence led to an implicit evaluation that is actually moderated by the relational information. Note, however, that we did not find a reversal in the explicit ratings of the stop-CSs whereas Moran and Bar-Anan did find this reversal. Hence, it seems unlikely that our instructions indeed conveyed the relational information more clearly than those of Moran and Bar-Anan. On the other hand, it is also possible that the use of a different implicit measure (the personalized IAT instead of the standard IAT/SPF) has led to this different result. Han and colleagues (2009) argue that the personalized IAT is less sensitive to extrapersonal associations (e.g., its measurement outcomes are less influenced by unrelated learning tasks). Building on this assumption, it could be argued that the personalized IAT is therefore more sensitive to the actual evaluations of the CSs, which are influenced by start/stop relational information. However, in order to sustain this assumption, more research is needed. More specifically, we plan to conduct a new experiment in which we use both a standard IAT and a personalized IAT to measure evaluations of the stop-CSs. The prediction is that the standard IAT will reveal more positive attitudes towards the stop-USpos CS than toward the stop-USneg CS (i.e., a standard EC effect) whereas the personalized IAT would reveal no or a reversed effect.

Assuming that our personalized IAT results do reflect a similar evaluation of the two stop-CSs, it would imply that implicit evaluations are to some extent determined by relational information, namely to the extent that it can counteract the effect of CS-US co-occurrences. At the level of mental processes, this suggests that propositions do mediate implicit evaluation (but see our earlier chapters for caveats regarding this statement). It is not clear, however, whether only propositions mediate implicit evaluation or whether both propositions and associations in memory play a role. Although it will be difficult to distinguish between these two positions empirically, future research on the impact or relational information on implicit evaluation will continue to reveal new facts about implicit evaluation which will further constrain theories about the mental processes that mediate implicit evaluation.

References

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task: A measure of association strengths. *Experimental Psychology, 56*, 329-343. doi: 10.1027/1618-3169.56.5.329

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230-241. http://www.ucm.es/info/psi/docs/journal/

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

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

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

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

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

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216. doi: 10.1037/0022-3514.85.2.197

Han, H. A., Czellar, S., Olson, M. A., & Fazio, R. H. (2010). Malleability of attitudes or malleability of the IAT? *Journal of Experimental Social Psychology*, *46*, 286-298. doi: 10.1016/j.jesp.2009.11.011

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

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

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

Moran, T. & Bar-Anan, Y. (2012). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion,* 1-10. doi: 10.1080/02699931.2012.732040

Olson, M., & Fazio, R. H. (2004). Reducing the Influence of Extrapersonal Associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality & Social Psychology, 86,* 653-667, doi: 10.1037/0022-3514.86.5.653

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

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

Appendix 1

In the first part of this study you are going to listen to two kinds of sound segments:

1. Human sound segment–vocal sound.
2. Musical sound segment–melody sound.

The two different segments will be repeated few times in different lengths. In order to make it easy to understand when each segment starts and stops, you will hear a ticking noise that indicates the break between two segments.

Parallel to the sound segments, four kinds of creatures will appear on the screen. Each type of creature will have one of four roles:

1. After the appearance of one kind of the creature, the musical sound segment will start.
2. After the appearance of one kind of the creature, the musical sound segment will be stopped.
3. After the appearance of one kind of the creatures, the human sound segment will start.
4. After the appearance of one kind of the creatures, the human sound segment will be stopped.

Your task is to learn what the permanent role of each type of the creatures is; at the end of this task your memory will be checked.

The four types of creatures that you are going to see are:

(pictures of the 16 creatures)

The study is about to start. Remember, your task is to learn what the permanent role of each type of creature is:

1. Begin the musical sound segment.
2. Stop the musical sound segment.
3. Begin the human sound segment.
4. Stop the human sound segment.

In the end of the study we will check what you have learned.

Appendix 2

In the first part of this research you will hear two different sound clips:

1. Human sound–vocal sound
2. Musical sound–melodic sound

These two different fragments will be presented a number of times in different lengths.

To clarify when each sound fragment starts and stops, a ticking sound will be played between the fragments. This ticking sound indicates a break between two consecutive fragments.

The appearance and disappearance of the human and musical sound WILL BE DETERMINED by four different families of creatures.

These four different families differ in color and shape of their head. You will see the Green Rectangles, the Purple Circles, the Red Squares and the Yellow Trapezoids.

Each creature will EITHER END OR START the reproduction of a sound.

More specifically, you will encounter the following situations:

1. A certain type of creature STARTS the musical fragment
2. A certain type of creature STOPS the musical fragment
3. A certain type of creature STARTS the human sound fragment
4. A certain type of creature STOPS the human sound fragment

Your task is to learn which specific type of creature is responsible for STARTING and STOPPING the human and the musical sound fragment. At the end of this task your memory will be tested.

The study will begin immediately. Remember, it is your task to learn which family of creatures (Green, Purple, Red, Yellow) STARTS or STOPS a specific sound (Human or Musical).

Please be therefore very attentive. If you fail on the memory task, you will have to repeat the entire learning phase.

GENERAL DISCUSSION

Implicit evaluation refers to those instances of evaluation that occur under conditions of automaticity (De Houwer, 2009a). Researchers have acknowledged the importance of implicit evaluation in predicting and changing behavior, and its role has been investigated across various domains (for an overview, see Gawronski & Payne, 2010). Besides documenting moderators that influence implicit evaluation, researchers have also tried to define the cognitive mediators of implicit evaluation. Surprisingly, there has been only one dominating mental process account of implicit evaluation ever since researchers directed their attention to this phenomenon (see Hughes, Barnes-Holmes & De Houwer, 2011, for a review). The longstanding assumption was that implicit evaluation is a product of mental associations. Associations are unqualified links between mental representations. Associations are assumed to be formed after the experience of pairings of stimuli. Pairings of stimuli have repeatedly been shown to cause a change in liking, an effect that is named evaluative conditioning (EC; for reviews see De Houwer, 2007; De Houwer, Thomas, & Baeyens, 2001; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). So, according to an associative view, if a neutral stimulus (conditioned stimulus, CS) is repeatedly paired with a valenced stimulus (unconditioned stimulus, US), individuals would form an association between the mental representations of the CS and the US. Later, when encountering the CS, the CS-US association would be excited automatically, which activates the US representation and with it the valence of the US. As a result, the implicit evaluation of the CS changes.

Although the associative conceptualization of implicit evaluation has served psychological science for many years and has given rise to many process models (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Fazio, 2007; Olson & Fazio, 2009), we argue that research could profit from an alternative approach to the study of implicit evaluation. A main reason why this shift in conceptualization is important is that, given the unqualified nature of the associative links, it is clear that they have limitations in representing more complex relations that could exist in the environment. It is evident that the way stimuli are related is an important element that, if taken into account, can completely change the way in which people respond to a stimulus. For example, imagine a substance that co-occurs with a disease: the mere co-occurrence of the substance and the disease does not tell us much about the properties of that specific substance. However, knowing that the substance causes, or prevents, the disease is a piece of information that becomes crucial in order to form a positive or negative evaluation of it. In order to be able to mentally represent relations such as, for example, "X causes Y", or "X prevents Y", one needs a conceptual construct that holds qualifying properties.

To this end, propositional accounts (De Houwer, 2009b; Mitchell, De Houwer, & Lovibond, 2009) postulate that every instance of associative learning is mediated by the formation of propositions, thus refuting the relevance of associations. Propositions are qualified links or statements about the world that possess a truth value. Therefore, propositions are mental constructs whose properties allow for the representation of any kind of relation between stimuli. So far, their importance in the acquisition of explicit evaluation has been recognized not only by propositional models, but also by dual process models (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006). However, dual process models typically still assume that implicit evaluation is exclusively the product of the automatic activation of associative links. In contrast, propositional models argue that also propositions can be activated automatically, and therefore can account for implicit evaluation without the need of employing the construct of associations. Our main research question stemmed from the inspiration provided by the propositional account of implicit evaluation. Using procedures from evaluative conditioning research (i.e., stimulus pairings between neutral and valenced stimuli), the aim of this dissertation was to investigate whether and to which extent relational information (i.e., the relation between the CS and the US) impacts implicit evaluation.

Overview of the studies and findings

In **Chapter 1**, we tested the assumption that the impact of stimulus pairings on implicit evaluation can be moderated by a context consisting of other pairings that imply a certain rule. Across three experiments, we tested this assumption by employing an implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998), an affective priming procedure (Fazio, Jackson, Dunton, & Williams, 1995), and a personalized IAT (Olson & Fazio, 2004), as implicit measures of evaluation.

In condition Opposite we operationalized the context by presenting Apos (cue A followed by a positive outcome), Bpos, ABneg (compound of cues A and B followed by a negative outcome), Cneg, Dneg, and CDpos. In condition Same, we presented Apos, Bpos, Abpos, Cneg, Dneg, and CDneg. In condition Opposite, the context rule implies that a single stimulus (e.g., A) is followed by a certain outcome (e.g., in the pairing Apos) when on its own, and by the opposite outcome when presented in compound with a second cue (e.g., in the pairing ABneg). Conversely, in condition Same a single cue was always followed by the same outcome, regardless of whether it was presented alone or in compound with another cue (e.g., in the pairings Apos and ABpos). The aim of this procedure was to create two contexts of stimulus pairings, each of which followed a different rule. After the context pairings, in both conditions the target pairings XFpos and YHneg were presented. Later, implicit and explicit evaluation for the single cues X and Y (our CSs) were measured. Note that the CSs X and Y were never presented on their own during the acquisition phase, but always in compound with other cues (F and H, respectively). We hypothesized that the relational information derived from the contexts would have an impact on the encoding of the pairings of X and Y, and that we would have observed a difference in evaluation between conditions Same and Opposite. Our results showed that the different contexts significantly moderated implicit evaluation. While in condition Same the positively paired cue (CSpos, X) was implicitly liked more than the negatively paired cue (CSneg, Y), we found no implicit preference for any of the cues in condition Opposite. We observe that this difference can be only due to the characteristics of the context pairings, that were the only elements that differed across conditions. Regardless of the mental mediators of this effect, we argue that relational information, specifically in form of contextual cues, moderates implicit evaluation.

In **Chapter 2**, we tested whether the type of implicit measure might moderate the impact of relational information on implicit evaluation. Different implicit measures are tasks that create different situations to which individuals have to respond in many ways. A measurement outcome is then calculated, based on the participant's performance on the task. Therefore, different measurement outcomes may be more or less sensitive to relational information, depending on the features of the situation that the implicit measure creates. In particular, in Chapter 2 we were interested in testing the sensitivity to relational information of three different implicit measures of evaluation: the IAT (Greenwald et al., 1998), the Affect Misattribution Procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005), and the evaluative priming task with naming responses (Spruyt, Hermans, De Houwer, Vandekerckhove, & Eelen, 2006). Experiments 1-3 employed the same design, and differed only in terms of the implicit measure of evaluation that was administered. Similarly to the experiments reported in Chapter 1, we adopted a procedure that contained stimulus pairings and relational information. In the initial instructions, a stimulus R was described as a special cue that had the function of reversing the outcome of the cue it was paired with. Therefore, if a certain cue was always followed by a negative outcome when presented alone (e.g., Bneg), it would be followed by the opposite outcome (e.g., BRpos) when appearing in compound with the cue R. In addition to the verbal instructions, four context pairings (Apos, ARneg, Bneg, and BRpos) were repeatedly presented, with the intent of showing the role of the cue R in action and clarifying its function. In addition, two target CSs, X and Y, were consistently presented in compound with the cue R, and followed by a positive or by a negative outcome (USs), respectively, resulting in the pairings XRpos and YRneg. Importantly, the cues X and Y were never presented on their own, but always in compound with the cue R (i.e., XR, YR). Clearly, context relational information predicts an outcome for the CSs that is opposite to that indicated by their mere co-occurrence with the USs. After the acquisition phase, participants had to indicate for the single cues X and Y whether they would have been followed by a positive or by a negative outcome if encountered again on their own. Later, participants performed an implicit measure (IAT in Experiment 1, AMP in Experiments 2 and 3), designed to measure implicit evaluation for the single target cues X (CSpos) and Y (CSneg). The results of the IAT showed a reversed EC effect of implicit evaluation, that is, a preference for the CSneg over the CSpos. This result shows an impact of relational information on implicit evaluation, and, given the direction of the effect, the magnitude of this impact is bigger than that of the mere CS-US co-occurrences. However, the results of the AMP diverged from those of the IAT, and showed only a null effect, that is, an absence of EC effect for any of the CSs. Across the two measures, the same manipulation resulted in different implicit measurement outcomes: whereas the IAT showed an effect in line with relational information, the AMP did not show any effect.

Finally, with Experiment 4 we aimed at testing the sensitivity of the evaluative priming with naming responses to relational information. Previous research (Spruyt et al., 2007) suggested that results in this implicit measure may be particularly sensitive to effects of spreading of activation of CS-US associations. Therefore, effects in this naming task might be mostly insensitive to the impact of relational information, that is, information that cannot be represented by mere associations. Experiment 4 employed the procedure of the long-delay condition of Experiment 3 of Peters and Gawronski (2011). Pictures of four male individuals A, B, C, and D (CSs) were paired with either positive or negative behavioral statements as USs, resulting in the pairings Apos, Bpos, Cneg, and Dneg. After these pairings were presented, participants were told that the information provided about individuals A and C was true (equivalence relation) and that about individuals B and D was false, and its implications had to be reversed (opposite relation). Therefore, this design produced four different CS conditions (positive-true, positive-false, negative-true, negative-false). After the acquisition phase, implicit and explicit evaluation were measured. As implicit measure we used an evaluative priming procedure with naming responses. Results of the implicit measure did not show a preference for any of the four CSs. Importantly, even for the truly-related CSs we did not find a preference, suggesting that the picture-picture naming task might have failed at tapping into any type of implicit evaluation of the CSs.

The aim of **Chapter 3** was to test whether relational information in form of verbal instructions, and the time at which the instructions are presented, moderates implicit evaluation. In Experiment 1, we conducted a conceptual replication of the studies of Peters and Gawronski (2011). Participants were presented with two nonword CSs, each of which was repeatedly paired either with a word positive in valence (USpos) or with a word negative in valence (USneg). Participants in condition Instructions Before were instructed, before the CS-US pairings, that each CS was a foreign word opposite in meaning to its paired US (opposite relation). Participants in condition Instructions After received the identical instructions, but only after all the CS-US pairings were presented. Later, implicit evaluation of the two CSs was assessed, employing an IAT. Interestingly, participants in condition Instructions Before showed a reversed EC effect of implicit evaluation, that is, an implicit preference for the CSneg over the CSpos. On the other hand, participants in condition Instructions After did not show a preference for any of the CSs, that is, they did not show any EC effect. Evidently, the point in time at which opposite relational information was made available moderated the implicit evaluation of the CSs, thus replicating the results of Peters and Gawronski (2011). In addition, whereas Peters and Gawronski presented relational information after each pairing in their short-delay condition, we obtained the same reversed EC effect employing only one presentation of the relational information, before the CS-US pairing series.

In Experiment 2 of Chapter 3, we aimed at testing whether the simple co-occurrence of a CS with a US already represents a cue that indicates a relation of equivalence between the two. To this end, we tried to change this spontaneous framing. Participants experienced three training series of pairings, in which two CSs were paired with either a USpos or with a USneg. Each series was followed by instructions describing a relation of opposition between each CS and its paired US. The aim of this manipulation was to create a context that encouraged participants to encode co-occurring stimuli as being opposite to each other. After the training series, a target CS-US series of pairings was presented. In order to test the resilience of the newly established framing of opposition to post-pairings instructions this series was followed either by equivalence instructions (describing equivalence between each CS and its paired US), by opposite instructions (describing opposition between each CS and its paired US) or by no instructions. Later, implicit evaluation for the CSs was assessed, adopting a standard IAT and a personalized IAT. The results showed that our manipulation was ineffective in changing the spontaneous framing of equivalence deriving from mere CS-US co-occurrences. However, the post-pairings instructions significantly moderated implicit evaluation of the CSs. On the implicit measures both conditions Equivalence Instructions and No Instructions showed a standard EC effect (a preference for the CSpos over the CSneg), whereas participants in condition Opposite Instructions showed an absence of preference for any of the CSs. On the one hand, these results show that we did not succeed in changing the spontaneous framing of co-occurring CSs and USs from equivalence to opposition. On the other hand, they show that post-pairings instructions that imply equivalence between CSs and USs do not add anything to the standard EC effect, that is obtained even without relational instructions, whereas opposite instructions can attenuate it until its complete dissolution, as shown by a series of contrast effect tests.

Finally, with Experiment 3 of Chapter 3, we aimed at testing the post-hoc assumption that, in a situation in which several contrasting relations are experienced, implicit evaluation is more strongly influenced by the first proposition about CS-US relations, whereas explicit evaluation is not. This would allow us to explain recent patterns of results (e.g., Peters & Gawronski, 2011) in light of a propositional account, and it would also shed light on the determinants of implicit and explicit evaluation. To this end, we presented participants with a series of CS-US pairings. In addition, participants were presented with verbal instructions describing the relation between the CSs and their paired USs. Participants received instructions both before and after the stimulus pairings.

These relations could be of equivalence (each CS is equivalent to its paired US) or opposition (each CS is opposite to its paired US). Therefore, in two conditions the relational information presented before and after the pairings was congruent (equivalence-equivalence, and opposition-opposition). Conversely, in two other conditions this information was incongruent (equivalence-opposition and opposition-equivalence). In the case of incongruent relations, the last one that was presented was indicated to be the valid one. In addition, a control condition in which no instructions were presented was also added to the design. The order of implicit and explicit measures of evaluation of the two CSs was counterbalanced across participants, allowing for a direct comparison of the two types of evaluation. Results showed that implicit evaluation was moderated by the content of the first relational instructions presented. On the contrary, explicit evaluation was not moderated by the first, but only by the second piece of relational information that was experienced. Therefore, the results confirmed our predictions.

Finally, **Chapter 4** consisted in a replication of a study reported by Moran and Bar-Anan (2012). In their study, they realized four CS conditions (CS1, CS2, CS3, and CS4). At the beginning of the experiment, participants were instructed that each CS had a specific role in the reproduction of one of two sounds. The CS1 always started the reproduction of a pleasant sound (USpos), and the CS2 always stopped its reproduction. The CS3 always started the reproduction of an aversive sound (USneg), and the CS4 always stopped it. Therefore, two CSs (CS1 and CS2) always co-occurred with an USpos, and two other CSs (CS3 and CS4) always co-occurred with a USneg. After this acquisition phase, Moran and Bar-Anan measured implicit and explicit evaluation of the four CSs, adopting as implicit measures an IAT (Greenwald et al., 1998) and a Sorting Paired Features task (SPF; Bar-Anan, Nosek, & Vianello, 2009). Interestingly, their results showed that relational information (i.e., the start/stop relations) did not moderate in any extent implicit evaluation. The CSpos were always evaluated more positively than the CSneg, regardless of whether they started or stopped the correspondent US.

In our replication, we added three, in our opinion important, modifications to the original design of Moran and Bar-Anan. First, in our initial instructions we emphasized the role of each CS in starting or stopping the experience of the USs, in order for them to be encoded as active agents in the modulation of each US experience. Second, we employed a series of two personalized IATs as implicit measures of evaluation, because it is possible that the personalized IAT is more sensitive to the implications of relational information (see Han, Czellar, Olson, and Fazio, 2009, for a set of studies on the effect of the personalization of the labels on IAT outcomes). Finally, unlike Moran and Bar-Anan we always presented the implicit measure of evaluation of the oppositely related CSs (CS2 and CS4) first. This procedure was adopted to ensure an optimal measurement of the crucial dependent variable of our manipulation, that is, the implicit evaluation in the presence of opposite relational information (see Greenwald, Nosek, and Banaji, 2003, for an in-depth analysis of the use of multiple IATs). In line with Moran and Bar-Anan's results, we observed a standard EC effect of implicit evaluation for the CSs equivalently related to the USs (the staring CSs). However, contrary to Moran and Bar-Anan, for the CSs that were oppositely related to the USs (the stopping CSs), no preference for any of the CSs was observed. This difference between conditions shows that relational information can moderate implicit evaluations also in the design by Moran and Bar-Anan, although without leading to a reversed EC effect.

To sum up, the experiments reported in the four empirical chapters investigate, across several aspects, how and when relational information can impact implicit evaluation. In the following section, we will summarize the most important findings and explain their implications for mental process models of evaluation.

Implications of our findings

As mentioned in the Introduction, the impact of relational information on implicit evaluation has been studied only recently (Peters & Gawronski, 2011; Moran & Bar-Anan, 2012; Zanon, De Houwer & Gast, 2012). Therefore, ideas about the mental mediators of these effects are still in a relatively early stage. Our findings have relevant implications for theories about these mediators. Across all experiments, we have manipulated relational information between the CSs and the USs by realizing two basic relations: equivalence and opposition. These relations have been implemented in different ways, employing context cues (Chapter 1; Chapter 2, Experiments 1-3), verbal instructions (Chapter 2, Experiment4; Chapter 3), or both (Chapter 4). Across all experiments, we have always kept the CS-US co-occurrences constant, manipulating the relations occurring between CSs and USs and assessing implicit evaluation of the CSs. In order for us to be able to investigate the impact of relations on implicit evaluation, the analysis of the effect of opposite relational information, and its comparison with that of equivalence information, are the crucial ones. That is, whereas the implications of equivalence instructions are in line with those of CS-US co-occurrences (and therefore do not allow one to distinguish between the separate impact of each source of information), the implications of opposite relational information are, by definition, contrary to those of the mere CS-US co-occurrences. Hence, given that the CS-US pairings are always kept constant, differences between conditions could tell us whether the content of the relational information moderated the acquisition of implicit evaluation of the CSs.

Most importantly, in several experiments, we observed that relational information did influence implicit evaluation. Our first important finding, that models must try to accommodate, is that opposite relational information can reduce (when presented after the CS-US pairings) or even reverse (when presented before or during the CS-US pairings) standard EC effects of implicit evaluation. Specifically, we have shown this effect both for relational information presented in form of a contextual rule (Chapter 1 and 2) and in form of verbal instructions (Chapter 3 and 4). A second important finding (reported in Chapter 3) is that the first experienced relational information has more impact on implicit than on explicit evaluation.

In general, these results are difficult to explain adopting a purely associative account of implicit evaluation. Single process associative models (e.g., Baeyens, et al., 1992; Fazio, 2007; Olson & Fazio, 2009) assume that the CS-US co-occurrences are encoded in form of mental associations, and that implicit evaluation is mediated exclusively by these associations. However, in our studies we have always kept the CS-US co-occurrences equal, and manipulated the content of the relations between CSs and USs, observing different effects on the implicit evaluation measures. Yet, given these conditions, an associative account would predict that the same CS-US mental associations would be always formed, and that the same implicit evaluations would be always observed, regardless of the content of the relational information (i.e., equivalence or opposition). Given our results, at a first glance purely associative models therefore seem inadequate to explain our findings. However, single process associative models could account for attenuations of EC effects due to the impact of (opposite) relational information if they assume that opposite relational information does block the process of CS-US association formation. Nevertheless, this assumption could explain an absence of EC effect, but could not account for the reported reversed EC effects. According to associative models, implicit evaluation must be based exclusively on associations, and on a reversed EC effect these associations must be formed between the mental representations of the CSs and the opposite of the USs. One could argue that participants may form these associations at the time of the encoding of the CS-US pairings, and therefore that the effect of stimulus pairings on associations depends on the interpretation of the experienced events. A mental representation, opposite to the physical US encountered, would be formed and associated to the mental representation of the CS. For instance, the pairing of the nonword BAYRAM with the word HAPPY in a context that implies a relation of opposition, could be mentally represented as a pairing of the BAYRAM and SAD, and hence lead to an association between the representations of BAYRAM and SAD in memory. Although such an account is feasible, it implies the existence of a propositional system that allows for the mental interpretation of events on the basis of relational information. As such, it would effectively transform the single process associative account into a dual process account. Consequently, this would rule out the hypothesis of a purely associative explanation of reversed EC effects of implicit evaluation. With regard to the assumption on the timing moderation on the impact of relational information on implicit and explicit evaluation, an associative account could argue that the first associations are greater in strength than secondly formed associations, and therefore are more effective in influencing implicit evaluation. This explanation suffers from the same issues of the first explanation: in case of opposite relational information presented first, a purely associative model would require to refer to propositional processes in order to explain the opposite association formation. To sum up, it is evident that purely associative models are severely constrained by the findings reported in the present dissertation. That is, in order to maintain the idea that associations always mediate implicit evaluations, very specific supplementary assumptions have to be made about the factors that influence the formation and activation of those associations.

Dual process models (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006) traditionally account for effects on implicit evaluation in a similar fashion as purely associative models. However, by postulating the existence of a propositional system, dual process models allow for association formation that is mediated by the (propositional) encoding of the CS-US pairings (i.e., taking into account relational information). Therefore, most recent dual process models (e.g., Gawronski & Bodenhausen, 2011) could account for reversed EC effects of implicit evaluation, assuming that people can form associations between the mental representation of the CSs and the mental representation of the opposite of the physically experienced US (formed via propositional processes at the time of encoding). In addition, they can explain the absence of EC effects when relational information is presented after the stimulus pairings by assuming that the CS-US associations are counteracted by propositional processes, formed on the basis of the (opposite) relational information. This mechanism would also explain the stronger impact of the first relational information on implicit evaluation than on explicit evaluation. The latter, based on propositional processes, could be more easily corrected after a second, valid, piece of relational information is made available. Conversely, the associations that guide implicit evaluation, once established could only be blocked by, incongruent, second information, and not reversed.

Finally, purely propositional processes (e.g., De Houwer, 2009b; Mitchell, De Houwer, & Lovibond, 2009) postulate that all changes in evaluations are based on propositions. Therefore, both information derived from CS-US co-occurrences and relational information are encoded in propositional format. It is therefore possible that multiple, even incongruent, propositions about the same CS are formed. Different effects of relational information on implicit evaluation could therefore be explained by the conflict of these incongruent propositions. The reversed EC effects of implicit evaluation, obtained when opposite relational information is presented before or during the stimulus pairings, would thus imply that the relational proposition has a stronger impact than the co-occurrence proposition. Conversely, the absence of EC effects when relational information is available only after the stimulus pairings would suggest that the two propositions have the same impact, and that they therefore cancel each other out. In addition, in order to address our second finding, a propositional model would have to assume that the proposition formed on the basis of the first relational information has a stronger effect on implicit than on explicit evaluation. Propositions formed afterwards would then be able to influence explicit evaluation but not implicit evaluation.

Limitations and future directions

In order to define possible future directions of research on this topic, it is important to note that, in the present dissertation, we report a series of functional findings, inspired by the functional-cognitive framework of De Houwer et al. (2012). According to this framework, the elements of the environment are potential moderators of evaluation, that is conceived as an effect. Consequently, stimuli that convey relational information can be functionally conceived as a type of moderator. We have shown that relational information can moderate implicit evaluation in a substantial extent. Therefore, our findings contribute to the body of knowledge on the possible moderators of implicit evaluation. At the same time, our functional findings constrain existing mental process accounts, which are responsible for defining the (mental) mediators of evaluation effects. Given the importance of evaluation in explaining and predicting behavior, future research should aim at extending our knowledge of the moderators of (implicit) evaluation. Psychological research would profit from such an approach, because the discovery of the functional moderators of evaluation would allow for new psychological theories to arise, and for the existing ones to adapt their assumptions in order to fit with the empirical evidence. Although we are still far from discarding existing accounts of the mental mediators of evaluation, it is evident that this line of research would lead to the improvements of the current mental process models, and consequently to a better explanation and prediction of human behavior.

In the experiments reported in this dissertation, we have consistently focused on the relations of equivalence and opposition. However, these are only two of the vast range of relations that could occur between stimuli in the environment (see Hayes, Barnes-Holmes & Roche, 2001). To name a few, there are also comparative (i.e.,” more than / less than”), hierarchic (i.e.,” a part of”), or temporal (i.e., “before / after”) relations. As the moderation of relations of oppositions and equivalence on implicit evaluation has been demonstrated in several experiments, it is reasonable to think that also other relations may serve as moderators for implicit evaluation. For instance, Molet, Macquet, and Charley (in press) have examined the impact of comparative relations on explicit evaluation. They presented three cues that differed only in size (small, medium, and big) and found that when the medium-sized cue was said to be paired with a USneg (a shock), the small-sized cue was explicitly evaluated more positively than the big-sized cue. However, when the medium-sized cue was said to be paired with a USpos (an amount of money), the big-sized cue was explicitly evaluated more positively than the small-size cue. Even though Molet et al. examined only explicit evaluation, these results are promising and show that evaluation can be moderated also by relational information other than equivalence and opposition. Potentially, also the moderation of other relations could be investigated. In any case, it would be very interesting to explore whether effects similar to those of Molet et al. could be obtained also on implicit evaluations.

Another possible moderator, that we have started to explore in Chapter 3, is the effect of the conditions under which relational information is made available. In line with the findings of Peters and Gawronski (2011), our results show that the moderation of the timing in which relational information is presented is particularly important, in that it can lead to very different implicit evaluations. In this class of moderators, another possible option for future studies could be the distinction between instructed and experienced relational information. For instance, one could compare the effects of relations experienced without the use of verbal instructions (e.g., through a matching-to-sample procedure, Hayes et al., 2001) to that of purely instructed relations (e.g., via verbal instructions). Another interesting property to explore would be the number of repetitions of pairings and relational information. for example, one could present participants with opposite relational information, followed by a small number of CS-Us pairings. Our previous findings have shown that, at this point in time, implicit evaluation is likely to be in line with relational information. However, it would be interesting to see whether implicit evaluation changes if many more CS-US pairings are presented. It would be possible that the effect of relational information becomes stronger, keeps equal or disappears[1].

In Chapter 4, we have seen how important consistency of relational information is in order to obtain a moderating effect on implicit evaluation. When participants were asked to focus on both relations and co-occurrence information (see Moran & Bar-Anan, 2012), relational information showed no moderating effect on implicit evaluation. However, our results of Chapter 4 may also suggest that, in order for relational information to be effective, participants must consistently focus on the relations occurring between stimuli, and not on the mere co-occurrence of the stimuli. One interesting experiment would be to present two groups with the same CS-US pairings and opposite relational information and to draw participants' attention to either (opposite) relational information or to co-occurrence information. A difference in implicit evaluation of the CSs between these two conditions would let us speculate about the way in which relational information is encoded. For instance, it would be possible that this process requires a certain level of attention drawn to relational information, as well as an amount of cognitive resources available. This latter assumption could be easily tested by giving one of the groups a secondary task to perform, at the same time in which relational information is presented.

On the other hand, it is also possible that individuals differ in terms of their ability to process relational information. Measures of interindividual differences (e.g., need for cognition, Cacioppo & Petty, 1982) may be a valuable addition to almost every experiment on this topic, and would be particularly useful to explain unclear results (e.g., null effects due to a number of participants who show to have not encoded relational information properly).

The manipulation of cognitive resources would also be interesting in order to test for the automaticity features of the acquisition and activation of relational information. It would test whether relational information moderates implicit evaluation in an efficient matter. Moors and De Houwer (2006) define automaticity as a term that refers to a number of features (e.g., uncontrollability, unconsciousness, efficiency, etc.) that are distinct and independent from each other. Therefore, a process or effect can be automatic according to one feature and non-automatic according to another. Hence, it is worth exploring automaticity features separately. For instance, another feature of automaticity that could be worth exploring is goal-independence. In an experiment in which both relational information and pairings of stimuli are presented, one could give participants the goal to focus on relations between stimuli, or the goal to focus on the co-occurrences, or simply no goal, and measure the resulting implicit evaluation. This could tell us whether the impact of relational information on evaluation is automatic in the sense of goal-independent. In addition, in another study relational and co-occurrence information could be presented either explicitly or subliminally, in order to explore whether relational information could moderate implicit evaluation also unconsciously.

In terms of the limitations of the present research, the experiments with null effects or attenuations of EC effects, are particularly open to improvements. Perhaps the null effects were due to the use of insensitive implicit measures either to implicit evaluations or to the impact of relational information on implicit evaluations. In Chapter 2, for instance, the AMP or the evaluative priming with naming task might not have picked up any trace of implicit evaluation. In this sense, future research could extend its scope to other implicit measures that we have not considered in the present dissertation (e.g., the Extrinsic Affective Simon Test, De Houwer, 2003; the Go/No-Go Association Test, Nosek & Banaji, 2001; the Approach-Avoid task, Rinck & Becker, 2007). Besides the most traditional implicit measures, the Implicit Relational Assessment Procedure (IRAP, see Barnes-Holmes, Barnes-Holmes, Power, Hayden, & Milner, 2006), is a recent implicit measure designed to tap in relational responses, which has lately been gaining popularity within evaluation researchers. This implicit measure is definitely worth exploring, in the domain of relational information and EC, and future research could profit from a comparison between its measurement outcomes and those of the more traditional implicit measures (e.g., IAT, personalized IAT, affective priming). Indeed, this effort could lead to an extensive taxonomy of the vast range of implicit measures of evaluation, according to their sensitivity to relational information, which would represent a very useful tool for evaluation researchers.

To conclude, research on the impact of relational information on implicit evaluation should aim at extending its body of functional moderators. However, in order to guide research, it would be important to prioritize the investigation of the moderators whose explanation maximally constrains the assumptions of the existing mental process models. We are confident that research on both the functional moderators and on the mental mediators of evaluation would eventually lead to an improvement of the actual mental process models and consequently to reach the ultimate goal of a more accurate explanation and prediction of human behavior.

References

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

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task: A measure of association strengths. *Experimental Psychology, 56*, 329-343. doi: 10.1027/1618-3169.56.5.329

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., Stewart, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist, 32,* 169-177. http://www.psihq.ie/irish-psychologist-journal-of-psychology

Cacioppo, John T. & Petty, Richard E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116-131 doi: 10.1037/0022-3514.42.1.116

De Houwer, J. (2003). The Extrinsic Affective Simon Task. Experimental Psychology, 50, 77-85. doi: 10.1027//1618-3169.50.2.77

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230-241. http://www.ucm.es/info/psi/docs/journal/

De Houwer, J. (2009a). How do people evaluate objects? A brief review. *Social and Personality Psychology Compass, 3*, 36-48. doi: 10.1111/j.1751-9004.2008.00162.x

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

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

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2012). *A functional-cognitive framework for attitude research*. Manuscript submitted for publication.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strengths. Social Cognition, 25, 603–37. doi: 10.1521/soco.2007.25.5.603

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

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

Gawronski, B., & Payne, B. K. (2010). Handbook of implicit social cognition: Measurement, theory, and applications**.** New York, NY: Guilford Press.

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

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

Han, H. A., Czellar, S., Olson, M. A., & Fazio, R. H. (2010). Malleability of attitudes or malleability of the IAT? *Journal of Experimental Social Psychology*, *46*, 286-298. doi: 10.1016/j.jesp.2009.11.011

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.

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

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorising in implicit attitude research: Propositional and behavioral alternatives. *Psychological Record, 61*, 465–496. http://thepsychologicalrecord.siuc.edu/

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

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326. doi: 10.1037/0033-2909.132.2.297

Molet, M., et al. Relational responding modulates and reverses affective ratings in evaluative conditioning. *Learning and Motivation* (2012). doi: http://dx.doi.org/10.1016/j.lmot.2012.11.001

Moran, T. & Bar-Anan, Y. (2012). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion,* 1-10. doi: 10.1080/02699931.2012.732040

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. Social Cognition, 19(6), 625-666. doi: 10.1521/soco.19.6.625.20886

Olson, M., & Fazio, R. H. (2004). Reducing the Influence of Extrapersonal Associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality & Social Psychology, 86(5),* 653-667, doi: 10.1037/0022-3514.86.5.653

Olson, M. A., & Fazio, R. H. (2009). Implicit and explicit measures of attitudes: The perspective of the MODE model. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), Attitudes: Insights from the new implicit measures (pp. 19-63). New York, NY: Psychology Press.

Payne, B. K., Cheng, C. M., Govorun, O., Stewart, B. D. (2005). An Inkblot for Attitudes: Affect Misattribution as Implicit Measurement. *Journal of Personality and Social Psychology*, 89, 277-293. doi: 10.1037/0022-3514.89.3.277

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

Rinck M., Becker E.S. (2007). Approach and avoidance in fear of spiders. *Journal of Behaviour Therapy and Experimental Psychiatry*, 38, 105–120. doi: 10.1016/j.jbtep.2006.10.001

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*, 995-1008

Spruyt, A., Hermans, D., De Houwer, J., Vandekerckhove, J., & Eelen, P. (2007). On the predictive validity of indirect attitude measures: Prediction of consumer choice behavior on the basis of affective priming in the picture-picture naming task. *Journal of Experimental Social Psychology, 43,* 599-610. doi: 10.1016/j.jesp.2006.06.009

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

Footnotes

[1] We thank Bertram Gawronski for drawing our attention to this idea.

NEDERLANDSTALIGE SAMENVATTING

De invloed van relationele informatie op impliciete evaluatie

Evaluatie is een belangrijke determinant van gedrag. Daarom speelt de studie hiervan een grote rol in psychologisch onderzoek. We definiëren evaluatie als het effect van stimuli op evaluatieve responsen (De Houwer, Gawronski & Barnes-Holmes, 2012). Traditioneel focuste onderzoek zich op zogenaamde expliciete evaluaties. Dit zijn intentionele en gecontroleerde evaluaties (De Houwer, 2009a). Impliciete evaluaties zijn automatische, ongecontroleerde evaluaties (De Houwer, 2009a), typisch gemeten door zogenaamde impliciete maten van evaluatie. Hun belang werd aangetoond in verschillende domeinen van psychologisch onderzoek (voor een overzicht, zie Gawronski & Payne, 2010).

Een typische manier om impliciete evaluaties te exploreren en te veranderen is door het paren van stimuli. De verandering in het leuk vinden van een neutrale stimulus (geconditioneerde stimulus, CS) nadat het herhaaldelijk gepaard werd met een positieve of negatieve stimulus (ongeconditioneerde stimulus, US) is een effect dat evaluatieve conditionering genoemd wordt (EC; De Houwer, 2007). Wanneer we kijken naar veranderingen in impliciete evaluaties, spreken we van EC van impliciete evaluaties. Traditioneel nemen vele associatieve modellen (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Fazio, 2007; Olson & Fazio, 2009) en twee-proces modellen (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006) aan dat impliciete evaluatie gemedieerd wordt door de associaties tussen de mentale representaties van de CSen en de USen, die gevormd werden na het veel voorkomen van CS-US paren. Eens deze ongekwalificeerde associaties gevormd zijn, kunnen ze automatisch geactiveerd worden wanneer de CS wordt aangeboden. Hieruit volgt dat de aanbieding van de CS automatisch resulteert in de activatie van de mentale representaie van de US, inclusief de representatie van de valentie van de US. Op die manier zal de CS een automatische impliciete evaluatie ontlokken.

Onderzoeker zijn echter pas recent gestart om na te gaan of ook de manier waarop CS en US gerelateerd zijn, een invloed kan hebben op impliciete evaluaties. We gebruiken de term “relationele informatie” om te verwijzen naar informatie over het type van relatie tussen gebeurtenissen. Het is bijvoorbeeld duidelijk dat of een substantie (CS) een ziekte (USneg) uitlokt of deze juist voorkomt een belangrijk gegeven is om deze substantie te evalueren. Daarom gaan propositionele modellen (e.g., De Houwer, 2009b; Mitchell, De Houwer, & Lovibond, 2009) ervan uit dat alle informatie bewaard wordt in de vorm van proposities. Deze proposities zijn gekwalificeerde verbindingen tussen mentale representaties (vb. “de CS is het effect van de US”). In tegenstelling tot associaties kunnen deze verbindingen dus ook relationele informatie bevatten. Van zodra er aangenomen wordt dat proposities automatisch geactiveerd kunnen worden, dan men postuleren dat ook proposities kunnen leiden tot impliciete evaluaties. De experimenten die beschreven zijn in het huidige proefwerk werden geïnspireerd door de propositionele benadering van impliciete evaluaties en richten zich tot het onderzoeken van de modererende rol van relationele informatie op het verwerven van impliciete evaluaties.

In Hoofdstuk 1 presenteerden we dezelfde CS-US paren in twee condities. In beide condities werden ook andere, zogenaamde context paren gepresenteerd. In de conditie *Tegenovergesteld*, impliceerden de context paren een regel die een tegenstelling suggereerde tussen de CSen en USen. In de conditie *Zelfde* daarentegen, impliceerden ze een regel die gelijkheid tussen de CSen en US suggereerden. Nadien werden de evaluaties van de CSen gemeten. We vonden een significante moderatie van de context regel die een standaard EC effect van impliciete evaluaties produceerde in de conditie *Zelfde* en een verzwakking van het EC effect van impliciete evaluaties in de conditie *Tegenovergesteld*.

In Hoofdstuk 2 testten we een potentiële moderator van het effect van relationele informatie op impliciete evaluaties, namelijk de eigenschappen van de gebruikte impliciete maat. In vier experimenten testten we de sensitiviteit voor relationele informatie van drie verschillende impliciete maten van evaluatie. De resultaten toonden aan dat hoewel de impliciete associatie test (IAT; Greenwald, McGhee, & Schwartz, 1998) een zekere sensitiviteit vertoonde, de affectieve misattributie procedure (AMP; Payne, Cheng, Govorun & Stewart,

2005) en de evaluatieve priming met het benoemen van responsen (Spruyt, Hermans, De Houwer, Vandekerckhove, & Eelen, 2006) geen significante EC effecten vertoonden. Onze resultaten laten ons daarom niet to om sterke claims te maken over de sensitiviteit van de maat van relationele informatie.

In Hoofdstuk 3 testten we hoe relationele informatie, in de vorm van verbale instructies, een impact kan hebben op impliciete en expliciete evaluaties. In Experiment 1 presenteerden we tegengestelde relationele informatie ofwel voor ofwel na een serie van CS-US paren. In Experiment 2 probeerden we de assumptie te testen dat het samen voorkomen van CS-US wordt aanzien als een cue die wijst op een equivalentie tussen de CSen en USen. Ten slotte, in Experiment 3, testten we de assumptie dat,in vergelijking met expliciete evaluatie, impliciete evaluatie meer beïnvloed wordt door de eerste gepresenteerde relationele informatie. Onze resultaten bevestigden deze hypothese. Bijkomend vonden we dat relationele informatie over het feit dat CS en US equivalent zijn, niets toevoegt aan het effect van het louter samen voorkomen van CS en US op impliciete evaluaties. Informatie over het feit dat CS en US tegengesteld zijn, had echter wel een effect dat tegengesteld is aan het effect van het louter samen voorkomen van CS en US.

Tot slot, in Hoofdstuk 4 repliceerden we de studies van een controversiële paper van Moran and Bar-Anan (2012). Zij onderzochten de moderatie van de start-stop relaties op impliciete evaluaties. Ze vonden geen bewijs voor de impact van relationele informatie op impliciete evaluatie. We brachten echter enkele modificaties aan het design toe die de rol van relationele informatie zouden kunnen versterken. Hoewel de resultaten niet sluitend waren, boden ze steun voor het idee dat relationele informatie ook in het opzet van Moran en Bar-Anan een invloed kan hebben op impliciete evaluaties.

Onze resultaten hebben belangrijke implicaties voor mentale proces modellen van impliciete evaluaties. Om onze resultaten te verklaren, moeten bestaande modellen van impliciete evaluatie zeer specifieke assumpties maken. Als dusdanig leggen onze resultaten duidelijke restricties op aan huidige en toekomstige modellen van impliciete evaluatie. Onze bevindingen zijn het meest problematisch voor puur associatieve modellen. Langs de andere kant, meer recente twee-processen modellen Gawronski & Bodenhausen, 2011) kunnen onze bevindingen verklaren op voorwaarde dat ze toelaten dat propositionele processen impliciete evaluaties kunnen beïnvloeden. Tot slot, propositionele modellen kunnen onze bevinden verklaren op voorwaarde dat verondersteld wordt dat proposities een automatische invloed kunnen hebben op preferenties.

Toekomstig onderzoek kan zich richten op de analyses van functionele moderators van evaluatie. We toonden aan dat relationele informatie inderdaad een belangrijke moderator is van impliciete evaluatie. Verder onderzoek zou onze huidige kennis verder kunnen uitbreiden door het exploreren van deze functionele moderator, bijvoorbeeld het in rekening brengen van verschillende relaties (bvb vergelijkende, temporele, hiërarchische), de automaticiteit van de relationele informatie, de condities onder welke de relationele informatie effectief is. Functionele bevindingen zouden het toelaten om nieuwe en meer accurate mentale processen theorieën van de mentale processen die impliciete evaluaties mediëren. Deze verbeteringen aan de huidige mentale processen modellen zouden dan leiden tot betere verklaringen en voorspellingen van het menselijk gedrag.

Referenties

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

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10,* 230-241. http://www.ucm.es/info/psi/docs/journal/

De Houwer, J. (2009a). How do people evaluate objects? A brief review. *Social and Personality Psychology Compass, 3*, 36-48. doi: 10.1111/j.1751-9004.2008.00162.x

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

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2012). *A functional-cognitive framework for attitude research*. Manuscript submitted for publication.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strengths. *Social Cognition*. 2007;25:603–37. doi: 10.1521/soco.2007.25.5.603

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

Gawronski, B., & Payne, B. K. (2010). Handbook of implicit social cognition: Measurement, theory, and applications**.** New York, NY: Guilford Press.

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

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

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

Moran, T. & Bar-Anan, Y. (2012). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion,* 1-10. doi: 10.1080/02699931.2012.732040

Olson, M. A., & Fazio, R. H. (2009). Implicit and explicit measures of attitudes: The perspective of the MODE model. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), Attitudes: Insights from the new implicit measures (pp. 19-63). New York, NY: Psychology Press.

Payne, B. K., Cheng, C. M., Govorun, O., Stewart, B. D. (2005). An Inkblot for Attitudes: Affect Misattribution as Implicit Measurement. *Journal of Personality and Social Psychology*, 89, 277-293. doi: 10.1037/0022-3514.89.3.277

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

Spruyt, A., Hermans, D., De Houwer, J., Vandekerckhove, J., & Eelen, P. (2007). On the predictive validity of indirect attitude measures: Prediction of consumer choice behavior on the basis of affective priming in the picture-picture naming task. *Journal of Experimental Social Psychology, 43,* 599-610. doi: 10.1016/j.jesp.2006.06.009