Article

Personality and Social Psychology Review
2019, Vol. 23(3) 267–284

DOI: 10.1177/1088868318795730
pspr.sagepub.com
SAGE

How Do Actions Influence Attitudes? An Inferential Account of the Impact of Action Performance on Stimulus Evaluation

Pieter Van Dessel[1], Sean Hughes[1], and Jan De Houwer[1]

Abstract

Over the past decade, an increasing number of studies have shown that the performance of specific actions (e.g., approach and avoidance) in response to a stimulus can lead to changes in how that stimulus is evaluated. In contrast to the reigning idea that these effects are mediated by the automatic formation and activation of associations in memory, we describe an inferential account that specifies the inferences underlying the effects and how these inferences are formed. We draw on predictive processing theories to explain the basic processes underlying inferential reasoning and their main characteristics. Our inferential account accommodates past findings, is supported by new findings, and leads to novel predictions as well as concrete recommendations for how action performance can be used to influence real-world behavior.

Keywords

attitudes, action effects, approach-avoidance, inferential account, predictive processing

It is almost axiomatic to claim that attitudes exert an important impact on behavior (Ajzen & Fishbein, 2005). Interestingly, however, there is also a substantial amount of research showing that behavior can have an impact on attitudes (Olson & Stone, 2005). For example, nodding one's head while listening to a message can improve liking of that message (Wells & Petty, 1980), selecting one object from two equally attractive alternatives can lead to more favorable evaluations of the object (Gawronski, Bodenhausen, & Becker, 2007), and making approach movements when viewing Chinese ideographs can result in more positive ratings of those stimuli (Cacioppo, Priester, & Berntson, 1993). In this article, we focus on situations where the performance of a specific action in relation to a stimulus influences the subsequent evaluation of that stimulus (i.e., evaluative stimulus-action effects).

There has recently been a surge in interest in evaluative stimulus-action effects, triggered in part by the seminal work of Kawakami, Phills, Steele, and Dovidio (2007) who found that repeated performance of approach or avoidance movements in response to images of Black and White individuals altered evaluations of in- and out-groups. This approach-avoidance (AA) training procedure has now been adopted in many studies with the typical outcome that repeated approach leads to more positive stimulus evaluations whereas repeated avoidance leads to more negative stimulus evaluations (Van Dessel, De Houwer, & Gast, 2016). The fact that AA training effects have been found with difficult to change behaviors (e.g., implicit prejudice: Kawakami et al., 2007; addictive behaviors: Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011) and that effects seem to occur under some of the conditions of automaticity (e.g., unintentionally: Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016; unconsciously: Kawakami et al., 2007; but see Van Dessel, De Houwer, Roets, & Gast, 2016) are important reasons for the increasing popularity of AA training studies. AA training effects have now been observed across many different domains in psychology, including social psychology (e.g., racial evaluations: Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011), clinical psychology (e.g., alcohol: Wiers et al., 2011; cigarettes: Wittekind, Feist, Schneider, Moritz, & Fritzsche, 2015; social anxiety: Taylor & Amir, 2012; depression: Becker et al., 2016), and educational psychology (e.g., mathematics: Kawakami, Steele, Cifa, Phills, & Dovidio, 2008).

Although a growing number of studies have shown that evaluative stimulus-action effects such as AA training effects are both robust and widely applicable, others have sometimes failed to obtain such findings (e.g., Becker, Jostmann, Wiers, & Holland, 2015; Krypotos, Arnaudova, Effting, Kindt, & Beckers, 2015; Vandenbosch & De Houwer, 2011;

---

[1] Ghent University, Belgium

Corresponding Author:
Pieter Van Dessel, Department of Experimental Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium.
Email: Pieter.vanDessel@UGent.be

Van Dessel, De Houwer, Roets, & Gast, 2016). These contradictory findings suggest that (a) there are important boundary conditions for these effects, (b) many of these conditions have yet to be discovered, and (c) establishing such conditions has so far proven difficult. One possible reason for these limitations is that the dominant theoretical accounts offered to explain this phenomenon are inaccurate. These accounts assume that a change in mental associations due to (repeated) pairings of stimuli and actions leads to changes in liking (association formation accounts: e.g., Kawakami et al., 2007). As we describe below, these dominant accounts provide an intuitive explanation of many evaluative stimulus-action effects, yet they also have important limitations. We therefore propose that the time is ripe for considering alternative explanations to association formation accounts that might provide a better understanding of evaluative stimulus-action effects (see Boddez, De Houwer, & Beckers, 2017; Hughes, Barnes-Holmes, & De Houwer, 2011, for similar arguments in the context of learning research in general). With this in mind, we offer a new perspective on evaluative stimulus-action effects that diverges from traditional accounts and builds on the idea that inferential (rather than associative) processes underlie changes in liking due to stimulus-action relations. Our model of evaluative stimulus-action effects is in accordance with the idea that people often infer stimulus evaluations from their actions (as specified in self-perception theory: Bem, 1972) and with the increasingly more popular view that, in contrast to what is often assumed, human cognition does not necessarily depend on two different types of mental processes such as automatic, associative and controlled, propositional processes (see Melnikoff & Bargh, 2018). Furthermore, our model is consistent with recent evidence indicating that propositional processes play an important role in several psychological phenomena that have long been considered as associative in nature (e.g., automatic evaluation: see De Houwer, 2014a; Mann & Ferguson, 2015; or intuitive judgments: see Kruglanski & Gigerenzer, 2011).

In the remainder of this article, we first discuss the strengths and limitations of association formation accounts of evaluative stimulus-action effects. We then describe the core concepts that make up our inferential account and outline the general processes that operate on these concepts to produce evaluative stimulus-action effects. Thereafter, we delineate the inference steps involved in these effects and potential moderators according to our model. In the General Discussion, we highlight the added explanatory, predictive, and influence value of our model. We close with a discussion of the potential limitations and future directions offered by our account.

Association Formation Accounts

According to dominant accounts of evaluative stimulus-action effects, the pairing of stimuli and actions leads to a co-activation of their corresponding mental representations which automatically creates an association between the two representations (Kawakami et al., 2007). This association is typically conceived of as an unqualified link that transmits activation from one representation to another (Shanks, 2007). Once a strong enough association has been established, presentation of the stimulus will result in activation of the stimulus representation, which will then increase activation of the action representation. If the action representation contains evaluative components, this can lead to an evaluative response to the stimulus that is in-line with the valence of the action. This explanation is similar to associative explanations of evaluative conditioning (EC) effects (i.e., evaluative changes resulting from the pairing of a stimulus with other, valenced, stimuli; Baeyens, Eelen, Crombez, & Van den Bergh, 1992). Pairing a stimulus with a valenced event (i.e., performance of a valenced action or presentation of a valenced stimulus) creates a link between valenced representations and stimulus representations which can—in turn—lead to automatic changes in stimulus evaluation. Several associative accounts have been proposed to explain evaluative stimulus-action effects, and these differ mainly in the specific action representations that are assumed to become associated with stimulus representations. The three most popular and well-described accounts assume that associations are formed between representations of the evaluative stimuli and (a) representations of evaluative action attributes such as the valenced words used to describe the actions (common-coding account of evaluative stimulus-action effects: Eder & Klauer, 2009), (b) positive representations of the self (self-anchoring account of AA training effects: Phills et al., 2011), or (c) motivational systems of approach and avoidance (motivational-systems accounts of AA training effects: Wiers et al., 2011).

In accordance with EC research (see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010, for an overview), the fact that evaluative stimulus-action effects can occur under conditions of automaticity (e.g., in the absence of stimulus awareness: Kawakami et al., 2007, or without an intention to let action performance influence stimulus evaluations: Van Dessel, De Houwer, Gast et al., 2016) has been considered as strong support for the idea that automatic association formation mechanisms underlie these effects. Specifically, these results are in-line with the view that evaluative stimulus-action effects are mediated by processing in automatic systems that operate on the basis of association formation (Strack & Deutsch, 2004). Further support for this idea has come from studies showing that stimulus-action pairings can alter difficult to change or spontaneous behaviors that are assumed to be tied into these systems (e.g., implicit prejudice: Kawakami et al., 2007; addictive behavior: Wiers et al., 2011). One study also showed that evaluative stimulus-action effects can be enhanced when there are a higher number of stimulus-action pairings, consistent with the prediction made by associative accounts that the number of pairings determines the strength of stimulus-action associations and, as a result, the evaluative effect (Woud, Becker, & Rinck, 2011).

Importantly, however, the findings described in the previous paragraph, such as the observation that evaluative stimulus-action effects can occur under conditions of automaticity, are no guarantee that association formation processes mediate these changes in liking (see Mitchell, De Houwer, & Lovibond, 2009). In fact, evidence suggests that current associative accounts of evaluative stimulus-action effects have difficulty explaining many of the observed evaluative stimulus-action effects. First, most associative accounts imply that there are few (if any) boundary conditions for evaluative stimulus-action effects. Repeated performance of valenced actions such as approach and avoidance in response to stimuli should automatically lead to association formation (e.g., formation of associations between stimulus representations and representations of valenced words: Eder & Klauer, 2009; representations of the self: Phills et al., 2011; or motivational systems: Wiers et al., 2011) and resulting changes in evaluation. Hence, these models have difficulty explaining why some, but not other, studies produce AA training effects (see Mertens, Van Dessel, & De Houwer, 2018, for a discussion). Note that some associative accounts that were developed outside of the literature on AA training effects have specified assumptions about possible boundary conditions of association formation (e.g., attention for the pairings: Wagner, 1981). This could potentially explain specific null effects in previous studies on evaluative stimulus-action effects. Nevertheless, current accounts of AA training effects provide little information about the assumed boundary conditions of association formation.

Second, recent studies have shown that evaluative stimulus-action effects are moderated by specific variables that are not easily explained by associative accounts. For example, awareness of stimulus-action contingencies has been found to be an important moderator of AA training effects (see Van Dessel, De Houwer, & Gast, 2016). This does not fit well with current associative accounts, which assume that pairings lead to automatic associative changes (e.g., in the absence of contingency awareness: Kawakami et al., 2007).

Third, other studies have shown that instructions about future actions can also create changes in stimulus evaluations and that these effects share important similarities with experience-based effects (e.g., unintentionality: Van Dessel, De Houwer, Gast et al., 2016). Similarly, mere observation of AA actions can also lead to evaluative stimulus-action effects (Van Dessel, Eder, & Hughes, 2018). Because instructions and observations do not involve pairings of stimuli and valenced actions, it is unclear how these effects might occur on the basis of associative processes (see Lovibond, 2003, for a detailed discussion of the limitations of associative learning models in accounting for learning via instructions). It is also unclear how associative accounts can explain evaluative stimulus-action effects that only involve *single* pairings of stimuli with valenced actions (e.g., Centerbar, Schnall, Clore, & Garvin, 2008) or pairings of stimuli with actions of neutral valence (Bem, 1972).

In sum, although it is difficult if not impossible to refute association formation models of evaluative stimulus-action effects as a class of models, there are several findings that do not readily fit with the association formation models currently available in the literature. With this in mind, the current article explores the merits of another type of model of evaluative stimulus-action effects that differs in important ways from the associative accounts outlined above.

The Inferential Account

Basic Building Blocks

Propositions. The core conceptual unit of our inferential account is the propositional representation. A propositional representation is a mental representation that constitutes a statement about the world (De Houwer, 2014a). Propositional representations contain relational information (i.e., information about how concepts are related; see Lagnado, Waldmann, Hagmayer, & Sloman, 2007) which distinguishes them from typical (unqualified) associative representations. Note that propositions could—in principle—be implemented in associative networks, provided that those networks are capable of encoding relational information (De Houwer, 2014a). Moreover, and unlike what is often assumed, propositions are not necessarily verbal, but can involve embodied or grounded representations (De Houwer, 2014b). Hence, nonhuman and nonverbal (e.g., infants) organisms can also store propositional information.

We argue that relational information is at the core of the inferences that mediate evaluative stimulus-action effects. For example, there is an important difference between the acquisition of propositional information that *"I am approaching stimulus A"* compared with *"Stimulus A approaches me."* "I," "approach," and "stimulus A" are present in both cases but the relation between the three concepts—that is, the specified role of each of the concepts—is fundamentally different. A recent study indicated that this difference in relational information can moderate evaluative learning: instructions stating that participants would perform an AA action in relation to a stimulus produced bigger changes in evaluations than instructions stating that the stimulus would perform an AA action in relation to participants (see Van Dessel, De Houwer, & Smith, 2018). Such (difference in) relational information is difficult to capture by models that operate on the basis of associative representations (Gentner, 2016; Hummel, 2010).

Inferences. We define an inference as a specific sub-type of propositional representation: it is a proposition (thus, it is relational rather than associative in nature) but one that is constructed on the basis of other propositional information.

The construction process that leads to the inference can be seen as an information generation procedure that involves the application of information generation (i.e., inference) rules to information that is currently entertained. Note that we use the term “inference” to describe the outcome of the computation process rather than the computation process itself. We refer to the computation process as “making an inference” or “inferential reasoning.” Our definition of an inference is broad in the sense that inferential reasoning can occur on the basis of a multitude of different inference rules. These rules can encode general statements about the world (e.g., if-then rules), but they can also constitute mere similarity metrics (e.g., analogical mapping rules: Gentner & Smith, 2013; Hahn & Chater, 1998). As we explain below, however, we make specific predictions about the processes underlying inferential reasoning and the inference rules that people use under specific circumstances, constraining our account.

Our definition of an inference implies that not all propositions are necessarily inferences (but all inferences are propositions). However, in the current model, we draw on the assumption of predictive processing theories that the activation of propositional information constitutes an inferential process that involves the prediction of information (i.e., construction of information that is compatible with activated information, see the following section). From this perspective, any activated propositional information can be construed as an inference. We will therefore always use the term inference (rather than proposition) in the continuation of this article. Note that in our model, all inferences can also be seen as “predictions” in the sense that they constitute information that is predicted (i.e., constructed on the basis of probabilistic information).

Evaluations. The outcome of the processing steps described in our account is a change in stimulus evaluation. In accordance with De Houwer, Gawronski, and Barnes-Holmes (2013), we use the term “evaluation” to refer to a behavioral phenomenon, that is, the impact of stimuli on evaluative responses. Note that this definition of evaluation avoids conflation of the to-be-explained behavior (i.e., the evaluative response) with the mental construct that is used to explain the behavior (i.e., the attitudinal representation). Stimulus evaluations can occur under the various conditions of automaticity (e.g., uncontrolled, unconscious, efficient, or fast; see Moors, 2016; i.e., implicit evaluation) or arise in a more deliberate and controlled manner (i.e., explicit evaluation). Our inferential model describes how the performance of actions in response to a stimulus can produce changes in both types of evaluations.

The Inferential Process

To clarify the basic mental processes underlying evaluative stimulus-action effects, we will first describe the nature of inferential reasoning in general. In doing so, we draw on the notion that predictive processing can provide the basis for inferential reasoning.[1] On the basis of this idea, we make three key assumptions that help explain evaluative stimulus-action effects, namely that inferential reasoning (a) strongly depends on momentary goals, (b) is highly contextual, and (c) is learning-dependent.

Predictive processing. Our approach is based on an idea that is at the core of many recent theories in various areas of psychological science (e.g., psychophysiology: George & Hawkins, 2009; perceptual psychology: Proulx, 2014; psychopathology: Fletcher & Frith, 2009; see Metzinger & Wiese, 2017, for an overview), namely the idea that predictive processing is the basis of cognitive processing. In this view, the mental system is seen as a “prediction machine” that is constantly anticipating events in the world around it to be able to respond to them quickly and accurately (Helmholtz, 1962). Bayesian approaches to cognitive processing assume that this comprises the continuous updating of a person’s generative model of the world through a process that involves computing probabilities—on the basis of Bayes’ (1958) theorem—with the aim of integrating and updating prior evidence for stored information (see Penny, 2012). This predictive inference mechanism is considered of great evolutionary importance because it helps optimize the use of energy expenditure (by reducing prediction error) and avoid entropy (i.e., disorder) while allowing organisms to respond to the environment in an optimal fashion (Friston, 2010). Importantly, this mechanism might also fit the architecture of the brain and its various substrates (see Bastos et al., 2012; George & Hawkins, 2009).

From this perspective, inferential reasoning essentially involves the construction of information that is compatible with activated information on the basis of a person’s mental model of the world. This generative model can be seen as an information network that represents information in a hierarchical manner such that information at higher levels can be used to predict compatible information at lower levels (Friston, 2008). The information is essentially propositional as it has a specific truth value (i.e., a probabilistic index for retrieval). Inferential reasoning thus involves drawing probabilistic samples from a pool of propositional information by applying inference rules to currently entertained information (Sanborn & Chater, 2016). For example, when information is entertained that one has performed an approach action in response to a certain stimulus, rules of analogical mapping may be applied to this information to compute compatible information (e.g., information about stimulus valence) with a certain precision.

It is important to note that predictive processing accounts assume that inferential reasoning is not necessarily slow and effortful, which deviates from certain other inferential reasoning theories. Rather, inferential reasoning (i.e., probabilistic construction of compatible information on the basis of activated information) can occur in a manner that is to a greater or lesser extent automatic. Specifically, when certain information is entertained, this will facilitate fast and easy activation of compatible information when the inference rule that supports this inference is well-practiced or more effortful activation of compatible information when application of the required inference rule is more difficult. Note that this account proposes the same general mechanism for automatic and controlled mental processes (i.e., the prediction of compatible information on the basis of prior evidence via the application of inference rules that are more or less difficult to apply). This is consistent with recent recommendations to explore alternatives to dual-process theories of human cognition (e.g., Melnikoff & Bargh, 2018).

Inferential reasoning is goal-dependent. A first essential question for any inferential reasoning theory is as follows: "*Why* do organisms engage in inferential reasoning in the first place?" Predictive processing theories assume that an essential feature of organisms is that they strive for homeostasis or the maintenance of optimal internal states (e.g., for the sake of survival). To achieve homeostasis, organisms have evolved such that they are able to represent desired states (i.e., goals). Inferential reasoning may serve the function to construct information that is compatible with goals, allowing for adaptive action (i.e., actions that allow one to achieve desired states: Pezzulo, Rigoli, & Friston, 2015). As a result, inferential reasoning may be critically dependent on the goals that one entertains.

An important feature of predictive processing theories is that actions are considered "active inferences" or inferences that act on the environment (Friston, 2010). Specifically, it is assumed that, when a desired state is activated (e.g., to be satiated), organisms will make predictions about available actions and their respective outcomes. When the desired end state is predicted with sufficient precision on the basis of a specific action (e.g., opening the fridge), this will cause the activation of information about proprioceptive states necessary for the action, which will lead to action execution (Chetverikov & Kristjánsson, 2016; Cisek & Pastor-Bernier, 2014). From this perspective, actions are considered to be essentially goal-directed in that they are emitted on the basis of a person's activated goals. We assume that active inferences share this feature with inferences that involve the construction of other information than proprioceptive information (e.g., sensory or verbal information). For example, not only inferences involved in action planning (e.g., information about proprioceptive states necessary for AA actions) but also inferences involved in action interpretation (e.g., information about the valence of AA actions) will depend on active representations of desired states.

Inferential reasoning is context-dependent. A second essential question for an inferential theory is as follows: "What inferential reasoning will organisms engage in?" The answer to this question is based on the assumption that organisms strive to minimize energy expenditure (Friston, 2010). Inferential reasoning that takes into account large amounts of information will therefore be unserviceable. Instead, inferential reasoning is considered to be highly contextual such that it involves the sampling of information on the basis of momentarily entertained information (Sanborn & Chater, 2016). This can explain why inferential reasoning is not necessarily optimal even though it might depend on optimizing rules (e.g., Bayesian updating of information). It is a common misconception that inferential reasoning is "cold," rational, and error-free (Moors, 2014). Rather, inferential reasoning can be irrational (in part) because currently entertained information strongly biases inferential reasoning. Depending on the context, people might entertain information that leads them to make inferences that are not logical in nature. For example, recent evidence suggests that contextual retrieval of information can lead to belief biases (see Banks, 2013) or probabilistic reasoning errors (see Sanborn & Chater, 2016). The context-dependence of inferential reasoning might also explain why people engage in suboptimal behavior. Specifically, action selection may depend on what information about desired end states is contextually available, such that only a subset of all relevant goals will inform a person's actions (see also Moors, Boddez, & De Houwer, 2017). From this perspective, inferential reasoning (and action selection) can be seen as a satisfying process; if activated information is good enough to achieve specific (contextually activated) goals, the mental system will stop sampling information and save its energy.

Inferential reasoning is learning-dependent. A third important assumption of many predictive processing accounts (but also other inferential reasoning accounts) that might help explain what inferences people make is that inferential reasoning involves the application of inference rules that have their roots in both phylogenetic and ontogenetic development. For example, a person may become more inclined to apply a logical modus ponens rule (if "A implies B" and "A" both hold, then we can deduce "B") to specific information if the application of this rule to similar information has led to good outcomes in the past (i.e., accurate predictions for achieving specific goals). The fulfillment of goals should trigger the updating of probabilities (on the basis of Bayes rule) such that application of certain inference rules will be more likely to be repeated in the future. Note that this mechanism can also lead to irrational inferences. For example, a previously learned rule that often leads to good outcomes may be applied even when application of this rule is not suited in the current environment. Hence, inferential processes can sometimes be biased because the inference rules that support them are not logically correct or are incorrectly applied. In fact, inferential reasoning may often depend on (heuristic) rules that do not always lead to optimal behavior (Evans, 2010; Kahneman, Slovic, & Tversky, 1982; Kruglanski & Gigerenzer, 2011). For example, people frequently use an availability heuristic

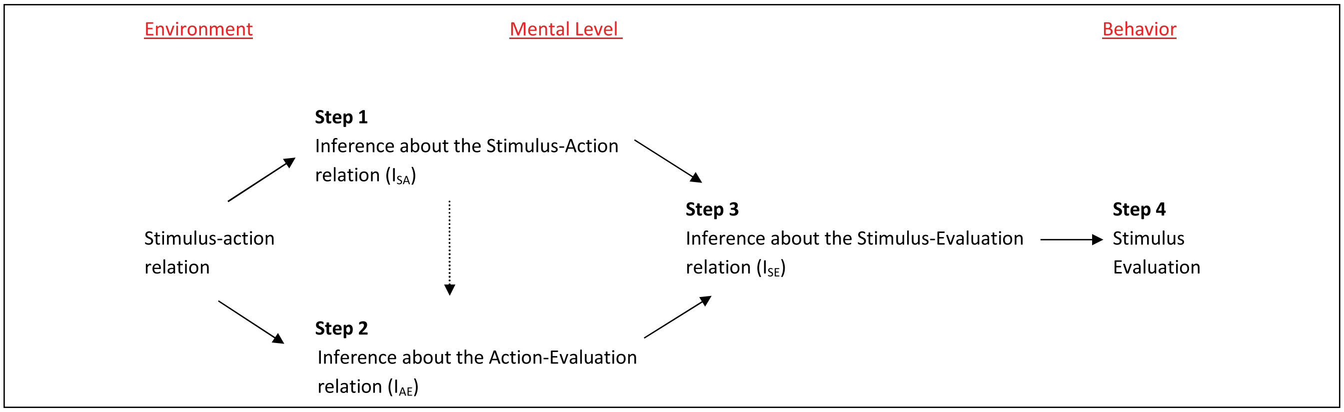

Figure 1. Schematic of the basic steps involved in the effects of actions on stimulus evaluation according to the inferential account of evaluative stimulus-action effects.

(i.e., they give disproportionate weight to easily available information) when making inferences. These (heuristic) rules may be strongly integrated in a person's cognitive system (and hence, easily used) because they are "ecologically" rational (i.e., they often lead to good outcomes in a person's typical environment, see Brighton & Gigerenzer, 2012, for an elaborate discussion).

Processing Steps: From Stimulus-Action Relations to Evaluations

Now that we have specified the nature of inferences and how these inferences are brought about in general, we will provide more detailed information about the inferential processing steps involved in evaluative stimulus-action effects specifically. We assume that the impact of action performance on stimulus evaluations is mediated by a four-step inferential process (see Figure 1) and that stimulus-action effects only arise when all steps are completed. For each step, we first describe the specific inference that is formed. Next, we clarify *how* this inference is formed and what factors moderate this inferential processing step (and subsequent evaluative stimulus-action effects) by drawing on the assumptions about the properties of inferential reasoning that we outlined above.

Step 1: Inference about a stimulus-action relation (I_{SA}). The inferential process is initiated when one or more actions are performed in response to a stimulus in the environment. Under certain conditions (see below), this will involve the formation of an inference that the stimulus and action are related to one another (e.g., "*Stimulus A is approached*" or *"I always avoid Stimulus B"*: Inference about a stimulus-action relation, I_{SA}). This first step inference can be about different facets of the relation (e.g., it can specify a single co-occurrence of stimulus and response or the strength of a statistical stimulus-response contingency; De Houwer, 2009).

Inferential process. In-line with predictive processing theories, we assume that people are very good at detecting relations in the environment because they continuously make predictions about the world around them (to support appropriate action) and readily update these predictions based on their interactions with the environment. However, people will not predict or detect every environmental contingency (this would be too energy consuming). Rather, a person's activated goals will determine the amount of attention that an environmental regularity (e.g., stimulus-action relation) will receive (allowing for more precise predictions: Feldman & Friston, 2010). Hence, performance of an action in response to a stimulus will lead to the formation of an inference about a relation between the action and the stimulus (I_{SA}) when this information is in line with one's current goals. Once registered, the activation level of this inference will depend on its inferred relevance to current goals, and this will determine the extent to which the inference biases the generation of other information. When the activation level of I_{SA} is sufficiently strong, this will provide the basis for evaluative stimulus-action effects.

Moderators. Our account assumes that any contextual factor that either facilitates or impedes the goal to detect stimulus-action relations will moderate Step 1. We therefore predict moderation of evaluative stimulus-action effects on the basis of manipulations of the external context (the state of the external environment) that achieve this aim. For example, effects should be facilitated when participants are informed that there are specific (stimulus-action) relations in the action task or that it is important to detect them. Enhanced effects should also be observed on the basis of more indirect instructions that facilitate the goal to detect stimulus-action relations (e.g., instructions that it is important to perform the experiment in a thoughtful manner) or by providing (performance-related) incentives. Furthermore, effects should be affected by contextual manipulations that either facilitate or impede a person's goal to attend to the stimulus, the action, or the relation between the two because this should influence the goal to detect stimulus-action relations. For example, we predict a facilitation of evaluative stimulus-action effects when stimulus identity is task relevant (see Van Dessel, De Houwer, Gast, Roets et al., 2016; Van Dessel et al., 2016), when the action task involves more trials (facilitating the goal to accurately predict the correct action on the basis of stimulus features: see Woud et al., 2011), and when the relation between action and stimulus is stronger (e.g., deterministic rather than probabilistic stimulus-response contingencies).

We also assume that internal context factors (internal states of the organism) can moderate Step 1. First, we predict moderation of evaluative stimulus-action effects by transient internal states of the organism. For example, when a person has more task motivation, this should facilitate the goal to detect stimulus-action relations and therefore enhance effects (see Laham, Kashima, Dix, Wheeler, & Levis, 2014; Zogmaister, Perugini, & Richetin, 2016). Second, we predict moderation by more stable internal states, in-line with the assumption that inferences strongly depend on a person's pre-existing beliefs (which are the result of their prior learning history). For example, a person who has learned that it is usually beneficial to register stimulus-action contingencies should show stronger effects. Individual difference factors such as general processing fluency or need for cognition should also facilitate effects because they improve the efficiency or extensiveness of inferential reasoning in general.

Note that our account assumes that the inference about a stimulus-action relation (I_{SA}) is a more proximal determinant of evaluative stimulus-action effects than the actual regularity in the environment. Hence, the subjective representation of this relation should have a stronger impact on evaluative stimulus-action effects than objective experiences of the stimulus-action relation. This accords with studies showing that (a) instructions that specify I_{SA} can lead to a change in stimulus evaluation even in the absence of actual action performance (Van Dessel, De Houwer, Gast, & Smith, 2015) and (b) evaluative stimulus-action effects are stronger when participants are able to report I_{SA} (moderation by awareness of stimulus-action contingencies: Van Dessel, De Houwer, & Gast, 2016). As we noted above, these two findings are difficult to explain on the basis of association formation theories. Note that, from our perspective, the observation that awareness of stimulus-action contingencies moderates AA training effects is due to the ease of retrieval of the inference about the relation between stimulus and action rather than a causal effect of awareness per se. Contingency awareness provides a good indication of whether I_{SA} has been formed in a sufficiently strong manner such that I_{SA} is able to support evaluative stimulus-action effects. As argued by Cleeremans (2014), awareness of acquired information may occur when activation of this information has acquired sufficient strength such that the information is predicted by the mental system itself (i.e., information is represented on a meta-level). Thus, although awareness does not cause the formation of inferences, inferences that are formed in the absence of awareness may be much more weakly represented such that they are less easily retrieved and therefore are less likely to be used during the inferential process.

Step 2: Inference about an action-evaluation relation (I_{AE}). The second step involved in evaluative stimulus-action effects is the construction of an inference that relates the performed action to a certain valence (inference about an action-evaluation relation, I_{AE}). This inference can refer to the valenced properties of the performed action (e.g., *"approaching IS positive")*, but it can also relate the action to valence in other ways (e.g., "*Pleasant stimuli are typically approached*").

Inferential process. When participants perform an action in response to a stimulus or when participants make inferences about a stimulus-action relation (I_{SA}; completion of Step 1 can influence Step 2), they will construct compatible information on the basis of their generative model of the world. Depending on a person's activated goals, this can involve the construal of information that refers to the performed action and that relates the performed action to a certain valence. For example, participants may retrieve detailed episodes of previous positive or negative experiences with this action (e.g., other moments when positive stimuli were approached) when this information is easily generated on the basis of current goals (e.g., the goal to evaluate). When activation of such information is sufficiently strong, it will provide the necessary input for Step 3 and hence determine evaluative stimulus-action effects.

Moderators. Our account assumes that any contextual factor that either facilitates or impedes a person's goal to draw inferences about a relation between the performed action and valence will moderate Step 2. We therefore predict facilitation of evaluative stimulus-action effects on the basis of instructions that provide information about a relation between the performed action and valence (see Van Dessel, Hughes, De Houwer, & Smith, 2018) or when providing instructions or other incentives to retrieve such information. We also predict facilitation when participants are incentivized to attend to (a) the action (see Step 1), (b) valence in general (e.g., by making valence task relevant), or (c) a relation between the action and valence (e.g., informing participants about evaluative properties of the action). For example, we recently found that using valenced terms such as "up/down" or "approach/avoid" to describe an action (which might communicate the importance of action valence) strongly moderates evaluative stimulus-action effects (Van Dessel, Eder, & Hughes, 2018).

It is important to note that attention can be oriented toward evaluative properties of specific actions that are not normally considered. For example, even though a person's history of approach and avoidance might be overwhelmingly related to good (approach) or bad (avoid) things, respectively, activation of incongruent information can be facilitated in specific contexts (e.g., in the context of an electric shock, avoidance can be positive). A recent experiment provided evidence that activation of such non-typical I_{AE} can lead to contrastive evaluative stimulus-action effects. In this study, participants evaluated neutral stimuli that were repeatedly avoided more positively when other, feared, stimuli also had to be avoided rather than approached (Mertens et al., 2018).

We assume that Step 2 strongly depends on a person's learning history such that a person who has had more (salient) experiences in their life history that relate specific actions to valence will more easily construct information about a relation between this action and valence and will therefore show stronger evaluative stimulus-action effects. Accordingly, some people might more easily retrieve information that approaching is positive whereas others might more easily retrieve information that approaching can sometimes be negative (e.g., they might have learned during their lifetime that approaching stimuli can be scary; Hsee, Tu, Lu, & Ruan, 2014), and this should strongly moderate AA training effects.

Step 3: Inference about a Stimulus-Evaluation relation (I_{SE}). The third step involves the construal of an inference about the evaluative properties of the target stimulus (e.g., "*Stimulus A is pleasant*": Inference about a stimulus-evaluation relation, I_{SE}). We assume that this inferential process will only lead to evaluative stimulus-action effects when it incorporates both inferences about a stimulus-action relation (I_{SA}) and an action-evaluation relation (I_{AE}). For example, when a person has inferred that (a) they approached stimulus A (I_{SA}) and (b) they typically approach stimuli they like (I_{AE}), they might apply an "affirm the consequent" inference rule, allowing them to infer that they like stimulus A (I_{SE}). Note that I_{SE} does not necessarily represent an unambiguous relation between stimulus and evaluation. For example, after a person retrieves information that (a) they approached stimulus A and (b) approaching is somehow related to positive valence, they might make the inference that stimulus A is somehow related to positive valence (transitive inference: Burt, 1911).

Inferential process. Depending on a person's activated goals, they may construct information about a relation between a stimulus and evaluation (I_{SE}) on the basis of available information. Importantly, the generation of I_{SE} is insufficient for Step 3. I_{AE} and I_{SA} need to be integrated in the inferential process such that they determine the activation level of I_{SE}. This will depend on (a) the activation level of I_{AE} and I_{SA} and (b) the availability of an inference rule that facilitates activation of I_{SE} on the basis of I_{AE} and I_{SA}. When I_{SE} is activated on the basis of I_{AE} and I_{SA}, this will trigger the integration of I_{AE} and I_{SA} in evaluation (Step 4), allowing for evaluative stimulus-action effects.

Moderators. We assume that the third inference step is dependent on (a) a person's goal to generate I_{SE} and (b) integration of I_{AE} and I_{SA} in this inferential process. With regard to the first determinant, we predict stronger evaluative stimulus-action effects when participants learn that it is important to retrieve evaluative information about the target stimulus (e.g., on the basis of instructions). Note that this can occur not only during the action task but also during the evaluation task (at which time the retrieval of evaluative information about the stimulus is task-relevant). We also predict facilitation of evaluative stimulus-action effects when an incentive is provided to attend to (a) the stimulus (see Step 1), (b) evaluation in general (see Step 2), or (c) the relation between the stimulus and evaluation.

Regarding the second determinant (integration of I_{AE} and I_{SA}), we predict enhanced evaluative stimulus-action effects when the formation of I_{AE} and I_{SA} is facilitated (see Steps 1 and 2) and when facilitating the use of inference rules that allow constructing I_{SE} on the basis of I_{AE} and I_{SA}. For example, stronger evaluative stimulus-action effects should be observed when participants are given experience in applying a rule to I_{AE} and I_{SA} to infer I_{SE} (e.g., via a pre-training during which participants are required to learn that relations between stimuli and valenced actions can inform them about stimulus valence). This manipulation should be most effective if the training occurred in a similar context (facilitating retrieval of the inference in the current context) and if making the inference led to positive outcomes (facilitating future use of the inference rule). Importantly, moderation by the first determinant will depend on the second determinant (i.e., whether participants integrate I_{AE} and I_{SA} in the inferential process) because merely generating I_{SE} is not sufficient for evaluative stimulus-action effects (i.e., changes in evaluation that result from performance of a specific action in relation to a stimulus). Hence, we predict smaller evaluative stimulus-action effects when contextual factors facilitate construal of I_{SE} (e.g., "Person A is positive") on the basis of other information than I_{AE} and I_{SA} (e.g., salient information about positive behaviors of that person). In-line with this latter prediction, a recent study showed that presentation of information that is highly diagnostic about stimulus valence (i.e., "Niffites are peaceful, civilized, benevolent, and law-abiding; Luupites are violent, savage, malicious, and lawless") before AA training can prevent AA training effects on evaluations of these groups (Van Dessel, De Houwer, Gast, & De Schryver, 2016).

Individual difference factors should moderate Step 3 because one can be more or less fluent in the application of inference rules (and integration of specific information in evaluative inferences). We predict that participants who have more experience in using information about their own actions for making evaluative inferences (e.g., because they previously engaged in experiments where such integration was useful) should show enhanced effects. Furthermore, because integration of I_{AE} and I_{SA} in I_{SE} requires rather elaborate inference rules, we predict reduced (or even nonexistent) evaluative stimulus-action effects in organisms with less developed abilities to follow such inference rules. For example, the ability to use language may be an important factor that strongly determines the inferences an individual can make (Gentner, 2016), and integration of I_{AE} and I_{SA} in I_{SE} may require this ability. Hence, nonhuman animals that cannot integrate semantic information in their inferences should not exhibit evaluative stimulus-action effects. Note, however, that the inference of I_{SE} on the basis of I_{AE} and I_{SA} is not logically valid. We therefore predict that participants who have received formal logic training will exhibit reduced effects (given the opportunity and motivation to be accurate in evaluation—see our discussion of Step 4) because these participants might more easily infer that it is not logically valid to infer stimulus evaluation on the basis of stimulus-dependent actions.

Step 4: Stimulus evaluation. In the fourth and final step, the activation of the inference about a stimulus-evaluation relation (I_{SE}) mediates a subsequent change in stimulus evaluation. When the stimulus that was involved in the action task is encountered, an evaluative response to the stimulus is emitted on the basis of the activation of I_{SE}. For example, a person who is asked to indicate their liking of a stimulus for which they have information available that relates this stimulus to positive valence will provide a more positive evaluation of the stimulus.

Inferential process. We assume that inferential processes underlie all cognitive processes, including stimulus evaluation. In-line with the assumption that action selection is inherently goal-directed, we postulate that evaluative responses are selected on the basis of their estimated potency to produce specific (contextually activated) desired outcomes. Hence, information that is constructed in Step 3 (I_{SE}) will inform action selection depending on a person's goal to use this information for evaluation. For example, when a person is asked to indicate their liking of a stimulus, I_{SE} can bias the generation of response-related information (e.g., representation of a button press that indicates "strong liking for a stimulus") because this accords with the goal to provide a "good enough" response (and complete the experiment/communicate their feelings/ . . .). The response will be emitted when this desired outcome is predicted with sufficient precision on the basis of the activated action representation.

In contrast to associative or dual-process accounts of evaluation (e.g., Gawronski & Bodenhausen, 2006), our account postulates that both evaluations that are emitted under certain conditions of automaticity (i.e., implicit evaluations) as well as more controlled (i.e., explicit) evaluations are the result of a single, inferential, process. Importantly, however, effects on implicit and explicit evaluations can dissociate because elements of specific procedures that are used to capture implicit and explicit evaluations can facilitate the activation of distinct goals. Procedures for measuring implicit evaluations typically require speeded responding and thus facilitate the corresponding goal to provide fast responses. As a result, we assume that implicit evaluation strongly depends on inferences that are readily available under those conditions (i.e., automatic activation of propositional information: see De Houwer, 2014a). For example, in an Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), I_{SE} can influence prediction (and resulting execution) of a categorization response to the evaluative stimulus with a response key that is valence-congruent (i.e., it is also used to categorize stimuli of congruent valence) in accordance with the goal to emit a fast and accurate response. In contrast, explicit evaluation measures more strongly facilitate activation of other goals such as the goal to be accurate in evaluation or the goal to present a positive image of themselves to the experimenter (i.e., self-presentation goals). As a result, information that requires multiple steps for retrieval might be more regularly contacted and integrated in explicit evaluation (see also Cunningham, Zelazo, Packer, & Van Bavel, 2007). For example, the goal to be accurate in one's explicit evaluations may facilitate retrieval of information that is less readily available but is considered more diagnostic about stimulus valence. This should especially be the case when more easily retrieved information is not considered very diagnostic about stimulus valence.

Moderators. The translation of I_{SE} into stimulus evaluation is assumed to strongly depend on a person's current goals. Hence, evaluative stimulus-action effects should typically be strongest when a goal to evaluate the stimulus is available (e.g., by instructing participants to evaluate the stimulus in an explicit evaluation task). Importantly, however, evaluative stimulus-action effects might also arise when participants are instructed to evaluate other stimuli than the specific stimulus that was involved in the action task because this evaluation goal can be used to activate I_{SE} and integrate this information in their response. Hence, our model can explain evaluative stimulus-action effects on both explicit and implicit evaluations (see Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016).

Because implicit and explicit evaluation tasks lead to the activation of different goals, we can predict specific conditions under which implicit and explicit evaluations can dissociate. First, we expect that dissociations between implicit and explicit evaluations might depend on the registered relation between stimulus and evaluation in I_{SE}. When participants contact information that represents an identity relation between stimulus and valence (e.g., "*Stimulus A IS good"*), this should have a stronger impact on explicit evaluation than information that links stimulus and valence in a less well-specified manner (e.g., *"Stimulus A is somehow related to good")* because it is more diagnostic about stimulus valence. In contrast, implicit evaluation may be strongly influenced by both types of information. For example, there is evidence suggesting that the propositions "I am good" and "I want to be good" both strongly impact self-evaluations as measured with an IAT (Remue, Hughes, De Houwer, & De Raedt, 2014) even though only the former reflects actual self-esteem. Second, we assume that explicit measures more strongly facilitate the goal to be accurate in the evaluation of the stimulus. Because the inference of evaluation on the basis of performed actions is not logically valid, participants who contact this conclusion should show reduced evaluative stimulus-action effects on explicit evaluations whereas the reduction of effects on implicit evaluations should be less pronounced. In accordance, one recent study found a dissociation between evaluative stimulus-action effects on implicit and explicit evaluation (Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016). In this study, participants first received highly diagnostic information about the valence of the stimuli before performing an AA training task. Notably, AA training influenced implicit evaluations measured with an IAT but not explicit evaluations measured with a self-report rating scale. This pattern might be observed because participants relied on information indicating that approaching and avoiding a stimulus is not a good basis for explicit evaluation when compared with more diagnostic information. In contrast, implicit evaluations might reflect AA training contingencies because the contingency information was easily available and facilitated quick responding (it was a "good enough" response). Finally, our account also predicts that effects on explicit and implicit evaluation should be more in agreement with each other when the measurement context facilitates adoption of similar goals (e.g., when one is asked to "go with their gut" when proving explicit liking ratings: Ranganath, Smith, & Nosek, 2008; or when information is provided before IAT administration that the IAT is used to measure attitudes: e.g., Echabe, 2013).

We also assume that integration of an inference about a stimulus-evaluation relation (I_{SE}) in the evaluative response should depend on individual difference factors that impact a person's goal to integrate I_{SE} in evaluation. For example, trait reactance should reduce evaluative stimulus-action effects because a person's activation of the goal to be reactant can interfere with the goal to integrate the learned information in Step 3 in evaluation. As preliminary support for this idea, a recent study found a strong correlation between AA training effects on evaluations of well-known social groups and personal reactance measured with a trait reactance scale (Van Dessel, De Houwer, Gast, Roets, & Smith, 2018). In contrast, demand compliance should increase evaluative stimulus-action effects because the goal to comply with experimenter's demands should facilitate the goal to integrate ISE in their evaluative response. Accordingly, a recent study found that demand compliance positively correlated with AA training and AA instruction effects in the context of novel stimuli (Van Dessel, Smith et al., 2018). One could argue that changes in stimulus evaluation that are due to demand compliance are less important because they do not reflect changes in a person's "genuine" liking of the target stimuli. However, it is difficult to establish what constitutes as "genuine" liking. Moreover, it is important to note that studies have found both instruction-based and experience-based evaluative stimulus-action effects for participants who do not show demand compliance, indicating that these effects do not necessarily depend on controlled, nonautomatic processes that involve the intentional use of the acquired information for evaluation (e.g., Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016).

The Merits and Limitations of the Inferential Account

Heuristic Value

In the current article, we have argued that people show evaluative stimulus-action effects because they make a specific inference about the evaluation of a stimulus (I_{SE}) on the basis of constructed inferences about stimulus-action relations ($I_{SA)}$ and action-evaluation relations ($I_{AE)}$ and integrate this inference in stimulus evaluation. This model has heuristic value. First, it can account for the known characteristics of evaluative stimulus-action effects that are often considered as support for association formation models. Specifically, our account can explain why AA training effects sometimes occur under automaticity conditions (e.g., an unintended impact of training on liking; see Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016) and why effects can be observed on difficult to change or spontaneous behaviors such as implicit stimulus evaluations (e.g., Kawakami et al., 2007). Our model encompasses these results because it acknowledges that inferences can occur in a more or less automatic manner and because it specifies the conditions under which inferences can lead to changes in implicit evaluation (see moderation of Step 4).

Second, the heuristic value of the inferential account exceeds that of current association formation models in that it can also explain the characteristics of evaluative stimulus-action effects that do not fit well with current associative accounts of these effects. First, it can explain why studies sometimes fail to find changes in stimulus evaluation on the basis of pairings of stimuli and valenced actions (e.g., following AA training: Becker et al., 2015; Vandenbosch & De Houwer, 2011). In contrast to the idea that effects are driven by the automatic installation of associative links between stimuli and actions, we assume that (a) an inferential process chain is required that (b) depends on a number of important moderators that either enhance or impede evaluative stimulus-action effects under specific circumstances. For example, the activation of nondominant information about the relation between an action and evaluation (I_{AE}) might explain some of the observed null findings (Mertens et al., 2018; see moderation of Step 3). Second, our account can explain why stronger evaluative stimulus-action effects are observed under specific conditions that are difficult to explain from an associative perspective, such as when participants are aware of stimulus-action contingencies (Van Dessel, De Houwer, & Gast, 2016; see moderation of Step 1). Third, our account can explain recent observations that effects can arise even in the absence of *actual* performance of stimulus-based actions but on the basis of mere instructions or observation of these actions (Van Dessel et al., 2015). These effects are explained by incorporating into our model a core assumption of propositional theories, namely the assumption that propositional information about regularities in the environment can be generated on the basis of instructions, observations, imagination, or experiences with those regularities (De Houwer, 2009). Moreover, because the acquisition of propositional information is a more proximal mediator of effects than action performance, our inferential account can further explain why instruction effects can be stronger than training effects (Hughes, Van Dessel, Smith, & De Houwer, 2018); it is easier to form inferences about stimulus-action relations when the contingencies are instructed compared with when one has to discover those contingencies through trial-and-error.

Finally, the heuristic value of our inferential account extends beyond effects involving repeated performance of actions in response to a stimulus that we have focused upon so far (i.e., training-based procedures such as AA training). In contrast to most association formation accounts of evaluative stimulus-action effects, our account does not postulate that changes in evaluation require a large number of pairings of actions and stimuli. Rather, our account can also explain findings in which actions influence evaluations that arise at the time of the action (e.g., Cacioppo et al., 1993). For example, stimuli viewed during arm flexion might be rated more positively than stimuli viewed during arm extension because participants infer a stimulus-action relation (I_{SA}; e.g., that they used an approach movement when viewing the stimulus) and an action-evaluation relation (I_{AE}; e.g., that approaching is positive) and they use this information to infer a stimulus-evaluation relation (I_{SE}) and their own evaluative responses. Moreover, our account can also explain evaluative stimulus-action effects that do not involve valenced actions (e.g., AA) such as effects that were obtained in the context of the self-perception hypothesis (Bem, 1972). Our account accords with the original idea that participants infer stimulus evaluations from their actions (e.g., smiling, approaching). For example, when a person is asked to smile when watching a cartoon, they may infer that the cartoon is funny (I_{SE}) on the basis of two inferences: (a) that they are smiling in the presence of the cartoon (I_{SA}) and (b) that they tend to smile at things they like (I_{AE}; e.g., Laird, 1974).

Similarly, our account can explain findings in the field of emotion research showing that arousing actions (e.g., doing exercise) can influence evaluations (e.g., of an attractive opposite-sex confederate: White, Fishbein, & Rutstein, 1981). As argued by predictive processing theories, organisms need to be good at monitoring internal states (Seth, 2013). Information about an organism's own physiological state in the presence of a specific stimulus (I_{SA}) and the positivity or negativity associated with this state in the past (I_{AE}) may bias information generation about evaluative features of the stimulus in the present (I_{SE}), which might influence stimulus evaluation. Related to this idea, our account can also accommodate findings that provided support for cognitive dissonance theory (Festinger, 1957), which states that a person's motive to maintain cognitive consistency can give rise to irrational and sometimes maladaptive behavior. For example, in a study by Festinger and Carlsmith (1959), participants rated boring tasks as more likable if they had to persuade others that the task was fun, and especially when they received a small relative to a big sum of money for doing so. In this instance, I_{SE} may also be inferred on the basis of information about a person's actions (e.g., "I must like the task because I talked positively about it"). In the high money condition, however, there is a second reason available for participants' actions (large sum of money; and hence the evidence for liking the task—I_{SE}—is considered less valid). Note, however, that the explanation of cognitive dissonance, emotional arousal, or self-perception effects was not the primary purpose of this article, and these explanations are therefore preliminary and require further investigation.

Predictive Value

On the basis of our proposed inferential reasoning framework, we can not only explain established moderators of evaluative stimulus-action effects but also predict new moderators. Each of the four inferential processing chain steps described above involves inferential reasoning that follows the described characteristics of inferential reasoning (goal-dependent, context-dependent, learning-dependent). Hence, in the previous section outlining the processing steps, we specified many new testable predictions regarding the moderation of evaluative stimulus-action effects by specific (external and internal) contextual variables. In the studies that we recently conducted in our lab, we have already tested several predictions that were derived from the basic ideas of this account. For example, we recently found that repeated performance of actions other than AA can also lead to changes in stimulus evaluations when these actions are described in valenced terms and that these changes depend on how positive or negative participants considered these actions (I_{AE}; Van Dessel, Eder, & Hughes, 2018). However, the inferential account also leads to several interesting predictions that still need to be tested. Hence, our model will stimulate new research that is bound to increase our understanding of evaluative stimulus-action effects. In the next sections, we briefly discuss two sets of predictions that we are currently testing in ongoing research.

Moderation of evaluative stimulus-action effects by the external context. According to our account, there are many ways to contextually moderate evaluative stimulus-action effects. On one hand, each of the proposed steps can be *facilitated* based on contextual factors that should enhance effects. For example, one assumption that we are currently investigating is whether verbally providing I_{AE} (e.g., the inference that people typically approach positive stimuli and avoid negative stimuli) can lead to stronger AA training effects (because this should facilitate Step 2; Van Dessel, Hughes, De Houwer, & Smith, 2018). On the other hand, each of the steps can also be *impeded* based on contextual factors that should reduce effects. In another study, we are investigating whether evaluative stimulus-action effects are reduced when participants are informed that inferring I_{SE} on the basis of I_{AE} and I_{SA} is not in accordance with formal logical rules (which should impede Step 3; Van Dessel, Hughes, De Houwer, & Smith, 2018).

Moderation of evaluative stimulus-action effects by the internal context. Our account assumes that characteristics of the organism also strongly influence evaluative stimulus-action effects. One key assumption is that these effects should depend on the information network that people bring with them to the experimental context. As we mentioned above, individual differences in the activation of an inference about an action-evaluation relation (I_{AE}) should moderate evaluative stimulus-action effects in specific directions. We are currently investigating whether participants' belief in I_{AE} (e.g., whether they consider approaching positive) moderates evaluative stimulus-action effects. We have also started to examine in a systematic way the role of individual differences in evaluative stimulus-action effects as they relate to the different inferential steps in our model (e.g., motivation, task experience).

Influence Value

Changing stimulus evaluations. The inferential account also has important implications for influencing real-world behavior. Many clinical interventions (e.g., exposure therapies) involve (AA) actions toward stimuli (e.g., C. R. Jones, Vilensky, Vasey, & Fazio, 2013). Our account highlights new ways of improving the impact of those interventions. Specifically, we predict that the manipulation of contextual factors can change the type, number, evaluative direction, and confidence with which inferences about evaluative stimulus properties are held, which will influence stimulus evaluations (and resulting changes in real-life behavior: Ajzen & Fishbein, 2005). In a recent study, we took a first step in applying our inferential account to improve AA training effects on evaluations of food products (Van Dessel, Hughes, & De Houwer, 2018). We predicted that a training task in which AA actions produced positive or negative *consequences* (i.e., an avatar representing themselves became more or less healthy) when performed in response to specific food products would lead to stronger effects on evaluations of the food products than typical AA training paradigms (which only manipulate stimulus-action contingencies). The rationale for this prediction is that consequences would help participants more easily generate inferences that certain foods are good and other foods are bad (I_{SE}) because negative consequences that follow approaching of a stimulus (and positive consequence that follow avoidance) are more likely to lead to inferences about the valence of the stimulus than the mere fact that one consistently approaches or avoids a stimulus. The results were in-line with our prediction: consequence-based AA training effects on implicit and explicit evaluations were stronger than effects of typical AA training. These findings also imply that AA training effects can occur even when participants approach and avoid each stimulus an equal number of times (contingencies in consequence-based AA training involved stimulus, response, and action consequence; e.g., approaching one food product always led to positive outcomes and avoiding it led to negative consequences), a result that is difficult to explain on the basis of current association formation accounts. Further in-line with the inferential account, we found that AA training effects were stronger when the consequences were relevant for participants' tasks goals (i.e., when participants were instructed to make an avatar as healthy as possible). Overall, these results are consistent with our prediction that (small) changes to existing AA training procedures can help participants make better inferences and thereby facilitate changes in stimulus evaluations. On the basis of our account, we could therefore have the potential to improve existing therapies that already use these procedures for changing evaluations (e.g., Taylor & Amir, 2012).

Note that our account stops at the translation from inferences into evaluations and does not specify how changes in evaluation can lead to changes in real-life behavior (Ajzen & Fishbein, 2005). Our account could be extended to this additional aim by specifying the inferences underlying the change in behavior that results from evaluations (as well as the necessary conditions for this inferential process). We believe that this extension could be useful and might even be feasible (given the abundance of research on this topic). However, it extends beyond the aims of the current article and will be an important future endeavor.

Changing unwanted behavior. Importantly, it is also possible that inferences formed on the basis of stimulus-action contingencies can impact real-life behavior without requiring mediation via changes in evaluation. Although our model focuses on changes in evaluations, it can also be easily adapted to account for (and potentially improve) behavior that is not mediated by changes in evaluation (i.e., the effect of stimuli on evaluative responses). The inferential account postulates that stimulus-action procedures are effective at changing (pathological) behavior because they facilitate the installation and retrieval of relevant inferences (rather than installing stimulus-action associations or "response tendencies"[2]). For example, when alcoholic patients repeatedly avoid alcoholic drinks (e.g., Wiers et al., 2011), they might make specific inferences (e.g., they infer that these drinks are to-be-avoided or that they are able to avoid alcoholic drinks). Once established, this propositional (rather than associative) knowledge might be contacted in other contexts, allowing these patients to refrain from drinking. Similarly, when people consistently avoid unhealthy foods, they might infer that these stimuli are bad for them, and this may inform food choices. In the study described in the previous paragraph (Van Dessel, Hughes, & De Houwer, 2018), we found evidence for this idea. Participants who had performed the consequence-based AA training in which approaching unhealthy foods led to negative outcomes and avoiding unhealthy foods led to positive outcomes, reported eating less unhealthily in the days after the intervention (but not participants who had performed typical AA training without consequences) and actually ate less unhealthy snacks in an ad libitum snack task. It is possible that these effects occurred because the consequence-based AA training more strongly facilitated the (adaptive) inference that it is good to refrain from unhealthy foods.

These preliminary results point to an important promise for training-based action interventions that has been claimed ever since these effects were first observed: that action training (especially AA training) may influence not only difficult-to-change implicit evaluations (e.g., implicit prejudice: Kawakami et al., 2007) but also difficult-to-change behaviors (e.g., addictive behavior: Wiers et al., 2011). Unlike the traditional idea that such change occurs because these tasks lead to associative changes via repeated pairings, we assume that these procedures actually help people make specific inferences. Because these inferences are self-generated (i.e., a person needs to infer the information on the basis of the contingency information themselves), the information may be more strongly integrated in a person's information network than information that is merely provided to a person (see also research on behavioral nudging: Benartzi et al., 2017). The person has already made the inference (and may have done so several times, depending on the extensiveness of the training) which might allow for the inferences to be easily repeated.

Note that current training procedures might not only be improved by including consequences of stimulus-action relations and making those consequences goal-relevant but also by changing other procedural details that promote the installation of novel (and adaptive) inferences. Such an inference-based therapy is much closer to therapies often used in clinical practice (e.g., cognitive behavioral therapy) than typical cognitive bias modification therapies that aim to establish stimulus-action tendencies. However, they might add an automatization component to the typical therapies used in clinical practice (because they involve training or repeated generation of certain inferences). This "inference training" could be an important new method for establishing important clinical effects (especially seeing as current "cognitive bias modification trainings" to change clinical behavior often do not produce beneficial effects; e.g., Cristea, Kok, & Cuijpers, 2015; A. Jones, Hardman, Lawrence, & Field, 2017).

Relation to Other Accounts

Our inferential account of evaluative stimulus-action effects is the first to provide an elaborate account of evaluative learning effects from an inferential (and propositional) perspective (i.e., an account that makes underlying processes explicit). In doing so, we have drawn on a number of more general accounts of human learning and mental processing. First, our account makes many of the same assumptions as single-process propositional accounts of learning (e.g., Mitchell et al., 2009). Stimulus-action effects are explained with reference to a single memory system that involves the activation of propositional information and use of this information in inferential reasoning. These inferential processes operate on the basis of information that encodes not only co-variation but also the relational properties of concepts. Second, our model draws on predictive processing models (e.g., Feldman & Friston, 2010). Although these models currently enjoy widespread appeal elsewhere in psychological science, this idea has yet to find its way into evaluative learning research. Our model integrates the idea that the making and updating of predictions (on the basis of Bayesian rules) might underlie (action) effects on stimulus evaluation and makes assumptions on the basis of dominant ideas in the predictive processing literature. For example, our account builds on the idea that activation level of information is an important concept for explaining human behavior in general and reasoning in particular (Sanborn & Chater, 2016; see also atomic components of thought-rational theory: Banks, 2013).

Our model combines single-process propositional accounts of learning with predictive processing accounts and applies this specifically to evaluative stimulus-action effects. The idea that people infer stimulus evaluations from their actions is in-line with self-perception theory (Bem, 1972). However, our account provides more detail about the underlying conditions. Moreover, it is broader to the extent that it not only explains effects that arise when emotional responses are ambiguous but also the abundance of recent findings on training-based effects. Furthermore, it provides an explanation for effects on implicit and explicit evaluation. In doing so, we build on recent findings and recent theorizing indicating that propositional processes underlie not only explicit but also implicit evaluation (De Houwer, 2014a; Mann & Ferguson, 2015). Again, however, we formalize this idea in an inferential model by building on general accounts of human cognition, specifying clear assumptions, and providing testable predictions.

Our inferential account also bears similarity to the theory of event coding (TEC; Hommel, Müsseler, Aschersleben, & Prinz, 2001) to the extent that both theories assume that action performance is critically dependent on anticipated action consequences. Crucially, however, the TEC (and the common-coding account of AA training effects of Eder & Klauer, 2009, which is derived from this theory) assumes that the automatic formation of associations between action representations and perceptual consequences mediates action performance. In contrast, our inferential account combines principles from predictive processing theory and TEC to explain evaluative behavior on the basis of inferential reasoning (see Butz, 2016, for an integrative theory of human cognition in general). A recent study that pitted predictions of the common-coding and inferential account of AA training effects against each other provided stronger support for the inferential account which predicted AA effects on the basis of mere action observation (Van Dessel, Eder, & Hughes, 2018).

Limitations of the Inferential Account

The most important limitation of our account may be that it is still a relatively general account at this point (though this can also be seen as a strength of theories: Gawronski & Bodenhausen, 2015; Meiser, 2011). We have specified many different boundary conditions and moderators (which increases the falsifiability of our account), the processing steps necessary to produce changes in evaluation, and important details about how inferential reasoning might occur. Yet, there are still unconstrained factors. Most importantly, we have not specified (all) the specific inferences (e.g., I_{AE}) that might be involved in evaluative stimulus-action effects. Our reason for this is that there is much variation in the specific inferences that individuals make (e.g., because this depends on the availability of information in a person's information network which may strongly differ between people). Indeed, in a recent study, we asked participants to indicate why they would rate an approached stimulus as more positive than an avoided stimulus. We found marked variability in the specific responses obtained. Many of the provided reasons were not highly elaborated (e.g., "avoiding a stimulus means that something is wrong with it") suggesting that participants often use heuristic rules (e.g., an availability heuristic) to infer stimulus evaluation on the basis of AA training (and instructions).

Because of the complexity and uncertainty with regard to the inferences people might contact, propositional or inferential models are sometimes thought to be inferior to more simple and/or domain specific models (e.g., associative models which assume that evaluative stimulus-action effects occur on the basis of simple links that are formed as the result of repeated pairings). Note, however, that parsimony should not come at the cost of explanatory or influence value. Our model can be considered useful because it has high heuristic and predictive value (i.e., it allows us to explain existing findings, predict novel findings, and influence human behavior in novel ways). Note also that although individual association formation accounts might be considered simple (e.g., Phills et al., 2011), these accounts cannot explain many of the relevant findings in this literature (see above). As a class, however, association formation models have a high degree of flexibility, which allows proponents of these models to always make post hoc adaptations to explain obtained results. However, different findings require different adaptations that are often logically inconsistent. In sum, our model leads to a whole set of a priori predictions for which it is difficult to see how a similar set of predictions could be made on the basis of any individual association formation model or subclass of logically consistent association formation models. It is also possible that a dual-process framework incorporating inferential and associative processes could explain evaluative stimulus-action effects. However, an account that can explain these effects on the basis of one coherent processing mechanism (i.e., inferential reasoning) should be preferred over an account that additionally postulates the existence of an entirely different second mechanism (e.g., association formation) for reasons of parsimony.

It is also important to note that by providing a predictive processing framework for our account, we are directing it away from accounts that require a strong "homunculus" factor. We provide an explanation of how evaluative stimulus-action effects might occur in a more or less automatic or uncontrolled manner and under which circumstances these effects should arise. In doing so, we clarify that effects that occur under specific automaticity conditions do not necessarily require associative explanations. Inferential accounts can explain all aspects of evaluative stimulus-action effects—not only those that are highly controlled (e.g., effects resulting from demand compliance). Moreover, we also clarify that inferential reasoning can operate on the basis of (neurologically plausible) mental mechanisms that have specific characteristics. This also opens up the possibility of eventually specifying a more elaborate mathematical model of these complex effects by implementing Bayesian probability calculus (Friston, 2003).

In providing this inferential framework for evaluative stimulus-action effects, we have also opened up new avenues for the construal of novel accounts of other (evaluative) learning phenomena (e.g., EC) that might operate on the basis of similar processes. These accounts distinguish themselves from (current) propositional accounts in that they provide an explanation based on the generation of inferences (a subset of propositions) and that they provide an elaborate explanation of how this mechanism might occur. Of course, it is important to appreciate that our account provides only an initial framework (applied to one specific type of effects). Because we do not want to extend our model too far beyond the data, we think that further specification should come on the basis of empirical findings.

Concluding Remarks

In the current article, we have described a model of evaluative stimulus-action effects that draws on inferential rather than simple associative processes. We have specified the process steps that underlie these effects as well as how the processes underlying these steps might work and under what conditions they operate. We hope that this new framework can help improve our understanding of evaluative stimulus-action effects (and other evaluative learning phenomena) and further improve the utility of (action-based) procedures in clinical domains.

Acknowledgments

We thank Andreas Eder, Yannick Boddez, Baptist Liefooghe, and Senne Braem for their comments on an earlier version of this manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Pieter Van Dessel is supported by a Postdoctoral fellowship of the Scientific Research Foundation, Flanders (FWO-Vlaanderen). Jan De Houwer is supported by Methusalem Grant BOF16/MET_V/002 of Ghent University.

Notes

1. Note that inferential models—as a class of models—do not necessarily postulate that inferences are based on predictive processing. We do, however, add this assumption to our own account because it adds precision and leads to many novel predictions.
2. Note that automatic response tendencies often do not logically relate to the pathological behaviors under investigation in cognitive bias modification research (Snelleman, Schoenmakers, & de Mheen, 2015; Spruyt et al., 2013) which contrasts with predictions of association formation accounts. In contrast to these accounts, our inferential account does not assume that the action tendencies produce the unwanted behavior. Rather, action tendencies (i.e., fast responses in an approach-avoidance measurement tasks) might be determined by the speed of predictions of stimulus-based actions. This can be based on learned contingencies in Step 1 of the process chain or other learned information (e.g., "I'm required to avoid alcohol") and should not directly cause the unwanted behavior.

ORCID iD

Pieter Van Dessel https://orcid.org/0000-0002-3401-780X

References

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes and attitude change* (pp. 173-221). Mahwah, NJ: Lawrence Erlbaum.

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

Banks, A. P. (2013). The influence of activation level on belief bias in relational reasoning. *Cognitive Science*, *37*, 544-577. doi:10.1111/cogs.12017

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*, 695-711. doi:10.1016/j.neuron.2012.10.038

Bayes, T. (1958). Studies in the history of probability and statistics: IX. Thomas Bayes' (1763) essay "Towards solving a problem in the doctrine of chances" in modernized notation. *Biometrika*, *45*, 296-315.

Becker, D., Jostmann, N. B., Wiers, R. W., & Holland, R. W. (2015). Approach avoidance training in the eating domain: Testing the effectiveness across three single session studies. *Appetite*, *85*, 58-65. doi:10.1016/j.appet.2014.11.017

Becker, E. S., Ferentzi, H., Ferrari, G., Möbius, M., Brugman, S., Custers, J., . . . Rinck, M. (2016). Always approach the bright side of life: A general positivity training reduces stress reactions in vulnerable individuals. *Behaviour Research and Therapy*, *40*, 57-71. doi:10.1007/s10608-015-9716-2

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (*Vol. 6*, pp. 1-62). New York, NY: Academic Press.

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., . . . Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, *28*, 1041-1055. doi:10.1177/0956797617702501

Boddez, Y., De Houwer, J., & Beckers, T. (2017). The inferential reasoning theory of causal learning: Towards a multi-process propositional account. In M. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 53-64). Oxford, UK: Oxford University Press.

Brighton, H., & Gigerenzer, G. (2012). Homo heuristicus: Less-is-more effects in adaptive cognition. *Malaysian Journal of Medical Science*, *19*, 6-16.

Burt, C. (1911). Experimental tests of higher mental processes and their relation to general intelligence. *Journal of Experimental Pedagogy*, *1*, 93-112.

Butz, M. V. (2016). Toward a unified sub-symbolic computational theory of cognition. *Frontiers in Psychology*, *7*, 925.

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes. II: Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology*, *65*, 5-17. doi:10.1037/0022-3514.65.1.5

Centerbar, D. B., Schnall, S., Clore, G. L., & Garvin, E. D. (2008). Affective incoherence: When affective concepts and embodied reactions clash. *Journal of Personality and Social Psychology*, *94*, 560-578. doi:10.1037/0022-3514.94.4.560

Chetverikov, A., & Kristjánsson, Á. (2016). On the joys of perceiving: Affect as feedback for perceptual predictions. *Acta Psychologica*, *169*, 1-10.

Cisek, P., & Pastor-Bernier, A. (2014). On the challenges and mechanisms of embodied decisions. *Philosophical Transactions of the Royal Society*, *369*, Article 20130479. doi:10.1098/rstb.2013.0479

Cleeremans, A. (2014). Connecting conscious and unconscious processing. *Cognitive Science*, *38*, 1286-1315.

Cristea, I. A., Kok, R. N., & Cuijpers, P. (2015). Efficacy of cognitive bias modification interventions in anxiety and depression: Meta-analysis. *The British Journal of Psychiatry*, *206*, 7-16. doi:10.1192/bjp.bp.114.146761

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, *25*, 736-760. doi:10.1521/soco.2007.25.5.736

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. (2014a). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*, 342-353. doi:10.1111/spc3.12111

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

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*, 252-287. doi:10.1080/10463283.2014.892320

Echabe, A. E. (2013). Relationship between implicit and explicit measures of attitudes: The impact of application conditions. *Europe's Journal of Psychology*, *9*, 231-245.

Eder, A. B., & Klauer, K. C. (2009). A common-coding account of the bidirectional evaluation–behavior link. *Journal of Experimental Psychology: General*, *138*, 218-235. doi:10.1037/a0015220

Evans, J. S. B. T. (2010). Intuition and reasoning: A dual process perspective. *Psychological Inquiry*, *21*, 313-326.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, Article 215. doi:10.3389/fnhum.2010.00215

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, *58*, 203-210.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48-58. doi:10.1038/nrn2536

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*, 1325-1352.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), e1000211.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127-138.

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. (2015). *Theory and explanation in social psychology*. New York, NY: Guilford Press.

Gawronski, B., Bodenhausen, G. V., & Becker, A. P. (2007). I like it, because I like myself: Associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology*, *43*, 221-232. doi:10.1016/j.jesp.2006.04.001

Gentner, D. (2016). Language as cognitive tool kit: How language supports relational thought. *American Psychologist*, *71*, 650-657.

Gentner, D., & Smith, L. A. (2013). Analogical learning and reasoning. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 668-681). New York, NY: Oxford University Press.

George, D., & Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. *PLoS Computational Biology*, *10*, e1000532. doi:10.1371/journal.pcbi.1000532

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (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

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition*, *65*, 197-230.

Helmholtz, H. (1962). *Handbuch der physiologischen optic* (J. P. C. Southall, Trans.). New York, NY: Dover.

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

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 849-937.

Hsee, C., Tu, Y., Lu, Z., & Ruan, B. (2014). Approaching aversion: Negative reactions toward approaching stimuli. *Journal of Personality and Social Psychology*, *106*, 699-712. doi:10.1037/a0036332

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

Hughes, S., Van Dessel, P., Smith, C. T., & De Houwer, J. (2018). *A comparative analysis of different evaluative learning procedures*. Manuscript in preparation.

Hummel, J. E. (2010). Symbolic versus associative learning. *Cognitive Science*, *346*, 958-965. doi:10.1111/j.1551-6709.2010.01096.x

Jones, A., Hardman, C. A., Lawrence, N., & Field, M. (2018). Cognitive training as a potential treatment for overweight and obesity: A critical review of the evidence. *Appetite*, *124*, 50-67. doi:10.1016/j.appet.2017.05.032

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

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

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

Kawakami, K., Steele, J. R., Cifa, C., Phills, C. E., & Dovidio, J. F. (2008). Approaching math increases math = me and math = pleasant. *Journal of Experimental Social Psychology*, *44*, 818-825. doi:10.1016/j.jesp.2007.07.009

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgements are based on common principles. *Psychological Review*, *118*, 97-109.

Krypotos, A.-M., Arnaudova, I., Effting, M., Kindt, M., & Beckers, T. (2015). Effects of approach-avoidance training on the extinction and return of fear responses. *PLoS ONE*, *10*(7), e0131581. doi:10.1371/journal.pone.0131581

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. 154-172). Oxford, UK: Oxford University Press.

Laham, S. M., Kashima, Y., Dix, J., Wheeler, M., & Levis, B. (2014). Elaborated contextual framing is necessary for action-based attitude acquisition. *Cognition and Emotion*, *28*, 1119-1126. doi:10.1080/02699931.2013.867833

Laird, J. D. (1974). Self-attribution of emotion: The effects of expressive behavior on the quality of emotional experience. *Journal of Personality and Social Psychology*, *29*, 475-486. doi:10.1037/h0036125

Lovibond, P. F. (2003). Causal beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 97-106.

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology*, *108*, 823-849.

Meiser, T. (2011). Much pain, little gain? Paradigm-specific models and methods in experimental psychology. *Perspectives on Psychological Science*, *62*, 183-191.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*, 280-293. doi:10.1016/j.tics.2018.02.001

Mertens, G., Van Dessel, P., & De Houwer, J. (2018). The contextual malleability of approach-avoidance training effects: Approaching or avoiding fear conditioned stimuli modulates effects of approach-avoidance training. *Cognition and Emotion*, *32*, 341-349. doi:10.1080/02699931.2017.1308315

Metzinger, T., & Wiese, W. (2017). *Philosophy and predictive processing*. Frankfurt, Germany: MIND Group.

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. (2014). Examining the mapping problem in dual process models. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 20-34). New York, NY: Guilford Press.

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*, *67*, 263-287.

Moors, A., Boddez, Y., & De Houwer, J. (2017). The power of goal-directed processes in the causation of emotional and other actions. *Emotion Review*, *9*, 310-318.

Olson, J. M., & Stone, J. (2005). The influence of behavior on attitudes. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *Handbook of attitudes and attitude change* (pp. 223-271). Mahwah, NJ: Lawrence Erlbaum.

Penny, W. (2012). Bayesian models of brain and behaviour. *Biomathematics*, *2012*, Article 785791. doi:10.5402/2012/785791

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17-35.

Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., & Inzlicht, M. (2011). Mind the gap: Increasing associations between the self and Blacks with approach behaviors. *Journal of Personality and Social Psychology*, *100*, 197-210. doi:10.1037/a0022159

Proulx, M. J. (2014). The perception of shape from shading in a new light. *PeerJ*, *2*, e363. doi:10.7717/peerj.363

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology*, *44*, 386-396.

Remue, J., Hughes, S., De Houwer, J., & De Raedt, R. (2014). To be or want to be: Disentangling the role of actual versus ideal self in implicit self-esteem. *PLoS ONE*, *9*, e108837.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *2012*, 883-893.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*, 565-573. doi:10.1016/j.tics.2013.09.007

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

Snelleman, M., Schoenmakers, T. M., & de Mheen, D. V. (2015). Attentional bias and approach/avoidance tendencies do not predict relapse or time to relapse in alcohol dependency. *Alcoholism-Clinical and Experimental Research*, *9*, 1734-1739. doi:10.1111/acer.12817

Spruyt, A., De Houwer, J., Tibboel, H., Verschuere, B., Crombez, G., Verbanck, P., . . . Noël, X. (2013). On the predictive validity of automatically activated approach/avoidance tendencies in abstaining alcohol-dependent patients. *Drug and Alcohol Dependence*, *127*, 81-86. doi:10.1016/j.drugalcdep.2012.06.019

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220-247. doi:10.1207/s15327957pspr0803_1

Taylor, C. T., & Amir, N. (2012). Modifying automatic approach action tendencies in individuals with elevated social anxiety symptoms. *Behaviour Research and Therapy*, *50*, 529-536. doi:10.1016/j.brat.2012.05.004

Vandenbosch, K., & De Houwer, J. (2011). Failures to induce implicit evaluations by means of approach-avoid training. *Cognition and Emotion*, *25*, 1311-1330. doi:10.1080/02699931.2011.596819

Van Dessel, P., De Houwer, J., & Gast, A. (2016). Approach-avoidance training effects are moderated by awareness of stimulus-action contingencies. *Personality and Social Psychology Bulletin*, *42*, 81-93. doi:10.1177/0146167215615335

Van Dessel, P., De Houwer, J., Gast, A., Roets, A., & Smith, C. T. (2018). *Comparing effects of AA instructions and AA training on racial prejudice*. Manuscript in preparation.

Van Dessel, P., De Houwer, J., Gast, A., & Smith, C. T. (2015). Instruction-based approach-avoidance effects: Changing stimulus evaluation via mere instruction to approach or avoid stimuli. *Experimental Psychology*, *62*, 161-169. doi:10.1027/1618-3169/a000282

Van Dessel, P., De Houwer, J., Gast, A., Smith, C. T., & De Schryver, M. (2016). Instructing implicit processes: When instructions to approach or avoid influence implicit but not explicit evaluation. *Journal of Experimental Social Psychology*, *63*, 1-9. doi:10.1016/j.jesp.2015.11.002

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

Van Dessel, P., De Houwer, J., & Smith, C. T. (2018). Relational information moderates approach-avoidance instruction effects on implicit evaluation. *Acta Psychologica*, *184*, 137-143. doi:10.1016/j.actpsy.2017.03.016

Van Dessel, P., Eder, A. B., & Hughes, S. (2018). Mechanisms underlying approach-avoidance training effects on stimulus evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1224-1241.

Van Dessel, P., Hughes, S., & De Houwer, J. (in press). Consequence-based approach-avoidance training: A new and improved method for changing unwanted behavior. Psychological Science.

Van Dessel, P., Hughes, S., De Houwer, J., & Smith, C. T. (2018). *Testing an inferential model of approach-avoidance training effects on stimulus evaluation*. Manuscript in preparation.

Van Dessel, P., Smith, C. T., De Houwer, J., Hussey, I., Hughes, S., De Neve, J., & Perugini, M. (2018). *The role of awareness in effects of different learning pathways on evaluations of novel stimuli*. Manuscript in preparation.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale, NJ: Lawrence Erlbaum.

Wells, G. L., & Petty, R. E. (1980). The effects of overt head movements on persuasion: Compatibility and incompatibility of responses. *Basic and Applied Social Psychology*, *1*, 219-230.

White, G. L., Fishbein, S., & Rutstein, J. (1981). Passionate love and misattribution of arousal. *Journal of Personality and Social Psychology*, *41*, 56-62.

Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science*, *22*, 490-497. doi:10.1177/0956797611400615

Wittekind, C. E., Feist, A., Schneider, B. C., Moritz, S., & Fritzsche, A. (2015). The approach-avoidance task as an online intervention in cigarette smoking: A pilot study. *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 115-120. doi:10.1016/j.jbtep.2014.08.006

Woud, M. L., Becker, E. S., & Rinck, M. (2011). Induction of implicit evaluation biases by approach-avoidance training: A commentary on Vandenbosch and De Houwer. *Cognition and Emotion*, *25*, 1331-1338. doi:10.1080/02699931.2011.628300

Zogmaister, C., Perugini, M., & Richetin, J. (2016). Motivation modulates the effect of approach on implicit preferences. *Cognition and Emotion*, *30*, 890-911. doi:10.1080/02699931.2015.1032892