Running head: ATTITUDINAL ENTROPY REVISITED

The Attitudinal Entropy Framework revisited: Increased conceptual precision is needed if the framework is to succeed

Pieter Van Dessel, Sean Hughes, Ian Hussey, and Jan De Houwer

Ghent University, Belgium

**Author Note**

PVD, SH, IH, & JDH, Department of Experimental Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, Ghent, Belgium. PVD is supported by a postdoctoral fellowship of the Scientific Research Foundation, Flanders (FWO-Vlaanderen). IH is supported by a postdoctoral fellowship from Ghent University. JDH and SH are supported by Methusalem Grant BOF16/MET\_V/002 of Ghent University to JDH. Correspondence concerning this article should be sent to [Pieter.vanDessel@UGent.be](mailto:Pieter.vanDessel@UGent.be)

The Attitudinal Entropy Framework Revisited: Increased Conceptual Precision is Needed if the Framework is to Succeed

Dalege, Borsboom, van Harreveld, and van der Maas (this issue) describe a novel framework for the conceptualization of attitudes that draws on principles from statistical mechanics. A core idea in their framework is that systems are often characterized by randomness (i.e., entropy) and that there is both heuristic and predictive value in applying the idea of entropy to the study of attitudes and related phenomena. We applaud their initiative: the attitudinal entropy framework provides an intriguing new perspective on theoretical questions and empirical findings in social psychology. It opens up new avenues for research in many areas and is a timely contribution given the growing popularity of predictive processing theories emphasizing entropy as an important factor in human cognition (for a recent overview see Metzinger & Wiese, 2017). These theories assume that people strive to minimize entropy by building a mental model of the world that enables them to optimally respond to the environment (Friston, 2010). Application of principles of predictive processing theories (e.g., entropy) to social cognition is currently lacking (see Van Dessel, Hughes, & De Houwer, 2018, for a discussion).

Nevertheless, and despite our shared appreciation for the framework, we believe that there is still room for improvement. We see four issues that need to be addressed going forward.

**Issue 1: The Framework Needs to be Better Situated Relative to other Psychological Theories**.

Dalege and colleagues devote much of their paper to discussing how findings from many areas of social psychology can be conceptualized and simulated by their framework (e.g., the mere thought effect, persuasion, heuristic cues, attitude ambivalence). Yet, relatively less attention is paid to how their framework fits in the larger eco-system of (social) psychological theories (i.e., how it extends beyond previous domain-specific models and accounts). For instance, it is unclear which of the predictions it makes are truly novel. Towards the end of their paper, the authors do select three models which they argue are “similar in focus to the AE” (p.41): the Iterative Reprocessing (IR) model (Cunningham & Zelazo, 2007), the Attitude as Constraint Satisfaction (ACS) model (Monroe & Read, 2008) and the Associative Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006). They highlight that their framework “is more in-line with the IR model and ACS model, which both assume that implicit and explicit evaluations are based on the same process”, unlike the APE model in which “explicit processing of the attitude object results from a process that is qualitatively different from implicit processing of the attitude” (p.41).

This single process perspective on attitudes also bears similarity to an inferential model of evaluative stimulus-action effects that we recently introduced (Van Dessel et al., 2018). Unlike the APE and other models the authors considered, our inferential account is one of the first to link predictive processing theories to the study of attitudes. It focuses on the inferences that underlie evaluative learning on the basis of stimulus-based actions (e.g., repeated approach or avoidance of a stimulus) and outlines how these inferences might arise based on predictive processing principles. [[1]](#footnote-2) Specifically, evaluative responding is considered to result from inferences about (the value of) action outcomes. These inferences are learning-, context, and goal-dependent, and reflect the (automatic) application of inference rules to activated information on the basis of a person’s belief network. This belief network can be seen as a generative model of the world that is continuously updated on the basis of available information.

The Attitudinal Entropy framework and our inferential model share several similarities with one another. For instance, the former argues that entropy (and its reduction) may play a key role determining the structure and properties of attitudes, a claim that is certainly compatible with the inferential account given its incorporation of predictive processing theory (see Issue 4). Second, the Attitudinal Entropy framework seems to share the position that implicit and explicit attitudes are based on a single type of mental process that involves inferential reasoning. For instance, Dalege and colleagues note that “weights between attitude elements generally arise based on inferences” (p.12). Moreover, assessing for entropy (which they conceptualize in part as consistency between attitude elements) presumably requires the mental system to be able to represent the truth value of attitude elements (and relations between these elements). This perspective is compatible with single process (propositional) models of attitudes and learning (De Houwer, 2009; Mitchell, De Houwer, & Lovibond, 2009) and diverges from models which distinguish between two types of attitudinal processes or systems: e.g., System 1 vs 2 (Kahneman, 2003), associative vs. rule-based processes (Smith & DeCoster, 2000), or associative and propositional processes (e.g., Gawronski & Bodenhausen, 2006). It also accords with recent recommendations to explore alternatives to dual-process theories of human cognition (e.g., Melnikoff & Bargh, 2018), a call which is especially relevant to attitude research where such theories remain dominant and often in the absence of clear empirical support (see Corneille & Stahl, 2018).

That said, there are several points of divergence between our inferential model and the attitudinal entropy framework. We will highlight these where appropriate in the following sections. For now, we conclude that clarification is needed about how the attitudinal entropy framework is situated in relation to other (social cognition) theories to elucidate connections and divergences and highlight the novel and important aspects of this framework.

**Issue 2: A More Nuanced Perspective and Account of Attitudes is Needed**

The Attitudinal Entropy framework would also benefit from the incorporation of recent developments at the conceptual level that have taken place within the attitudes literature. De Houwer, Gawronski, & Barnes-Holmes (2013) recently offered a meta-theoretical framework according to which attitude research can be conceptualized as the scientific study of evaluation. Evaluation is defined not in terms of mental constructs but in terms of elements in the environment, more specifically, as the effect of stimuli on evaluative responses. From this perspective, attitude research provides answers to two questions: (1) “Which elements in the environment moderate evaluation?” and (2) “What mental processes and representations mediate evaluation?” Research on the first question provides explanations of evaluative responses in terms of elements in the environment (functional level of analysis); research on the second question offers explanations of evaluation in terms of mental processes and representations (mental level of analysis). These two levels of analysis are mutually supportive, in that better explanations at one level lead to better explanations at the other level. However, their mutually supportive relation requires a clear distinction between the concepts of their explanans (that which explains) and explanandum (that which is to be explained), which are conflated if behaviors are treated as proxies for mental constructs (e.g., when they are viewed as attitudes or attitude elements).

One concrete example of this conflation of behavior and mediating mental mechanism can be found in Dalege and colleagues’ simulation 1a of effects observed on the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The authors argue that IAT effects are unstable and of limited predictive utility at the individual level, but that mean scores between participants show higher stability and predictive validity. It is worth noting that this premise is not reflective of a broad understanding of the literature on implicit measures: only a small number of recent papers have demonstrated this stability and predictive utility at the group level, and, within some domains at least, the IAT has repeatedly demonstrated predictive utility at the individual level (e.g., the domain of suicidality: Barnes et al., 2016; Nock & Banaji, 2007; Nock, Park, Finn, Deliberto, Dour, & Banaji, 2010; Randall, Rowe, Dong, Nock, & Colman, 2013; or intergroup behavior: Kurdi et al., in press). This aside, a key point here is that Dalege and colleagues attempt to model the behavioral effects observed on implicit measures (explanandum), while their model is discussed as an account of the attitudes that are the mediating mental mechanisms of such overt behaviour (explanans). This issue is compounded in simulation 1b where the unit of analysis that nodes within the network refer to is switched from attitude elements within a cognitive system (an intrapersonal model) to individuals in a group (an interpersonal model). The authors argue that these simulations, when seen from a distance, model phenomena relating to implicit attitudes. However, upon scrutiny, a lack of clarity around what level of analysis is being modeled (i.e, behaviour vs. mental level, intrapersonal vs. interpersonal) raises more questions than it answers.

The inferential model that we recently specified (Van Dessel et al., 2018) adopts the perspective of De Houwer et al. (2013). Specifically, we model evaluations (rather than attitudes), which we define as overt behavioral responses. This ensures that there is no conflation between the behaviors that need to be explained and the mental constructs that are used to explain these behaviors (inferences). We believe that adopting such a perspective would also benefit the Attitudinal Entropy framework. It would provide it with a clear definition of attitudes (something that Dalege and colleagues acknowledge that they are still searching for; “the exact nature of attitudinal elements needs to be further investigated”, p. 42), increase conceptual clarity within their framework (ensure that the thing used to explain [attitudes] is kept separate from the thing that needs to be explained [evaluative responses]), and allow for clear, testable predictions about the moderation of evaluative responses by specific contextual variables. Indeed, as it is conceived now, the author’s framework focuses mainly on the mental level of analysis (explanations of evaluation in terms of mental processes and representations) and says relatively little about the functional level of analysis (explanations of evaluative responses in terms of elements in the environment). It also assumes a more or less direct link between mental mechanisms (attitudes) and evaluative responding. Yet decades of research suggests that such a relation is unwarranted, and that the strength, nature, and number of evaluative responses is moderated by a wide variety of contextual (environmental) factors (Schwarz, 2007). Separation of the functional and cognitive level of explanation will improve both the depth and breadth of the model, for instance, allowing it to clearly stipulate how attitude elements might arise (from interaction with the environment), which is currently an important missing element.

**Issue 3: A More Nuanced Perspective and Account of Automaticity is Needed**

The Attitudinal Entropy framework would also benefit from a more nuanced perspective on automaticity. Automaticity can be seen as an umbrella term for a group of operating conditions (speed, efficiency, awareness, intent) under which a mental process operates (or as a set of conditions which allow for a process to occur or not) (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009; Moors & De Houwer, 2006). When applied to the study of attitudes, this means that attitudes can be encoded, stored, or retrieved under some but not other automaticity conditions, and that researchers need to clearly specify in what sense their measures are ‘implicit’ or ‘explicit’.

In their target article, Dalege and colleagues s suggest that “implicit measures…limit attention directed at the attitude object by measuring attitudes without directly asking individuals to introspect” (p.18) and that “the construct measured by implicit measures is itself more internally inconsistent than the construct measured by explicit measures because the former by their very nature direct less attention towards the attitude object than the latter” (p.20). This perspective is problematic because it reduces automaticity to a single dimension (attention) and presupposes that measures such as the IAT are ‘implicit’ simply because they reduce attention to the attitude object. Yet, attention is only one of several dimensions relevant to automaticity, and one that is not necessarily reduced in implicit measures. Research shows that people are often very aware of the attitude object and their responses to it in the IAT (Hahn, Judd, Hirsh, & Blair, 2014). Other research has found that attention to different aspects of the attitude object (e.g., its different semantic categories) influences the type of responses emitted towards that object (Gawronski, Cunningham, LeBel, & Deutsch, 2010) and that attention to certain (evaluative) properties of the attitude object is sometimes necessary to observe implicit attitudes (Spruyt & Tibboel, 2015). Still more work shows that ‘top-down’ processes (e.g., related to goals: Spruyt, Tibboel, De Schryver, & De Houwer, 2017) play a more important role in implicit attitudes than is often appreciated. Simply put, ‘attention’ is just one component of automaticity, there is no guarantee that implicit measures reduce it, nor is there any guarantee that it is the most important feature of automaticity when it comes to implicit attitudes. Moreover, it is possible that different automaticity features play a role in different types of (implicit) evaluations and that the same evaluation can operate under certain conditions in certain situations and different conditions in different situations. We would therefore encourage Dalege and colleagues to consider a more nuanced perspective on this issue and integrate it into their framework (e.g., how would the framework operationalize and deal with other automaticity conditions such as speed, efficiency, and intent?). Doing so would likely expand and improve the sophistication of the predictions that their model would make about implicit measures (now predictions about the stability and consistency of implicit measures are limited to how much a person thinks about the attitude object).

**Issue 4: The Conceptualization of (Attitudinal) Entropy Needs Further Refinement**

Finally, the Attitudinal Entropy framework would benefit from additional clarity around the entropy concept. Dalege and colleagues describe entropy as a characteristic property of attitudes (i.e., its uncertainty). Boltzmann entropy is formalized as the proportion of attitude elements in distinct states at a certain point in time whereas Gibbs entropy is formalized as the consistency of these configurations over time. Thinking about (or attention to) the attitude object is assumed to reduce Gibbs entropy. This formalization has merits because it is more elaborate compared to other (e.g., predictive processing) theories. In fact, it has been noted that the conceptualization of entropy in the predictive processing framework is implausible and requires more work (e.g., Otworowska, Van Rooij, & Kwisthout, 2018). Moreover, it allows simulation of human evaluative behavior in a nodal network with a good fit to several past findings and potential for novel predictions.

Importantly, however, even though consistency detection lies at the basis of the attitudinal entropy theory, it is not specified how this process occurs. Attitude elements are modeled as nodes that can only be switched on or off and are thus stripped from any (relational) content (e.g., the content of beliefs), making it difficult to see how consistency between attitude elements could be determined. The assumption that only the (momentary) valence of attitude elements (modeled as a binary variable) is compared in this process is unfeasible given that it is not specified how the valence of attitude elements (not only beliefs but also behaviors and feelings) is determined. Moreover, studies show that content-related characteristics of information about attitude objects (e.g., its diagnosticity or believability: Cone, Mann, & Ferguson, 2018) determines evaluation more than the amount of positive and negative information. For instance, Cone and Ferguson (2015) found that participants exhibited negative implicit and explicit evaluations of a person named Bob when they learned many pieces of positive information about Bob but one piece of negative information that was more diagnostic of Bob’s true character (e.g., that Bob was a child molester).

Finally, while the attitudinal entropy framework provides a number of answers by drawling parallels with physical and biological systems, the application of the concept of entropy to psychological systems also differs in a key way that may undermine the utility of the analogy. The authors argue that attitudes are determined by an attempt to reduce entropy within the system, and that the idea that reduction of entropy is the defining feature of living systems ultimately “might help answer the question why it is that we think: to reduce the entropy of our mental representations” (p. 45). While the possibility to draw conceptual parallels between physical, biological, and psychological levels of scientific analysis is appealing, it is important to note a key difference between different applications of the concept. In physical and biological systems, entropy is a descriptive term and requires no further explanation beyond the characterization of its action. In contrast, Dalege and colleagues characterise entropy reduction as an explanatory concept that requires further explanation of its action. Specifically, they argue not only that attitudes are determined through the process of entropy reduction, but also that the mental system is motivated to reduce entropy because that entropy causes distress. This introduction of an additional variable, distress, and the differential use of the concept in descriptive versus explanatory manners, differentiates use of the concept between levels of analysis. Without an explanation of the motivational role of entropy, it is possible that the current framework pushes the question of attitudes back from explaining attitudes to explaining entropy and distress. On the one hand, this may provide testable targets for future studies, but on the other hand the introduction of this additional variable may not be necessary in the first place. We suggest that a more parsimonious strategy would be to adhere more closely to the analogical strategy Dalege and colleagues propose in their introduction and treat entropy as an entirely descriptive concept that needs no further appeal to distress or other variables that justify its action.

To move forward, the current framework might benefit from the integration of basic principles from other (e.g., inferential reasoning) models. First, the framework might integrate ideas about how inferences could arise and about their characteristics (e.g., context-dependence of inferences: Van Dessel et al., 2018) to allow for a more encompassing computation of attitude consistency. This might facilitate explanation of several findings that were not discussed by the authors yet do not readily fit within the attitudinal entropy framework (e.g., that implicit evaluations are sometimes found to predict certain behavior better than explicit evaluations: Banaji & Greenwald, 2013). Second, the motivational role of attitudinal entropy might be elucidated on the basis of current theorizing. In our inferential reasoning model (Van Dessel et al., 2018), we only briefly refer to entropy as a motivational factor in the context of belief updating. We consider entropy not as a characteristic of an attitude (what would be the delineating factor of a configuration of attitude elements?) but of a more general belief system. This idea draws on predictive processing theories in which entropy reduction motivates inferences and behavior to allow for the conservation of mental energy (Friston, 2010). In the spirit of the attitudinal entropy model, we could model entropy in our inferential theory (e.g., as the extent to which inferences require fewer computation steps) and elucidate that entropy might be a factor that determines the circumstances under which a person’s belief system might be updated. Inferred value of information (e.g., for our survival or our self-concept) might also play an important role in inferential reasoning and belief updating such that entropy is not the only principle determining inferences and belief updating (which might be problematic: Otworowska et al., 2018).

**Predictions tested**

While this commentary has primarilally focused on conceptual matters, we also had the opportunity to test two of the framework’s predictions that Dalege and colleagues argue flow from their model with data we already had at hand. We used data from the Attitudes 2.0 dataset (Hussey et al., 2018) to assess predictions number 1b and 3. Data to test other predictions was not at hand. This large dataset (number of experimental sessions > 409,000) represents a single large study of implicit and explicit attitudes that were conducted on the Project Implicit website (https://implicit.harvard.edu). Subsets of this dataset have been used in previous research (e.g., Nosek & Hansen, 2008), and the full dataset is being curated for public release and publication (Hussey et al., 2018). Participants in the study completed one of 190 different IATs assessing attitudes within a wide range of attitude domains including politics, ideologies, popular culture figures, and everyday preferences (total *N* available for analysis = 155913). Self-report attitude scales also assessed multiple attitude features, such as “gut feelings” versus “actual feelings” towards the pairs of concepts used in the IAT. Relevant subsets of this data were employed to test two of the hypotheses that Dalege and colleagues put forward. Data and code for the analyses conducted below are available on the OSF (osf.io/XXXX).

**Prediction 1b: “Scores on implicit measures assessing attitudes, individuals regularly think about, are expected to have higher internal consistency and stability than scores on implicit measures assessing attitudes, individuals think only infrequently about” (p.20).** We calculated internal consistency values for each type of IAT (both Cronbach’s α and McDonald’s ω*t*). Participants were also asked how frequently they thought about the two concept categories that were used in the IAT they completed (e.g., Democrats and Republicans). For each type of IAT (*k* IATs = 190, mean *n* per IAT = 1641), mean frequency ratings were also calculated, resulting in 190 pairs of internal consistency values and mean frequency ratings. When these pairs were entered into linear regression analyses, this demonstrated that the self-reported frequency with which participants thought about the concepts employed in the IATs was predictive of the internal consistency of the IAT’s internal consistency between domains, as predicted by Dalege and colleagues. This relationship held across both metrics of internal consistency (α: β = 0.23, *p* = .024; ω*t*: β = 0.24, *p* = .021).

**Prediction 3: “Attitudes are expected to be less polarized than when individuals are given more time to answer the questions.”** The Attitudes 2.0 dataset also contains self-report ratings of both “gut feelings” and “actual feelings” upon reflection of the 190 concept category pairs. We employed these items to assess the hypothesis that deliberative evaluations are more extreme than gut evaluations. Self-report ratings for each evaluation type were recoded as absolute scores, so that positive scores represent deviation from neutrality/ambivalence without regard to whether those evaluations were positive or negative. A hierarchical linear regression model that accounted for the nesting of evaluations within of concept category domains demonstrated evidence against this prediction: “Gut” evaluations were found to be more extreme on average than “actual” evaluations (β = -0.07, *p* < .001).

As such, analyses using a very large existing dataset provide mixed evidence for the predictions that Dalege and colleagues put forth for the framework. Supportive evidence was found for one prediction, however an effect in the opposite direction to that predicted was found for another. Additional tests of the authors’ other predictions are of course warranted.

**Concluding Remarks**

The Attitudinal Entropy framework interfaces concepts from statistical mechanics (entropy) and social psychology (attitudes) to offer an intriguing new perspective on the latter that has both heuristic and predictive value. Nonetheless, several issues still need to be addressed. The authors will need to situate their framework in the larger eco-system of social psychological theories and specify the unique and truly novel predictions that it makes relative to existing domain-specific accounts. At the same time they will need to add conceptual precision at the level of attitudes, automaticity, and entropy if this framework is to stand the test of time.

References

Banaji, M., & Greenwald, A. (2013). Blindspot: Hidden biases of good people. New York: Delcorte Press.

Barnes, S. M., Bahraini, N. H., Forster, J. E., Stearns-Yoder, K. A., Hostetter, T. A., Smith, G., ... (2016). Moving beyond self-report: Implicit associations about death/Life prospectively predict suicidal behavior among veterans. *Suicide and Life-Threatening Behavior*. https://doi.org/10.1111/sltb.12265

Cone, J., & Ferguson, M. J. (2015). He Did What ? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*, 37–57.

Cone, J., Mann, T. C., & Ferguson, M. J. (2018). Can we change our implicit minds? New evidence for how, when, and why implicit impressions can be rapidly revised. *Advances in Social Psychology*.

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

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences, 11*, 97-104.

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., 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

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

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

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience,* 11:127–38.

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., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorization influence spontaneous evaluations of multiply categorizable objects? *Cognition and Emotion, 24*, 1008-1025.

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*, 1464–80.

Hahn, A., Judd, C. M., Hirsh, H. K., Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General, 143*, 1369-92.

Hussey, I., Hughes, S., Lai, C., Ebersole, C., Axt, J, & Nosek, B. A. (2018). Attitudes 2.0: A large dataset for investigating relations among implicit and explicit attitudes and identity. *Unpublished manuscript.*

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review, 93,* 1449-147.

Kurdi, B., Seitchik, A., Axt, J., Carroll, T. J., Karapetyan, A., Kaushik, N., … (in press). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. The American Psychologist.

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.

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

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *The Behavioral and Brain Sciences*, *32*, 183–198. doi: 10.1017/s0140525x09000855

Monroe, B. M., & Read, S. J. (2008). A general connectionist model of attitude structure and change: the ACS (Attitudes as Constraint Satisfaction) model. *Psychological Review, 115*, 733-759.

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

Nock, M. K., & Banaji, M. R. (2007). Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. *Journal of Consulting and Clinical Psychology, 75*(5), 707–715. doi: 10.1037/0022-006X.75.5.707

Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. *Psychological Science, 21*(4), 511–517. doi: 10.1177/0956797610364762

Nosek, B. A., & Hansen, J. J. (2008). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion*, *22*(4), 553–594. doi: 10.1080/02699930701438186

Otworowska, M., Van Rooij, I., & Kwisthout, J. (2018). Maximizing entropy of the Predictive Processing framework. doi: 10.31234/osf.io/5zam7

Randall, J. R., Rowe, B. H., Dong, K. A., Nock, M. K., & Colman, I. (2013). Assessment of self-harm risk using implicit thoughts. *Psychological Assessment, 25*(3), 714–721. doi: 10.1037/a0032391

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition*, *25,* 638-656.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108–131.

Spruyt, A., & Tibboel, H. (2015). On the automaticity of the evaluative priming effect in the valent/non-valent categorization task. *PLoS ONE 10*: e0121564.

Spruyt, A., Tibboel, H., De Schryver, M., & De Houwer, J. (2017). Automatic stimulus evaluation depends on goal relevance. *Emotion, 18,* 332-341

Van Dessel, P., Hughes, S., & De Houwer, J. (2018). How do actions influence attitudes?   
An inferential account of the impact of action performance on stimulus evaluation. Manuscript invited for revision at *Personality and Social Psychology Review*. Preprint available at: https://osf.io/kb3wq/

1. Although our inferential model mainly focuses on evaluative stimulus-action effects it can easily be (and already has been) generalized to explain other pathways via which evaluative behavior is established or changed (for one such example in the context of evaluative conditioning see De Houwer, 2018). [↑](#footnote-ref-2)