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 and IH are supported by Postdoctoral fellowships of the Scientific Research Foundation, Flanders (FWO-Vlaanderen). JDH is supported by Methusalem Grant BOF16/MET\_V/002 of Ghent University. Correspondence concerning this article should be sent to 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).

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., 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., which of the predictions it makes are truly novel or how it extends beyond previous domain-specific models and accounts). 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 & Reed, 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, Hughes, & De Houwer, 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. 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 latter (e.g., the inferential account suggests that model updating may be based on weighing of the extent to which integration of new information increases entropy [which is unfeasible] compared to other factors such as the inferred value of the information).[[1]](#footnote-2) Second, the Attitudinal Entropy framework takes the position that implicit and explicit attitudes are based on a single type of mental process which involves inferential reasoning. For instance, Dalege and colleagues note that “weights between attitude elements generally arise based on inferences” (p.12) while assessing for entropy (which they conceptualize in part as consistency between attitude elements) presumably requires the system to be able to evaluate how different pieces of information in the network are related to one another (and if such relations are true or not). Such a perspective is compatible with single process (propositional) models of attitudes and learning (De Houwer, 2014; 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 (Smith & DeCoster, 2000), or associative and propositional processes (e.g., Gawronki & 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, XXX.

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

The Attitudinal Entropy framework would also benefit by incorporating recent developments at the conceptual level that have taken place within the attitudes literature. For instance, although they provide detailed information about the factors that may influence attitudes (entropy), create a statistical model to simulate attitudes (using weights, thresholds, nodes, and connections), and refer to different states of attitudes (micro vs. macro state), they never actually define what constitutes an attitude (at least in mental terms). Likewise, their statistical approach treats beliefs, feelings, and behaviors as different elements that make up an attitude. However, scientific progress and theorizing dictates that we (a) first define a conceptual unit before it is operated on and (b) ensure that the concept used to explain (attitudes) is distinct from the phenomenon that needs to be explained (behavior). By not defining the concept that is to be operated on (attitudes) and perhaps more importantly, conflating it with the to-be-explained phenomenon (behavior), the framework may hamper rather than facilitate research on attitudes (for reasons why see De Houwer, 2011).

De Houwer, Gawronski, & Barnes-Holmes (2013) recently offered a meta-theoretical framework that can help resolve these issues. According to their framework, 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? (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 (i.e., as they are when they are viewed as attitudinal elements).

One concrete example of this conflation of behaviour and mediating mental mechanism can be found in simulation 1a, where the authors attempt to demonstrate that the AE model can account for the effects observed on the implicit association test, which they argue are unstable and of limited predictive utility at the individual level but whose 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 indeed, within some domains at least, the IAT has repeatedly demonstrated predictive utility at the individual level (e.g., within suicidality; Nock et al., 2010; Randall et al., 2013; Cha et al., 2018; additional REFS). This aside, a key point here is that, here, Dalege and colleagues attempt to model the behavioural effects observed on implicit measures (explanandum), where their model is elsewhere 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, without explanation, from attitude elements (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 whether what level of analysis is being modeled (behaviour vs. mental level, intrapersonal vs. interpersonal) raises more questions than it answers.

Our inferential model (Van Dessel et al., 2018) adopts such a perspective. Specifically, we model evaluations (rather than attitudes) which we define as 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 (REF). XXX

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

The Attitudinal Entropy framework would benefit from a more nuance perspective on automaticity. In their paper Dalege and colleagues seem to adopt a traditional two-system perspective (System 1 vs. 2) that carves automaticity into two qualitatively distinct classes: implicit vs. explicit. However, this early view on automaticity as a binary, simple, “all-or-nothing” concept has given way to a more nuanced (decompositional) one. Automaticity is now 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, 2014; 2016). This means mental processes are not ‘implicit’ or ‘explicit’: the same mental process can - in principle - operate under certain automaticity in one situation (e.g., occur quickly and without intent) and different conditions in other situations (e.g., occur efficiently or without awareness). 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’.

Yet in their target article the authors seem to equate implicit with just one automaticity condition (attention) and 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 measures by explicit measures because the former by their very nature direct less attention towards the attitude object than the latter” (p.20). We believe that such a perspective is problematic on many fronts. First, it reduces automaticity to a single dimension (attention) and presupposes that measures such as the Implicit Association Test (IAT), Affective Misattribution Procedure (AMP), and others 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 (REF). Others have found that attention to different aspects of the attitude object (e.g., specific attitude elements) influences the type of responses emitted towards that object (see Gawronski, Cunningham, LeBel, & Deutsch, 2010) and that attention to certain properties of the attitude object is actually necessary to observe the attitude (REF). Still more work from our lab shows that ‘top-down’ processes such as goals (Spruyt, Tibboel, De Schryver, & De Houwer, 2017), motivation (REF), and XX, also play more of a 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 a guarantee that it is the most important feature of automaticity when it comes to implicit attitudes. It’s possible that different automaticity features play a role in different types of (implicit) evaluations and that even the same evaluation can operate under certain conditions in certain situations and different conditions in different situations. Thus we would 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 is limited to how much a person thinks about the attitude object).

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

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 and it allows simulation of human evaluative behavior in a nodal network with a good fit to 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, making it difficult to see how consistency between attitude elements could be determined. The assumption that only the (momentary) valence of attitude elements 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 the relational content of information about attitude objects influences evaluation (e.g., Van Dessel, De Houwer, & Smith, 2018) and important moderators of evaluation (e.g., diagnosticity of available information: see Cone & Ferguson, 2015) necessarily involve the validation of relational content of information.

Although Dalege and colleagues seem to assume that inferential reasoning processes determine attitude formation, it is unclear why inferential reasoning processes are not incorporated in their calculation of entropy. At the very least, this calculation should take into account that evaluative responses not only depend on the amount of thinking about an attitude object but also the direction of thinking. The attitudinal entropy framework might benefit from a clearer specification of inferential reasoning processes that elucidates how people make specific inferences (in the context of information validation). Integration of basic principles from inferential reasoning models (e.g., context-dependence of inferences: Van Dessel et al., 2018) might 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).

A more general inferential reasoning model of evaluations might be described that benefits from the elaborate conceptualization (and the specific method of computation) for entropy in the attitudinal entropy model. In this endeavor, we believe that it might be best to consider entropy not as a characteristic of an attitude (what would be the delineating factor of a configuration of attitude elements anyway?) but of a more general belief system. Modeling entropy might then allow estimation of the circumstances under which a person’s belief system might be updated. This could help elucidate how entropy might play a role in evaluative behavior and, by extension, behavior in general.

**Predictions tested**

In their article, Dalege and colleagues present multiple predictions that they argue flow from their model, and which may be tested in future work. As it happens, several of these predictions are testable using existing data. We used data from the Attitudes 2.0 dataset to assess predictions number 1b and 3. This large study (N experimental sessions > 409,000) of implicit and explicit attitudes was conducted on the Project Implicit website. 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., in prep). Participants in the study completed one of 190 IATs assessing attitudes within a wide range of attitude domains including politics, ideologies, popular culture figures, and everyday preferences (total *N* for analysis = 155913, mean N per IAT type = 1641). Self-report attitude scales also assessed multiple attitude features. Relevant subsets of these items and data were employed to test two of the hypotheses Dalege and colleagues put forward.

**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.” Internal consistency values were calculated 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 contained that were used within the IAT (e.g., Democrats and Republicans). For each type of IAT, mean frequency ratings were also calculated, resulting in 190 pairs of internal consistency values and mean frequency ratings. Linear regressions 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, for 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.” Although not a perfect test of this hypothesis, we observed that the Attitudes 2.0 dataset contains self-report ratings of both “gut feelings” and “actual feelings [upon reflection]” of the 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 model that accounted for 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).

**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.

*Ian: Additional points that are currently either brief or absent from the manuscript*

*- They are explicitly silent on where attitude elements come from, thereby neglecting the role of the environment in establishing these attitude elements. We touch on this, but do not elaborate its implications beyond the need for research at the functional level. With regard to the model specifically, one of the authors core arguments regarding Gibbs entropy and attitude strength is that “From the Proposition I.1, it follows that the natural state of an attitude is neutral or ambivalent and that consistent attitudes should be rare”, but that this is not the case because “attitude elements are not independent of each…and because of this dependency attitudes can assume low entropy macrostates.” This non independence is asserted without strong rationale, but a parsimonious explanation for why attitudes are often consistent is that attitude elements are not merely non-independent with this model, but are each formed on the basis of environmental regularities. Where the environment is regular, multiple related (non independent) attitude elements may be formed: for example, “snakes are scary”, “snakes bite”, and “I run away from snakes” are non independent not just in the modeling sense, but also in the sense that these propositions were installed by environmental interactions that possessed these properties.*

*- We touch on the issue of the article focusing too heavily on the phenomena which it can speak to and neglecting consideration of that which it cannot, e.g., the multiple other features of automaticity. This theory-related point could be expanded upon at the model-specific level by asking how sensitive are the simulations are to variation among their parameters? E.g., what range of alternative similar models were run for each simulation, and how closely to results match? The results as presented provide no information about whether these simulations represent extremely specific and fragile cases of the underlying model, or whether they are general dynamics of the model under a range of suitable parameters. James often leveled this accusation at computational models and it could be reiterated here.*

References

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

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.*

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.

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.

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

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

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](http://dx.doi.org/10.1016/j.actpsy.2017.03.016" \t "_blank).

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)