The Attitudinal Entropy (AE) Framework: Clarifications, Extensions, and Future Directions

Dalege, J.; Borsboom, D.; van Harreveld, F.; Lunansky, G.; van der Maas, H.L.J.

Citation for published version (APA):

Routledge
Taylor & Francis Group

REPLY

⭓ OPEN ACCESS ⓜ Check for updates

The Attitudinal Entropy (AE) Framework: Clarifications, Extensions, and Future Directions

Jonas Dalege, Denny Borsboom, Frenk van Harreveld, Gabriela Lunansky, and Han L. J. van der Maas

Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

In our target article, we formulated a general theory of attitudes – the Attitudinal Entropy (AE) framework (Dalege, Borsboom, van Harreveld, & van der Maas, this issue). The AE framework rests on analogical modeling of attitudes using concepts from statistical mechanics and is built on three basic principles. First, attitude inconsistency and instability are two distinct but related indicators of attitudinal entropy – a measure of randomness and unpredictability derived from thermodynamics. Second, energy (in a metaphorical sense) of attitudinal configurations is used to locally evaluate the global entropy of an attitude. Third, attention and thought have an analogous effect on attitudinal entropy as (inverse) temperature has on thermodynamic systems – attention and thought reduce attitudinal entropy. In our target article, we show how this framework can be used to explain a host of established phenomena and to make novel predictions to test the theory empirically.

The commentaries on our target article generally agreed that the AE framework has potential to achieve its main task – capturing the immense complexity of attitude by a limited number of basic principles. This encourages us to further refine the AE framework, so that it will eventually become what we envision it can be: A formalized theory on attitudes that truly explains many established phenomena and that makes unambiguous predictions.

In this reply we take a first step in the refinement of the AE framework by addressing the commentaries' critiques. First, we discuss the value of formulating a formalized and parsimonious theory of attitudes and by this make the aim of the AE framework more explicit. Second, we take the opportunity to clarify the principles of the AE framework. Third, we show that the AE framework explains several findings from the dual-process literature in a more parsimonious way. Fourth, we discuss critiques regarding the question whether the AE framework makes truly novel predictions. Fifth, we discuss the implications for measurement of the AE framework, which hold that measurement fundamentally influences attitudinal processes. Sixth, we discuss some potential extensions of the AE framework to the interpersonal realm. Seventh, to make the formalisms of the AE framework more

accessible and applicable, we present an online app in which the basics of the AE framework are implemented.

The aim of the AE framework

When developing the AE framework our aim was as simple as it was ambitious: the formulation of a theory of attitudes relying on few basic principles yet able to explain several of the complex findings in the vast attitude literature. In our view only such a theory is a practical one: The more parsimonious and constrained a theory is, the more concrete predictions it makes, the more understanding of the modeled system it provides and the more promise for eventual control over the modeled system it holds. While we do not think that the AE framework is the only possibility to arrive at such a theory of attitudes, we do hold that psychology would do well to invest more energy in the development of formalized theories that can be used to explain, model, simulate, and analyze empirical phenomena.

One crucial test for any general theory is whether it can explain established phenomena. A central aim of our target article was therefore to provide a first intensive investigation of whether the AE framework is up to this task. Some commentators were not entirely convinced by the value of this aim. For example, while March, Olson, and Gaertner (this issue) are supportive of our aim to derive a formalized theory on attitudes, they state that "[i]t is far more valuable … when such models make novel predictions [than to explain established findings]" (p. 199). However, in our view investigating whether a theory is able to explain established phenomena is important, because, any theory that *only* makes novel predictions without accounting for established findings is in danger to contribute to a highly fragmented research field, because it does not specify how these novel predictions relate to established phenomena (e.g., Kruglanski, 2001; Vallacher & Nowak, 1997).

Clarifying the aim of the AE framework seemed also important to us, because unfortunately Daley (this issue) misunderstood the aim of the AE framework. While he sees value in linking attitude research to the literature on

entropy, he argues that implementing the AE framework in the Ising (1925) model is not a compelling argument for the correctness of the AE framework. Of course, it is true that implementing the AE framework in the Ising model by itself does not support the correctness of the AE framework, as Daley correctly suggests. However, we never made such a claim. Instead, we implemented the AE framework in the Ising model, because this model lends itself very well to the principles of the AE framework and therefore represents a straightforward opportunity to formalize these principles. Thus, the Ising model implements basic principles, as articulated in the AE framework, which can be motivated on independent substantive terms.

The fact that the Ising model can be used to implement these principles in such a straightforward fashion is, in our view, surprising and noteworthy. However, of course we do not interpret the fact that the AE framework can be implemented in this way as evidence for the model's empirical adequacy, as Daley appears to suggest. Support for the correctness of the AE framework instead comes from the fact that it can successfully reproduce established phenomena. Another point of confusion apparent in Daley's commentary concerns the universality of the Ising model. While the universality of the Ising model poses a problem for underspecified theories, the AE framework considerably constrains the behavior of the Ising model by providing a clear mapping between the mathematical parameters in the Ising model and the substantive interpretations of these parameters. For example, the main principle of the AE framework that attention and thought reduce attitudinal entropy considerably reduces the flexibility of the Ising model, because low attention *must* result in more random and unstable behavior by keeping the dependence parameter (i.e., the parameter used to model attention and thought) in the Ising model at a low value. Thus, while the unconstrained Ising model can definitely fit not just one elephant, but an entire herd of them, as Daley correctly suggest, this does *not* mean that the constrained Ising model we use can fit any conceivable data. On the contrary, the model has clear truth conditions and falsifying instances; for example, if the mere thought effect turned out not to exist (under the conditions specified by the AE framework), the AE framework would be in serious trouble.

An interesting issue regarding the level of explanation that the AE framework provides was raised by Van Dessel, De Houwer, Hughes, and Hussey (this issue). They argue that the AE framework has much value at the descriptive level and at the functional level of explanation, but that its cognitive explanations are limited. Indeed, the focus of the AE framework lies at the level of functional explanations and we agree with van Dessel et al. that future study of the AE framework can focus on providing less abstract cognitive explanations of its central principles (e.g., what is the cognitive process by which attention and thought reduce attitudinal entropy?). In our view, however, van Dessel et al. somewhat undervalue the cognitive explanations the AE framework already provides. They, for example, argue that the AE framework does not provide an explanation for the motivation to reduce entropy other than that it causes distress. While it

seems that we did not communicate this issue optimally in our target article, there are ample reasons why individuals are motivated to reduce attitudinal entropy. For example, the functions typically associated with attitudes cannot be fulfilled by high-entropy attitudes. Based on this argument, it is our view that the AE framework fairs fairly well in explaining the motivation to reduce attitudinal entropy. A related question, however, might be addressed by a more detailed cognitive explanation: What determines that individuals direct attention and thought at some attitudes (and thereby lower the entropy of these attitudes) but not others?

Principles and definitions of the AE framework

The central principle of the AE framework holds that inconsistency and instability reflect attitudinal entropy, which results in attitudes naturally being driven towards inconsistency and instability. While Petty and Briñol (this issue) agree that this principle generates an interesting connection between psychological processes and physical laws, they worry that this principle leads us to focus too much on instable and inconsistent attitudes. We are glad that Petty and Briñol pointed out this issue, because we did not want to suggest that the AE framework focuses on instable and inconsistent attitudes, but that linking inconsistency and instability of attitudes to entropy has implications not only for inconsistent and unstable attitudes but also for consistent and stable attitudes – and all other attitudes, for that matter. The crucial point of the AE framework's central principle is that humans are in a constant struggle of keeping their attitudes from drifting to their natural state of high entropy. The AE framework therefore can account for unstable and inconsistent attitudes (high entropy attitudes to which the individual does not direct much attention and thought), very stable and consistent attitudes (low entropy attitudes that are constantly on the individual's mind) and attitudes anywhere in between (moderate entropy attitudes to which the individual directs some attention and thought). In fact, according to the AE framework, all of these variants of attitudes *must* exist, because a given individual cannot direct attention and thought to all attitude objects at the same time.

Petty and Briñol (this issue) further question whether defining stability and inconsistency of attitudes as related indicators of attitudinal entropy leads to the theoretical lumping of stability and consistency of attitudes. We are glad for being able to clarify this point, because we agree that lumping stability and consistency would not be a good idea. While the AE framework does indeed hold that inconsistency and instability are both indicators of entropy, this does not imply that inconsistency and instability are the same thing; it merely means that they are related constructs. As we stated in our target article, low Gibbs entropy (which relates to stability of attitudes) *creates* the possibility for low Boltzmann entropy (which relates to consistency of attitudes). Thus, from the perspective of the AE framework, stability can exist independently from consistency (attitudes can be stable and ambivalent), but consistency cannot not exist without stability.

Another point that requires some clarification is what micro- and macrostates specifically represent in the AE framework. March et al. (this issue) argue that defining microstates of attitudes as the elements making up the attitude would better align with classic expectancy-value models (e.g., Fishbein & Ajzen, 1975) than our definition. We apparently did not communicate this issue optimally, as our definition aimed to do exactly that: Definition I of the AE framework holds that "The configuration of the attitude elements constitutes the microstate of the attitude" (p. 176) where, as we also explicitly stated, the configuration of an attitude depends on the specific states of the given elements. We intentionally used this definition *because* it aligns well with classic theories on attitude structure, like expectancy-value models. Our definition of attitudinal microstates is thus exactly in accordance with what March et al. advocate.

Similarly, March et al. provide some arguments that seemingly go against the second definition of the AE framework, which holds that the macrostate of an attitude is represented by the number of positive vs. negative attitude elements. They argue that this definition of macrostates does not take relationships between elements into account. We are afraid that, again, we may not have communicated the theory optimally, because in a nontrivial sense, the relations between attitude elements in a network structure are the explanatory workhorse of the theory (see also Dalege et al., 2016) – in fact, in the AE framework, increasing the dependency between attitude elements is the *main* process that reduces attitudinal entropy. March et al. also claim that the AE framework does not capture the potential tension between negative and positive elements at the macrolevel, which is contrary to the formulation of the theory: the tension between negative and positive elements is actually represented by the entropy of the attitudinal macrostate – the higher the number of conflicting attitude elements, the higher the Boltzmann entropy.

Another interesting question raised by some commentators was whether the AE framework could integrate different bases of attitudes (March et al., this issue) and differing importance of attitude elements (Monroe, this issue). While the simplified Ising models we use in our target article do not take these issues into account – in fact, each node in these models is exchangeable – the Causal Attitude Network (CAN) model (Dalege et al., 2016), on which the AE framework is based, does integrate these issues. First, structural importance of nodes (i.e., centrality) is expected to reveal different bases of attitudes (e.g., affective versus cognitive bases; c.f. Edwards, 1990, Fabrigar & Petty, 1999). For example, in the CAN model a network in which affective nodes have more connections (and are therefore more central) than cognitive nodes represents an affect-based attitude. We refer the interested reader to Dalege et al. (2016) and Dalege, Borsboom, van Harreveld, and van der Maas (2017) for an in-depth discussion of the meaning of centrality in attitude networks. Second, Monroe (this issue) is correct in pointing out that the importance of attitude elements is not directly weighted in the contribution to the attitude's macrostate within the AE framework. While it might eventually

become apparent that the AE framework needs to explicitly take weighting of attitude elements into account, structural importance of attitude elements already indirectly leads to differences in the contribution to the attitude's macrostate. As a simple example, take a network with four nodes – nodes 2 and 3 are both strongly connected to node 1 and node 4 is not connected to any node. Node 1 is thus more structurally important than node 4 and the macrostate of the attitude is also more dependent on node 1 than node 4. The reason for this is that the states of nodes 2 and 3 are largely dependent on the state of node 1, so if node 1 is negative (positive) it is likely that nodes 2 and 3 are also negative (positive). The state of node 4, on the other hand, has no impact on the states of other nodes and is therefore less predictive of the macrostate of the attitude.

Another point that requires additional clarification is the process through which a given node changes its state due to the influence of other nodes. In his commentary, Monroe (this issue) argues that whether a node changes boils down to "a probabilistic roll of the dice" (p. 201). This statement, however, does not fully encompass our proposed process through which nodes change (or remain in) their state. The updating of states in the AE framework is based on Glauber (1963) dynamics and the basic workings of this process are the following. During each iteration a node is randomly chosen. Then the current energy of this node is calculated. This energy depends on the extent to which the node's state aligns with the connection the node has to other nodes and with its threshold (the disposition of the node). Then the energy of the node is calculated for its opposite state. If the node's opposite state is lower in energy than its current state, the node is likely to change its state and the likelihood depends on the difference in energy and on the value of the dependence parameter. The higher the dependence parameter, the more deterministic the system becomes. Only under a low dependence parameter is the changing of a node's state comparable to a roll of dice. In our view this is one of the strengths of the AE framework: It can account for random and unpredictable behavior *but* also for organized and predictable behavior.

In general, while Monroe (this issue) judges the attempt of the AE framework to explain attitudinal phenomena in a more formalized and rigorous way as a necessary step towards theoretical progress, he also questions how realistic some modeling choices of the AE framework are. He contrasts the AE framework with his connectionist neural network model of attitudes – the Attitudes as Constraint Satisfaction (ACS) model (Monroe & Read, 2008) – and argues that the ACS model is a more realistic model of attitude dynamics than the AE framework. We agree that the ACS model fairs well in providing a realistic model of how dynamics of neural networks relate to attitude dynamics. However, a realistic model of any given process is not necessarily a good theory. For example, neural network models hardly fulfill the central aim of theorizing to make the basic elements of the theory as simple as possible. This does, of course, not mean that they are not useful – to the contrary, building a copy of the brain can deliver many insights, but
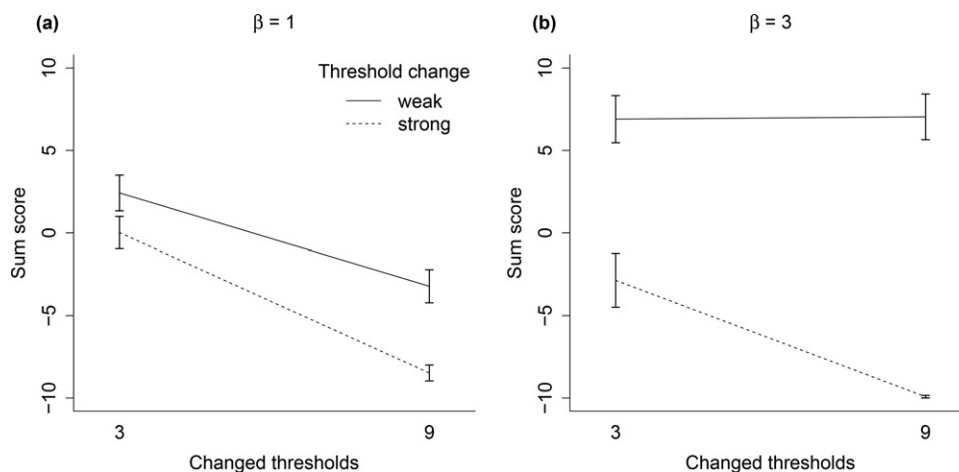
Figure 1. Effects of number and strength of threshold change under low (a) and high (b) dependence parameter. Error bars represent +- 2 standard errors around the mean.

to truly build a theory on these insights one needs to simplify the modeled processes. The AE framework represents such an attempt and is therefore also complementary rather than in competition with neural network models like the ACS model. One can think of the AE framework as the attempt to provide a map of relevant attitudinal processes, while neural network models try to rebuild the whole planet of attitude processes. In our view the additional value of neural network models is that they can be used to investigate where more parsimonious theories like the AE framework break down. A fruitful avenue for integration of the AE framework and neural network models like the ACS model would therefore be a detailed investigation which predictions are provided by the ACS model that do not follow from the AE framework. One area where the ACS model probably has more predictive power than the AE framework is the learning of attitudes. Currently, the AE framework does not specify how attitude networks develop and ongoing investigations of our lab focus on integrating Hebbian learning as used in the ACS model into the AE framework.

Beyond dual-process models

A common theme in several commentaries was the question whether the AE framework is able to integrate insights from dual-process models. While Van Dessel et al. (this issue) were in support of the AE framework as an alternative to dual-process models, Petty and Briñol (this issue) and March et al. (this issue) were not entirely convinced about the explanatory scope of the AE framework regarding findings from the dual-process paradigm. In our view the AE framework does a good job in explaining findings from the dual-process tradition. In the target article we already modeled a central finding from the dual-process approach to persuasion and in this section we first further explore how well the AE framework fares in explaining findings from this literature. We then show that the AE framework is also able to explain state-like and trait-like behavior of attitudes.

We were very glad that two prominent researchers from the dual-process tradition in persuasion research agreed that the AE framework is able to explain the general finding that

low involvement leads to heuristic processing while high involvement leads to systematic processing (Petty & Briñol, this issue). Petty and Briñol, however, also challenge us to explain more nuanced effects from the dual-process tradition and we are glad for the opportunity to do so. Specifically, they argue that the AE framework does not predict the finding that under low involvement number of arguments determines attitude change, while under high involvement quality of arguments determines attitude change (Petty & Cacioppo, 1984). Because the AE framework is a formalized theory we are able to test whether this claim is correct.

Simulation R1: Modeling impact of argument quantity versus argument quality

To test whether the AE framework predicts the effects of quantity versus quality of arguments under low versus high involvement, we set up a simulation similar to Simulation 3 of our target article. As in Simulation 3, we first created a low involvement group represented by a low dependence parameter ($\beta = 1$) and a high involvement group represented by a high dependence parameter ($\beta = 3$) and initialized all individuals' attitudes to have a positive disposition (all thresholds = 0.2). We straightforwardly modeled the effect of number of arguments by changing either the thresholds of three or nine nodes. The quality of arguments was modeled by different impact on the thresholds (weak arguments changed the thresholds to −0.2 and strong arguments changed the thresholds to −0.7). We then simulated 100 individuals in each of the six conditions (each individual was modeled using 1000 iterations of Glauber dynamics; in the first 500 iterations thresholds were set to the positive disposition to initialize the positive pre-persuasion attitude and in the second 500 iterations thresholds were set to the specific values of the given condition).

Figure 1 shows that the AE framework reproduced the global effect identified by Petty and Cacioppo (1984) and the three-way interaction on the sum score of the attitude elements at the 1000th iteration was significant, $F(1, 792) = 7.52$, $p = .006$, $\eta_p^2 = .01$. Figure 1(a) shows the

influence of the different threshold change conditions in the low dependence parameter condition. In this condition a strong main effect is observed for the number of changed thresholds (reproducing the finding that under low involvement argument quantity mostly determines attitude change), $F(1, 396) = 237.27$, $p < .001$, $\eta_p^2 = .37$. While also significant, the main effect of strength of the threshold change was substantially lower (reproducing the finding that under low involvement argument quality has less impact on attitude change), $F(1, 396) = 69.63$, $p < .001$, $\eta_p^2 = .15$. We also observed a small interaction effect between number and strength of threshold change, $F(1, 396) = 9.71$, $p = .002$, $\eta_p^2 = .02$, which was mostly driven by the stronger impact of strength of threshold change when number of arguments was high. Figure 1(b) shows the influence of the different threshold change conditions in the high dependence parameter condition. In this condition a strong main effect is observed for the strength of the threshold change (reproducing the finding that under high involvement argument quality mostly determines attitude change), $F(1, 396) = 429.93$, $p < .001$, $\eta_p^2 = .52$. While also significant, the main effect of the number of changed thresholds was substantially lower (reproducing the finding that under high involvement argument quantity has less impact on attitude change), $F(1, 396) = 28.84$, $p < .001$, $\eta_p^2 = .07$. We also observed a small interaction effect between number and strength of threshold change, $F(1, 396) = 31.22$, $p < .001$, $\eta_p^2 = .07$, which was mostly driven by the stronger impact of strength of threshold change when number of arguments was high. The different effects of quantity versus quality of arguments under low versus high involvement thus in fact follow from the AE framework.

A noteworthy observation in addition to the global pattern of the results in Figure 1 is that the AE framework predicts that when weak arguments are administered to highly involved individuals, the number of arguments does not matter. If the arguments were too weak, persuasion would just not be effective no matter the number of arguments. In contrast, Petty and Briñol (this issue) argue that under high involvement a large number of weak arguments leads to even *less* persuasion than a low number of weak arguments. This is thus indeed not what the AE framework would predict and a convincing empirical demonstration of this effect would indicate a limitation of the AE framework. A careful examination of the study by Petty and Cacioppo (1984) on which Petty and Briñol build their argument, however, reveals that the effect is far from robust – in fact, the effect did not even reach the conventional threshold for statistical significance so additional research is necessary to evaluate whether the effect in fact exists at all. Absent such evidence, in our view Simulation R1 lends further support to the claim that the AE framework can explain findings from the persuasion literature in a more parsimonious and constrained way than dual-process models like the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986) and the Heuristic-Systematic Model (HSM; Chaiken, Liberman, & Eagly, 1989).

March et al. (this issue) use a dual-process approach to what they see as state-like and trait-like attitudes. In their view, weaker attitudes are more state-like, because they are fluctuating and highly context-dependent, while stronger attitudes are more trait-like, because they are stable and relatively immune to context effects. It is our view that such a description of attitudes has a lot of merit and that such characterizations represent the endpoints on a continuum. In fact, the AE framework suggests that attitudes range from being high in entropy (and therefore highly fluctuating and unpredictable) to attitudes being low in entropy (and therefore stable and predictable). Therefore, while March et al. are correct in stating that the AE framework treats attitudes as states, the AE framework is able to integrate theories assuming attitudes to be temporary constructions (e.g., Schwarz, 2007) with theories assuming attitudes to be stable representations in memory (e.g., Fazio, 2007).

From the perspective of the AE framework, whether attitudes are more state-like or more trait-like is determined by the dependence parameter. While the AE framework does away with the idea of a stable representation of a summary evaluation in memory, attitudes in low entropy states (i.e., attitudes to which individuals direct much attention and thought) behave exactly as what would be expected from a stored representation of the summary evaluation. Under a high dependence parameter, attitude elements keep each other in check and therefore the macrostate of the attitude remains in the same state. Thus, while in principle the macrostate of the attitude is consistently constructed, the construction process under a high dependence parameter always results in a similar global evaluation and would therefore *appear* as a stored representation. In our target article, we show that variations in the dependence parameter also explain variations in attitude strength, such as resistance to change and impact on behavior (see also Dalege et al., 2016; Dalege, Borsboom, van Harreveld, & van der Maas, 2018). Given that the AE framework provides a parsimonious explanation of both weak and strong attitudes, a dual-process explanation of strong and weak attitudes may be superfluous.

In conclusion, while dual-process models have considerably advanced the attitude literature by illuminating, for example, the different effects of persuasion under low versus high involvement, the value of the dual-process description may ultimately be heuristic rather than explanatory. While indeed persuasion has different effects under low and high involvement, and while weak attitudes are quite different from strong attitudes, these differences may not arise because of different processes but rather reflect one and the same process operating under different conditions.

Does the AE framework provide novel predictions?

The commentators substantially diverged on the question whether the AE framework leads to novel predictions. On the one hand, Van Dessel et al. (this issue) went so far as to actually test two predictions of the AE framework, Rios and Roth (this issue) argue that the AE framework cannot only

illuminate intrapersonal but also interpersonal processes and Petty and Briñol (this issue) see the novel predictions as one of the AE framework's strengths. On the other hand, March et al. (this issue) seem to argue the extremely strong thesis that the AE framework does not make any novel predictions at all, as they contrast the AE framework with "models [that] make novel predictions about heretofore unobserved phenomena" (p. 199). However, it is unclear what could lead March et al. to this conclusion; the only evidence appears to be that, in their view, two out of 17 predictions stated in our target article were either already tested or trivial; even if this is true, there would seem to be 15 predictions remaining. In this section, we present a detailed analysis on the question whether the AE framework makes novel predictions.

A first point regarding the novelty of predictions is that judging whether a prediction is novel seems to be a rather difficult task. As case in point, compare how Petty and Briñol (this issue) judge the novelty of the AE framework's prediction that an opposite mere-thought effect exists (i.e., forcing individuals to quickly answer questions is expected to result in less extreme attitudes than when the questions are administered in an usual way) to March et al.'s (this issue) judgment. While Petty and Briñol called this prediction "intriguing" and judged it to be counterintuitive and unlikely to be confirmed, March et al. called the prediction neither "new [n]or surprising" and are convinced that this prediction could easily be demonstrated.

Why do established experts in the attitude field diverge so strongly on whether or not an effect would likely be demonstrated? We suggest that one important reason that researchers working in the same area can differ so substantially in their judgment of whether a prediction is novel might be that evaluating predictions based on verbal theories result in far too much flexibility. By turning the meaning of words in their intended direction, both March et al. (this issue) and Petty and Briñol (this issue) are able to provide valid reasons for *both* the existence *and* non-existence of the opposite mere-thought effect. This, in our view, illustrates the Achilles' heel of current mainstream theorizing in psychology: its reliance on verbal stories makes theoretical predictions far too dependent on unarticulated interpretations of theoretical constructs. In our view, a crucial strength of the AE framework is that it provides predictions that *must* follow, and this also gives more value to these predictions, because the predictions follow *independently* of what the developers of the model might think. Such unambiguous predictions are in our view valuable, because testing such predictions represents a more diagnostic test than testing a prediction that is only loosely based on a theory and dependent on this or that theorist's interpretation of the model.

Interestingly, Van Dessel et al. (this issue) conducted a first test of the opposite mere thought effect. They compared the extremity of attitude ratings, which asked individuals to report their reflected feelings, to the extremity of attitude ratings, which asked individuals to report their gut feelings. Van Dessel et al. found that these gut feelings were slightly more extreme than reflected feelings (note, however, that this was a very small effect), which is not in line with the predicted opposite mere thought effect. However, the test by Van Dessel et al. relies on the strong (and in our view not always realistic) assumption that individuals have reliable knowledge about their high entropy attitudes. Given the assumption that the state of high attitudinal entropy arises when individuals do not pay attention to the attitude object, it would not seem evident that people have introspective knowledge about such attitude states. Thus, while the test of Van Dessel does provide indirect evidence against the opposite mere thought effect, we think that the jury is still out on whether this effect exists, and further, more direct empirical tests, are warranted.

Another issue that came up in the commentaries regarding the novelty of the AE framework's predictions was that some of them have already been tested and confirmed (March et al., this issue; Petty & Briñol, this issue). We acknowledge that a subset of the AE framework's predictions indeed have already been tested, and we thank the commentators for pointing this out. However, it is important to note that whether a prediction was tested before or after we derived the AE framework has no bearing on the predictive power of the AE framework or the degree to which the reported effects confer evidence on the theory: if it turns out that the Ancient Greeks already observed that heavy and light objects fall at the same speed, that should not diminish the evidential force of this finding vis-à-vis Galileo's theory of free fall. Thus, in our view, while the fact that some predictions of the AE framework have already been tested does show our unawareness of this part of the literature, this does not detract from the fact that such confirmations provide evidence for rather than against the AE framework.

To conclude, while the novelty of some of the AE framework's predictions might be debatable, most of the AE framework's predictions have potential to advance the attitude literature by (1) providing more precision than what could be derived from other theories (e.g., predictions regarding boundary conditions of the mere thought effect and the effectiveness of persuasion) and (2) by predicting entirely novel effects. Many of these effects are in the context of the measurement of attitudes and we discuss the AE framework's implications for measurement in detail in the next section.

Let's go all psychology on attitude measurement

The probably most controversial implication of the AE framework is that measurement has fundamental influence on attitudes – the way attitudes are measured influences their entropy. While this implication met with doubts from some commentators (March et al., this issue; Petty & Briñol, this issue), other commentators appeared more enthusiastic about this implication and provided some first empirical support for the measurement implications of the AE framework (Van Dessel et al., this issue).

In contrast to March et al. (this issue), who stated that their "knee-jerk reaction to [the measurement implications

of the AE framework] … is that unless we've gone 'full-physics,' where attitudes are akin to quantum particles whose very measurement affects them (as in Heisenberg's uncertainty principle), it risks false-reification to claim that measurement determines some attitudinal property" (p. 197), we think that one should keep an open mind about the possibility that measurement affects attitudinal processes unless convincing evidence accumulates for attitudes being immune to measurement effects. In our view, however, the evidence *for* measurement effects on attitudes is substantial. Measurement of attitudes, just as most psychological measurements, sets in motion a variety of processes (e.g., Cannell, Miller, & Oksenberg, 1981; Strack & Martin, 1987; Thurstone, 1927; Tourangeau, Rips, & Rasinski, 2000) and it is unlikely that this has no impact at all on the measured system. Furthermore, the reference to quantum physics by March et al. actually supports our theorizing, because there is strong empirical evidence that at least some cognitive processes follow quantum probabilities (e.g., Pothos & Busemeyer, 2013) and, especially relevant to the current discussion this also holds for measurement effects on judgments (e.g., Busemeyer & Wang, 2018; Wang & Busemeyer, 2013; Wang, Solloway, Shiffrin, & Busemeyer, 2014). Additionally, one of the pioneers of quantum mechanics – Niels Bohr – probably used ideas from William James on thought-processes to derive quantum mechanical principles (Stapp, 1993; Plotnitsky, 2012). The reason why the theory of quantum mechanics might be built on insights from William James on thought-processes is that the nature of the human mind lends itself well to the issues discovered in quantum physics. While in physics it is highly counterintuitive that particles are fuzzy and do not have a definite position in space, that they can interfere with themselves, or that measurement affects the nature of particles, it is immediately apparent by introspection that such concepts are very fitting for mental processes. Thoughts are fuzzy and often hard to pin down, thinking one thought precludes thinking other thoughts, and providing a concrete answer to an attitude questionnaire forces one to reduce the fuzziness of one's thoughts to a single and concrete response. So, in our view appreciating the impact that psychological measurement might have on psychological processes has nothing to do with going full-physics, but simply means that we go full-psychology.

All in all, it seems at the very least worthwhile to explore the consequences of assuming that measurement affects attitudinal processes. From the perspective of the AE framework, the most important factor in measurement is the extent to which the measurement instrument directs attention to the attitude object. Measurement instruments limiting attention to the attitude object affect entropy-reduction less than measurement instruments not limiting attention to the attitude object.[1] All else being equal, implicit measures therefore tap attitudes in states higher in entropy than explicit measures. From this follows that scores on implicit measures must principally be less consistent and less stable than scores on explicit measures. While there is a large number of findings in the attitude literature on implicit measures, the probably most robust results on implicit measures are (1) that they show low internal consistency and low temporal stability (e.g., Bar-Anan & Nosek, 2014; Gawronski, Morrison, Phills, & Galdi, 2017; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005) and (2) that their value in predicting behavior is limited (e.g., Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). Any theory on what is measured by implicit measures should therefore first and foremost address these findings – otherwise it would be severely limited in its empirical grounding. It is our view that most theories on implicit measures fall short of this criterion and because of this our understanding of what is measured by implicit measures has remained elusive. Because the AE framework straightforwardly explains the most robust findings on implicit attitudes, the AE framework shows strong potential to illuminate one of the most researched and yet most elusive constructs of the last two decennia of attitude research – that of implicit attitudes.

Linking the measurement of attitudes to other established phenomena in the attitude literature, such as the mere thought effect, not only provides more understanding of what it means to measure attitudes implicitly, but also provides a host of novel predictions regarding the measurement of attitudes. For this reason, we appreciate it that Van Dessel et al. (this issue) already tested one of these predictions – specifically, the prediction that reliability of implicit measures is expected to be higher for attitude objects individuals think frequently about, than for attitude objects individuals do not think often about. In a straightforward test, Van Dessel et al. (this issue) provided first support for this prediction. The result of this test is, of course, not conclusive support for the measurement implications of the AE framework, but it shows that that the AE framework is able to advance our understanding of implicit measurement. We are therefore looking forward to more empirical investigations of the AE framework's measurement implications – in our view, the question to what extent measurement affects attitudinal processes is a rather open one and better understanding of this issue will lead to theoretical progress in the understanding of attitudes.

Extending the AE framework to the interpersonal realm

While the former sections of this reply mostly focused on clarifying and discussing principles and implications of the AE framework, we also want to use the opportunity of this reply to discuss possible extensions of the AE framework. Specifically, we want to reflect on Rios and Roth's (this issue) thoughtful analysis of how one could apply the AE framework to the interpersonal realm.

In our target article, we already raised the question to what extent the different levels of attitudinal entropy reduction are socially instigated. Based on Rios and Roth's (this

---

[1]There is also evidence that mere measurement of intentions leads to stronger correspondence between intentions and behavior (e.g., Wood et al., 2016). This effect is also in line with the AE framework's implication that measurement of attitudes affects attitudinal entropy.

issue) commentary, the answer to this question seems to be quite a lot. In the AE framework, the influence of the social environment can take two routes. First, the social environment might directly affect the information on which the attitude is based (by affecting the thresholds in the attitude network). Such a situation would arise when one, for example, belongs to a homogeneous social network where most individuals have the same political views (cf., Visser & Mirabile, 2004). Second, the social environment might affect the motivation for attitudinal entropy reduction (by affecting the dependence parameter of the attitude network). Such a situation would arise when, for example, a given topic receives much attention in the media. Other situations pointed out by Rios and Roth that might affect entropy reduction are wanting to confirm to the group norm (cf., Carlson & Settle, 2016) and the motivation to not look like a hypocrite (cf., Barden, Rucker, & Petty, 2005; Barden, Rucker, Petty, & Rios, 2014). In our view, a potentially fruitful avenue for future research is to investigate how dependent the different levels of attitudinal entropy reduction are on the social environment. It seems likely that high entropy reduction is often instigated by the social environment, as, for example, the typical determinants of attitude importance, such as values and social identification (Boninger, Krosnick, & Berent, 1995), have clear social connotations. In our view, an intriguing possibility for the determinants of attitudinal entropy reduction would be that paying attention and directing some thinking to the attitude object have the default effect of moderately reducing attitudinal entropy and that higher entropy reduction is mostly instigated by the social environment – the situations Rios and Roth discuss are obvious examples for this, but also having to make a decision (a likely core determinant of attitudinal entropy reduction) does generally not happen in a social vacuum.

Another question that we raised in our target article was what determines tolerance of attitudinal entropy. While it seems that attitudinal entropy reduction is a natural response in many situations, there are also situations that make attitudinal entropy reduction *less* likely. Rios and Roth (this issue) also provide some ideas regarding these determinants of tolerance toward attitudinal entropy, such as being motivated to adopt different attitudes depending on the context (Sinclair, Lowery, Hardin, & Colangelo, 2005) and to integrate different viewpoints (Antonio, Chang, Hakuta, Kenny, Levin, & Milem, 2004). In our target article, we briefly discussed empirical evidence of cultural differences in attitudinal entropy reduction. An exciting possibility for cultural differences in attitudinal entropy reduction might be that the low levels of attitudinal entropy are more or less culturally invariant, while the higher levels are dependent on the cultural context and that the higher levels of attitudinal entropy reduction take different forms in different cultures. For example, in individualistic cultures higher levels of attitudinal entropy reduction might be more determined by intrapersonal processes, while they might be more determined by interpersonal processes in collectivistic cultures (cf., Hoshino-Browne et al., 2005; Kitayama, Snibbe, Markus, & Suzuki, 2004).

A final point that we want to make regarding the interpersonal consequences of attitudinal entropy reduction is that we wholeheartedly agree with Rios and Roth (this issue) that attitudinal entropy reduction can have negative consequences (both on the individual level and on the group level). As Rios and Roth point out, high attitudinal entropy reduction in groups might be indicative of group think (Janis, 1982) and also on the individual level low entropy attitudes can have negative consequences (such as being too rigid and extreme). Better understanding the determinants of attitudinal entropy reduction might therefore also help in providing interventions to combat these negative consequences.

Attitudinal Entropy framework – the app

While we believe its formalized nature is one of the AE framework's assets, we also acknowledge that most social psychologists are not familiar with working with formalized theories. To make the AE framework more accessible, we therefore implemented the basic workings in an online app using the interactive web application Shiny (Chang, 2018) that runs on the programming platform R (R Core Team, 2013). The app can be assessed at https://jdalege.shinyapps.io/AttitudinalEntropyFramework/ and we will keep updating the app so that most insights from the AE framework are eventually implemented in the app.

Currently, the app has two main tabs. In the first tab, the user can simulate the equilibrium distribution of a given network under two different dependence parameters, which are used to model different amounts of attention and thought directed at the attitude (similar to the simulations on the mere thought effect in the target article). Currently, the user can vary the size of the networks between 4 and 10 nodes, the values of all thresholds between -1 and 1, the connectivity of all edges in the network between 0, 0.05, and 0.1, and the dependence parameter between 0 and 3. Figure 2 shows a screen shot of this tab in which we replicated the first simulation of the target article on the mere thought effect (Simulation 2a).

In the second tab, the user can simulate the dynamic behavior of the AE framework (similar to the simulations on persuasion in the target article and in the current reply). The starting thresholds are currently automatically set to 0.2 and the user can vary the amount (between 0 and 10) and strength (between −1 and 0) of the thresholds that will be changed by the persuasion, the dependence parameter between 0 and 3, and the number of simulated individuals between 10 and 100 (note that a high number of individuals results in a rather long time of computation). Figure 3 shows a screen shot of this tab in which we replicated the low dependence parameter/3 weakly changed thresholds condition of Simulation R1.

An essential next step

To conclude, we want to reiterate our claim that an essential next step for attitude research and Psychology in general is
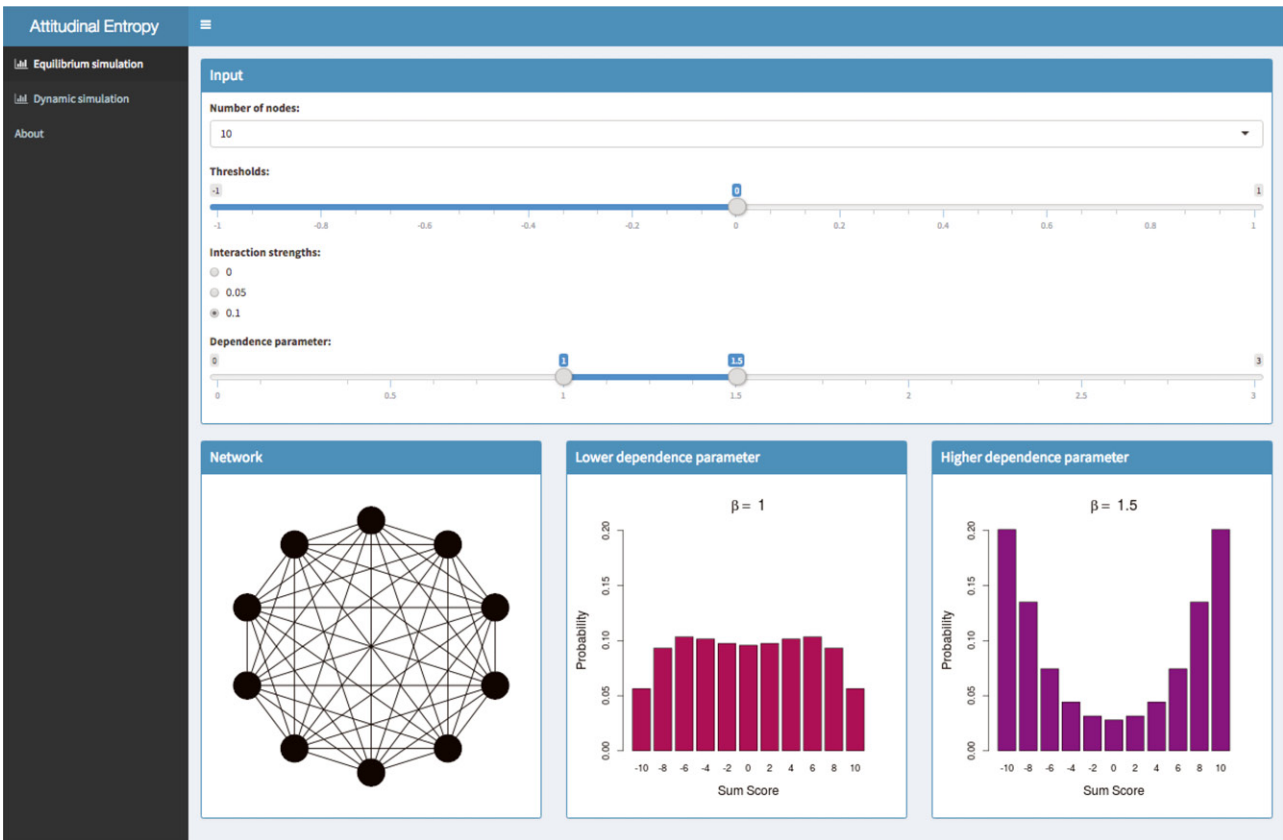
Figure 2. Screenshot of the online application in which Simulation 2a of our target article is replicated.
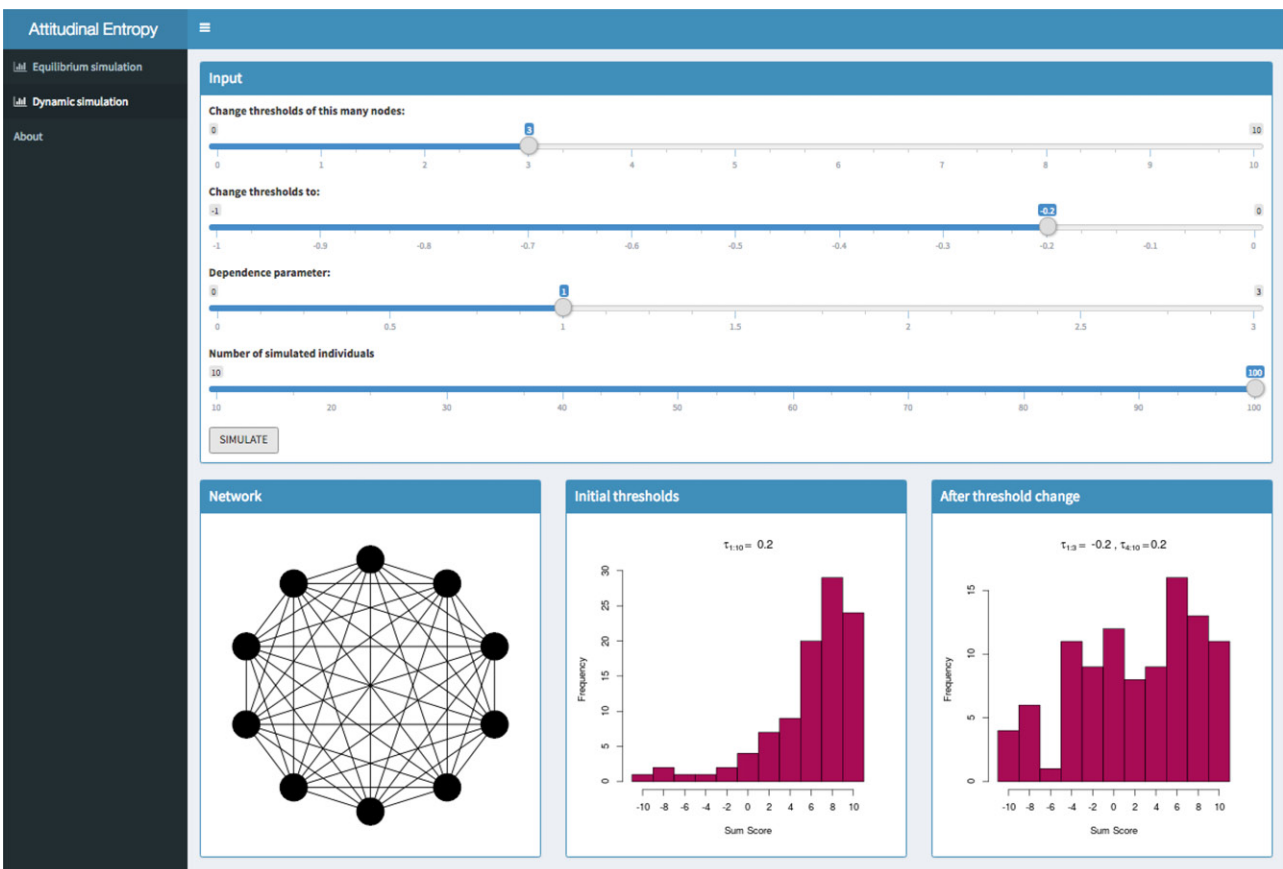

Figure 3. Screenshot of the online application in which the low dependence parameter/3 weakly changed thresholds condition of Simulation R1 is replicated.

to develop theories that (a) rely on few fundamental principles, (b) are formally specified and (c) make unambiguous empirical predictions. We remain, of course, open to the possibility that the specifics of the AE framework might be overthrown by a similarly well-specified theory, but as we show in this reply there are ample reasons to think that the AE framework rests on the correct building blocks to construct a comprehensive theory of attitudes. In our view, the main tasks for future research on the AE framework will revolve around two issues. First, the testing of predictions derived from the AE framework to establish a firmer empirical basis for the AE framework and second, studying the boundary conditions of the AE framework. Based on the parsimonious nature of the AE framework it is certain that at some point the explanatory scope of the AE framework will break down and the explanatory power of the AE framework will be determined by how fast this point is reached. Given the many established findings the AE framework is able to explain, we are optimistic that this power is rather substantial. We therefore stand by our conclusion of the target article that the answer to the question why we think might be "to reduce the entropy of our mental representations" (p. 190). However, we also agree with March et al. (this issue) that thinking is always in the service of doing, but the question remains how the cognitive system accomplishes to think in a functional way. Our tentative answer to this question is, however, straightforward and instructive: thinking is for reducing entropy, and reducing entropy is for doing.

References

Antonio, A. L., Chang, M. J., Hakuta, K., Kenny, D. A., Levin, S., & Milem, J. F. (2004). Effects of racial diversity on complex thinking in college students. *Psychological Science*, 15, 507–510.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, 46, 668–688.

Barden, J., Rucker, D. D., & Petty, R. E. (2005). "Saying one thing and doing another": Examining the impact of event order on hypocrisy judgments of others. *Personality and Social Psychology Bulletin*, 31, 1463–1474.

Barden, J., Rucker, D. D., Petty, R. E., & Rios, K. (2014). Order of actions mitigates hypocrisy judgments for ingroup more than outgroup members. *Group Processes and Intergroup Relations*, 17, 590–601.

Boninger, D. S., Krosnick, J. A., & Berent, M. K. (1995). Origins of attitude importance: self-interest, social identification, and value relevance. *Journal of Personality and Social Psychology*, 68, 61–80.

Busemeyer, J. R., & Wang, Z. (2018). Hilbert space multidimensional theory. *Psychological Review*, 125, 572–591.

Cannell, C., Miller, P., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 389–437). San Francisco, CA: Jossey-Bass.

Carlson, T. N., & Settle, J. E. (2016). Political chameleons: An exploration of conformity in political discussions. *Political Behavior*, 38, 817–859.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing with and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York, NY: Guilford Press.

Chang, W. (2018). shiny: Web Application Framework for R (R package Version (1.1.0) [Computer Software]. Available at https://cran.r-project.org/web/packages/shiny/index.html

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude (CAN) model. *Psychological Review*, 123, 2–22.

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. (2017). Network analysis on attitudes: A brief tutorial. *Social Psychological and Personality Science*, 8, 528–537.

Dalege, J., Borsboom, D., van Harreveld, F., & van der Maas, H. L. (2018). A network perspective on attitude strength: Testing the connectivity hypothesis. *Social Psychological and Personality Science*. Advance online publication. doi:10.1177/1948550618781062

Edwards, K. (1990). The interplay of affect and cognition in attitude formation and change. *Journal of Personality and Social Psychology*, 59, 202–216.

Fabrigar, L. R., & Petty, R. E. (1999). The role of the affective and cognitive bases of attitudes in susceptibility to affectively and cognitively based persuasion. *Personality and Social Psychology Bulletin*, 25, 363–381.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, 25, 603–637.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading: Addison-Wesley.

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, 43, 300–312.

Glauber, R. J. (1963). Time-dependent statistics of the Ising model. *Journal of Mathematical Physics*, 4, 294–307.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31, 1369–1385.

Hoshino-Browne, E., Zanna, A. S., Spencer, S. J., Zanna, M. P., Kitayama, S., & Lackenbauer, S. (2005). On the cultural guises of cognitive dissonance: The case of easterners and westerners. *Journal of Personality and Social Psychology*, 89, 294–310.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus [Contribution to the theory of ferromagnetism.] *Zeitschrift für Physik*, 31, 253–258.

Janis, I. L. (1982). *Psychological studies of policy decisions and fiascos*. Boston, MA: Houghton Mifflin

Kitayama, S., Snibbe, A. C., Markus, H. R., & Suzuki, T. (2004). Is there any "free" choice? Self and dissonance in two cultures. *Psychological Science*, 15, 527–533.

Kruglanski, A. W. (2001). That "vision thing": The state of theory in Social and Personality Psychology at the edge of the new millennium. *Journal of Personality and Social Psychology*, 80, 871–875.

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.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105, 171–192.

Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on response to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, 46, 69–81.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123–205.

Plotnitsky, A. (2012). *Niels Bohr and complementarity: An Introduction*. Berlin: Springer.

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, 36, 255–274.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Sinclair, S., Lowery, B. S., Hardin, C. D., & Colangelo, A. (2005). Social tuning of automatic racial attitudes: The role of affiliative motivation. *Journal of Personality and Social Psychology*, *89*, 583–592.

Stapp, H. P. (1993). *Mind, Matter, and Quantum Mechanics*. Berlin: Springer.

Strack, F., & Martin, L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In H. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 123–148). New York: Springer-Verlag.

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

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, England: Cambridge University Press.

Vallacher, R, R., & Nowak, A. (1997). The emergence of dynamical social psychology. *Psychological Inquiry*, *8*, 73–99.

Visser, P. S., & Mirabile, R. R. (2004). Attitudes in the social context: the impact of social network composition on individual-level attitude strength. *Journal of Personality and Social Psychology*, *87*, 779–795.

Wang, Z., & Busemeyer, J. R. (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science*, *5*, 689–710.

Wang, Z., Solloway, T., Shiffrin, R.M., & Busemeyer, J.R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of sciences of the United States of America*, *111*, 9431–9436.

Wood, C., Conner, M., Miles, E., Sandberg, T., Taylor, N., Godin, G., & Sheeran, P. (2016). The impact of asking intention or self-prediction questions on subsequent behavior: a meta-analysis. *Personality and Social Psychology Review*, *20*, 245–268.