Published as: March, D. S., Olson, M. A., & Gaertner, L. (2018). On Logical and Mathematical Boxes: Does the Attitudinal Entropy Framework Expand our Understanding of Attitudes? *Psychological Inquiry, 29,* 196-199.

FINAL DRAFT: MAY CONTAIN ERRORS

On Logical and Mathematical Boxes:

Does the Attitudinal Entropy Framework Expand our Understanding of Attitudes?

David S. March, Michael A. Olson, & Lowell Gaertner

University of Tennessee

"You can catch phenomena in a logical box or a mathematical box. The logical box is coarse but strong. The mathematical box is fine-grained but flimsy. The mathematical box is a beautiful way of wrapping up a problem, but it will not hold the phenomena unless they have been caught in a logical box to begin with."

- J. R. Platt, 1964 (p. 352)

People commonly understood the physical world through psychological processes in ancient times, sometimes metaphorically but often literally: the sun and climate were manifestations of Ra, the Egyptian sun god, and fire was the embodiment of the god Agni in Hindu tradition. The pendulum has since swung the other way: the natural sciences have gradually de-anthropomorphized the physical world, and with great success. So, it isn't surprising that people now seek to understand the psychological world through physical processes, sometimes metaphorically but often literally.

We do, of course, see the utility in advancing mathematical models of human mental processes. For example, modeling the electrical communications of neurons and the movements of neurotransmitters onto the properties of physical movement has led to advances in understanding how the brain works (e.g., Busemeyer & Wang, 2015). And we understand the temptation to continue the exercise of reducing psychological processes to physical ones (e.g., Pentland, 2015; also see work on biomedicalization). However, we wonder whether the pendulum has swung too far in this direction (for a related argument, see Kvaale, Haslam, & Gottdiener, 2013).

Despite our concerns about its premise, the Attitude Entropy (AE) model has potential to provide an engaging challenge to conventional approaches to attitudes research. In order for a

novel model to challenge orthodoxy, it must engage directly with it: the model's concepts must interact with current concepts in specific and clearly articulated ways. Thus, we first address the issue of how clearly and compellingly the AE model is communicated, and how variables central to the AE model engage with variables central to the attitudes and persuasion literature. In order to offer advances in understanding of any phenomena, a novel model must (ideally) explain existing phenomena and make predictions about as-yet unobserved phenomena. Thus, we address secondly whether the AE model appears to offer such advances. In the domains of attitude measurement, attitude strength, and attitude consistency, we remain open but skeptical about the explanatory insight offered by the AE model.

Clarity & Alignment with Attitude Concepts

Our first concern regards the nature and clarity of the relevant variables in the AE model and how they align with established variables in attitudes research. The AE's conceptual variables do harken current conceptualizations of attitudes, but in hard-to-align ways. For instance, according to their definition, microstates are more than merely the elements of the attitude. Were they merely elements absent any configuration, they might align well with notions of attitude *attributes* from expectancy-value perspectives (e.g., Fishbein & Ajzen, 1975). Instead, microstates are defined as *configurations* of the attitude elements. This view could align with—or at least engage—historical treatments of attitudes as sums of the expected values of the attributes the object is believed to have. In order to be configured in any particular way, some attention and processing must be allocated to the interrelations of attitudinal elements. Historically, it has been understood that such attention and processing is allocated to the extent that the attitudinal elements are discrepant (e.g., people attend more to messages regarding objects about which they are ambivalent; e.g., Maio, Bell, & Esses, 1996; people are motivated

to resolve inconsistencies; e.g., Cooper & Fazio, 1984; Festinger, 1957). Do microstates engage consistency considerations? "Configuration" does seem to imply relationships among elements, but it is unclear what those relationships are, and what processes act upon them.

Macrostates, on the other hand, are defined as "the number of positive vs. negative elements" without explicit considerations of relationships among elements. If macrostates are purported to describe a more global rather than local state, then it is difficult to reconcile the application of the concepts to attitudes as typically considered. Is there no consideration of the potential tension between positive and negative elements at the macrostate level? The authors claim that "global evaluation of an object is strongly related to the macrostate of an attitude." However, a global evaluation, at least as generally conceptualized, is more likely a product of reconciling potentially discrepant attitude-relevant information, a "summary evaluation" (Fazio, 1990) that distills relevant object-related information down to a single object-evaluation association. Or, it might involve the summation of attitude attributes, as in expectancy-value frameworks (e.g., Fishbein & Ajzen, 1975). Yet, it seems some such work must have already occurred at the microstate level if microstates include configurations. Thus, regarding functional properties of the attitude, how it is represented in memory, and what work is performed on the attitude or its elements, it is difficult to align micro and macro states of attitudes to existing conceptualizations of attitude representation.

We are also uncertain as to the distinction between the two different types of entropy as they apply to attitudes. The AE introduces Gibbs entropy as reflecting the stability of the attitude as a whole system, regardless of the underlying stability of its elements. Force applied to such a system is said to reduce Gibbs entropy, which results in a more stable summary attitude. Alternatively, Boltzmann entropy refers to quantity of combinations of elements that make up an

attitude that can be constructed that result in the same summary attitude. If we are understanding correctly (which we may not), if one likes chocolate, and is certain of that liking, they have low Gibbs entropy toward chocolate. Also, their attitude toward chocolate is made up of many elements, in this case more positive than negative, resulting in few combinations of these elements that would render a negative summary attitude. That most combinations of these elements result in the same evaluation indicates a low Boltzmann entropy toward chocolate. Put brusquely, this is confusing, and we don't know what this tells us about attitudes other than renaming concepts that already exist.

There were other instances where greater clarity would be appreciated. For example, the authors refer to the "energy of the attitudinal representation" (p. 3). Among models of memory in social cognition, representation is typically not described in terms of energy-related variables. It also stretched our thinking to ponder the statement, "the natural state of an attitude is neutral or ambivalent" (p. 8). Beyond the ambiguity surrounding the phrase, "natural state," the AE model does not appear to distinguish between neutral attitudes and ambivalent attitudes. This is an important distinction in attitudes research, beginning with Converse's (1970) distinction between "non-attitudes" and attitudes, and refined by Fazio's (1990) discussion of the "non-attitude-attitude continuum." People who feel neutrally (i.e., have no opinion) about an object tend to think and behave very differently from people who are ambivalent toward it. It would be useful to see the AE model explicitly compare and contrast an attitude with 0 positive and 0 negative elements (no opinion) to one with, say, 1 positive and 1 negative element, or 3 positive and 3 negative elements, etc. Psychologically, we know that these are very different sorts of attitudes.

Nor does the model seem to distinguish cognitive vs. affective bases of attitudes. Stronger attitudes (e.g., those capable of automatic activation) tend to be more affective in nature

(Rocklage & Fazio, 2018), and are often better predictors of behavior (Abelson, Kinder, Peters, & Fiske, 1982). Likewise, of experimentally-created attitudes, those founded on affect tend to be stronger (e.g., held with more confidence) than those founded on cognition (Edwards, 1990). When cognitive and affective components of attitudes are in conflict, affective components tend to "win out" in behavioral prediction (Lavine, Thomsen, Zanna, & Borgida, 1998). Given what we know about the general superiority of affective components regarding strength and behavioral prediction, it would be useful to know how the AE model accounts for the affective-cognitive distinction.

Advances in Understanding of Attitudinal Phenomena

The AE model proposes to explain dissociations between implicit and explicit measures of attitudes. Citing evidence of their low individual-level reliability, the authors propose that implicit measures are more likely to tap high entropy attitudes (or attitudes in a high entropy state), and that explicit measures, as a result of focused attention and thought, tap attitudes in a low entropy state. Our knee-jerk reaction to this claim 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. Implicit measures are generally more likely to tap affective reactions (e.g., Kendrick & Olson, 2012), while explicit measures are likely to tap propositional beliefs (there is even debate about the extent to which certain implicit measures assess affective vs. cognitive attitudinal components, e.g., Bar-Anon & Nosek, 2012; March & Graham, 2015). Implicit and explicit measures better align when people approach the explicit measures with an affective focus (Smith & Nosek, 2011). Furthermore, the predictive power of implicit vs. explicit measures on behavior has been shown to be at least partially the result of social desirability

concerns. This type of "control" over which component (implicit vs explicit) of an attitude is expressed does not seem to mesh with the idea that implicit attitudes are necessarily in a higher entropy state. There are a host of other differences between implicit and explicit measures—including differences in reliability—that the AE model appears to overlook, that account for their dissociation (Olson & Fazio, 2009).

More broadly, it appears that the AE conceptualizes attitudes as states rather than traits. This harkens another debate in the attitudes literature: the extent to which attitudes should be viewed as temporary constructions vs. stable representations in memory (Fazio, 2007; Schwarz, 2007). The resolution of this debate, at least to us, is that they can be either, depending on their strength. Strong attitudes are trait-like: they persist over time, are resistant to change, and reliably predict behavior. Among stronger attitudes, there should be little impact on the attitude of implicit vs. explicit measurement. A staunch coffee aficionado, for example, will have likely resolved inconsistencies among dependent attitudinal elements, and carry with them a low entropy attitude regardless of context. One could ask this person explicitly: "How much do you like coffee?" or have the person take a coffee IAT. We suspect that either way of inquiring about such a strong attitude would have little effect on its entropy.

Weaker attitudes are state-like, and are affected by temporarily accessible constructs, often as a function of context (e.g., weather, mood, e.g., Messner & Wänke, 2011; Schwarz & Clore, 1983). Prompting a respondent with an attitudinal inquiry may cause a weak or non-existent attitude to "consolidate" into a more integrated attitude, thus lowering its entropy (Fazio, Herr, & Olney, 1984). People consider objects and form attitudes toward them when they believe it is functional to do so. In cases where they have only minimal experience with an object, and no expected future interaction with it, there may be some attitudinal "elements" in

memory, but not an attitude—no summary evaluation of it. It would be considered a weak attitude by most definitions. However, when prompted with the question, "Do you like/dislike this object?", a consolidation process has been shown to ensue: people consolidate whatever information they have about the object into an evaluation that can be used to respond to the query. Research on attitudinal self-perception demonstrates just this (Fazio, 1987). Thus, there may be something to going 'full-physics,' only in a more limited scope than the AE model implies.

On the other hand, one interesting and potentially novel aspect of the AE model regards attitude *change* potential in the case of strongly-held attitudes. It says something about the field that the vast majority of "attitude change" research would be more accurately labeled "attitude formation" research. Studies of massive shifts in attitudes, cult indoctrination, and the like have largely been abandoned as our methods have increasingly retreated to the laboratory (Cialdini, 2009). How do strong attitudes change? The conventional wisdom is that they mostly don't. The question is an important one, and the AE model has potential to provide novel insight by way of its dependence parameter. Specifically, the claim is that reducing the dependence parameter has potential to make the individual more open to change. This harkens novel work by Baron (2000) on "intense indoctrination." In broadly surveying real-world scenarios involving massive shifts in attitudes, Baron describes the first step in such shifts as "softening up," where the individual is isolated from normal attitude support structures (i.e., family and friends), and deprived the cognitive resources (e.g., through sleep deprivation) to properly consider one's own attitudes and the elements that comprise them. Perhaps the dependence parameter captures and distills some essence of Baron's thinking on the matter. In any case, as the AE model does a service to the field by expressly considering possible ways to change strongly held attitudes.

We were disappointed with the AE Framework to the extent to which it is being offered as a novel conceptual tool to understand attitudinal inconsistency. Aside from terminology/notation, we see little conceptual difference between the AE Framework and Leon Festinger's (1957) Theory of Cognitive Dissonance. To paraphrase Festinger:

1. If two cognitive elements are relevant, they can be either dissonant (i.e., inconsistent) or consonant (i.e., consistent).

2. The degree of dissonance increases as the importance of the elements increases.

3. Dissonance is psychologically uncomfortable and gives rise to pressure to reduce the discomfort by reasserting consonance among the cognitive elements, which can be achieved by changing behavioral cognitive elements, changing environmental cognitive elements, and/or adding new cognitive elements. That is, persons can reduce dissonance by re-thinking aspects of the initially inconsistent elements.

4. Each of the latter routes of change have unique resistance and the element with less resistance is more likely to change.

The AE framework offers three propositions:

1. Inconsistency of an attitude is the Boltzmann entropy of the attitude.

2. Energy of the attitudinal representation serves as a local processing possibility to evaluate the global Boltzman entropy of an attitude. Attitude elements are likely to change when the opposite state has lower energy.

3. Focusing attention on the attitude object and thinking about the attitude object reduces the Gibbs entropy of attitude by increasing the networks dependence parameter.

Ignoring the underlying models, the AE propositions can be phrased as follows:

1. Entropy is inconsistency.

2. There is a drive to reduce inconsistency.

3. Thought reduces inconsistency by linking attitudinal elements together and shifting them toward consistency.

What does the AE framework reveal about human functioning that was not revealed by Cognitive Dissonance Theory? The authors attempt to reify their framework with simulations of existing effects. For example, they model the Mere Thought Effect (Tesser, 1978) which is the tendency for thinking about an attitude object to yield a more extreme attitude toward the object. But the simulations do not require novel predictions from the AE framework. The simulations simply restate the basic Mere Thought Effect via mechanical models. This is interesting, but there does not appear to be any new concepts offered. On page 24, the authors suggest that "the AE framework predicts that also an opposite mere thought effect exists, in the sense that when individuals are asked to very quickly answer attitude questions, attitudes are expected to be *less* polarized than when individuals are given more time to answer the questions." This is not new or surprising. If more thought leads to more polarization and people can have more/deeper thoughts when given more than less time, it seems to follow directly from the mere thought effect that less time would yield less thought and a less polarized attitude.

This general concern—that we were hard-pressed to uncover conceptual novelty—occurred to us throughout the paper. For example, on page 27, the authors predict that "persuasion is more effective for strong attitudes when the dependence parameter is lowered (e.g., by reducing attention directed at the attitude object) before the persuasion is employed." Indeed, the Cognitive Response approach of the 1960s and 1970s conceptualized such "reduced attention directed at the attitude object" as distraction (Festinger & Maccoby, 1964; Osterhouse & Brock, 1970).

Conclusion

It is laudable for a new model to attempt to account for well-documented findings as evidence of its predictive capacity. It is far more valuable, however, when such models make novel predictions about heretofore unobserved phenomena and offer novel conceptual insight. Our impression, to return to the opening quote from Platt (1964), is that the mathematical box offered by the AE framework has yet to yield much beyond what has already been caught by the logical box of the existing attitudes literature.

We began our commentary by expressing concern that models of the physical world are at risk of being over-applied to the psychological world. It strikes us as evidence of the over-application of physical to psychological models that the authors of the target article should conclude with the claim that the answer to the question of why we think is "to reduce the entropy of our mental representations." Instead, as attitudes and social cognition researchers, we subscribe to the doctrine that attitudes are inherently functional insofar as they are "ready aids" to action in a complex world (Smith, Bruner, & White, 1956, p. 41), and that "thinking is for doing" (Fiske, 1992, p. 877).

References

Abelson, R. P., Kinder, D. R., Peters, M. D., & Fiske, S. T. (1982). Affective and semantic components in political person perception. *Journal of Personality and Social Psychology*, *42*, 619-630.

Baron, R. S. (2000). Arousal, capacity, and intense indoctrination. *Personality and Social Psychology Review*, *4*, 238-254.

Busemeyer, J. R., & Wang, Z. (2015). What is quantum cognition, and how is it applied to psychology? *Current Directions in Psychological Science, 24*, 163-169.

Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science, 4*, 5-6.

Converse, P. E. (1970). Attitudes and non-attitudes: Continuation of a dialogue. In E. R. Tufte (Ed.), *The quantitative analysis of social problems* (pp. 168-189). Reading MA: Addison-Wesley.

Cooper, J., & Fazio, R. H. (1984). A new look at dissonance. *Advances in Experimental Social Psychology*, *17*, 229-268.

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

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

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In *Advances in experimental social psychology* (Vol. 23, pp. 75-109). Academic Press.

Fazio, R. H. (1987). Self-perception theory: A current perspective. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social Influence: The Ontario Symposium* (vol. 5, pp. 129-150). Hillsdale, NJ: Erlbaum.

Fazio, R. H., Herr, P. M., & Olney, T. J. (1984). Attitude accessibility following a self-perception process. *Journal of Personality and Social Psychology*, *47*, 277-286.

Festinger, L. (1957). A theory of cognitive dissonance. *Evanston, Il: Row, Peterson*.

Festinger, L., & Maccoby, N. (1964). On resistance to persuasive communications. *The Journal of Abnormal and Social Psychology*, *68*, 359-366.

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

Fiske, S. T. (1992). Thinking is for doing: portraits of social cognition from daguerreotype to laserphoto. *Journal of Personality and Social Psychology*, *63*, 877-889.

Kendrick, R. V., & Olson, M. A. (2012). When feeling right leads to being right in the reporting of implicitly-formed attitudes, or how I learned to stop worrying and trust my gut. *Journal of Experimental Social Psychology*, *48*, 1316-1321.

Kvaale, E.P., Haslam, N., & Gottdiener, W.H. (2013). The 'side effects' of medicalization: A meta-analytic review of how biogenetic explanations affect stigma. *Clinical Psychology Review*, *33*, 782-794.

Lavine, H., Thomsen, C. J., Zanna, M. P., & Borgida, E. (1998). On the primacy of affect in the determination of attitudes and behavior: The moderating role of affective-cognitive ambivalence. *Journal of Experimental Social Psychology*, *34*, 398-421.

Maio, G. R., Bell, D. W., & Esses, V. M. (1996). Ambivalence and persuasion: The processing of messages about immigrant groups. *Journal of Experimental Social Psychology, 32*, 513-536.

March, D. S., Gaertner, L., & Olson, M. A. (2018). On the prioritized processing of threat in a dual implicit process model of evaluation. *Psychological Inquiry*, *29*, 1-13.

March, D. S. & Graham, R. (2015). Exploring implicit ingroup and outgroup bias toward Hispanics. *Group Processes and Intergroup Relations, 18,* 89-103.

Messner, C., & Wänke, M. (2011). Good weather for Schwarz and Clore. *Emotion, 11*, 436-437.

Olson, M. A., & Fazio, R. H. (2009). Implicit and explicit measures of attitudes: The perspective of the MODE model. In Petty, R. E., Fazio, R. H., & Briñol, P. (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 19-64). Mahwah, NJ: Erlbaum.

Osterhouse, R. A., & Brock, T. C. (1970). Distraction increases yielding to propaganda by inhibiting counterarguing. *Journal of Personality and Social Psychology*, *15*, 344-358.

Pentland, A. (2015). *Social Physics: How social networks can make us smarter*. Penguin.

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347-353.

Rocklage, M. D., & Fazio, R. H. (2018). Attitude accessibility as a function of emotionality. *Personality and Social Psychology Bulletin*, *44*, 508-520.

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

Schwarz, N. & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513-523.

Smith, M. B., Bruner,J. S., & White, R. W. (1956). *Opinions and personality.* New York: Wiley.

Smith, C. T., & Nosek, B. A. (2011). Affective focus increases the concordance between implicit and explicit attitudes. *Social Psychology*, *42*, 300-313.

Tesser, A. (1978). Self-generated attitude change. *Advances in Experimental Social Psychology, 11*, 289-338.