Journal of Motivation, Emotion, and Personality
Vol. 14 (2026), pp.1–6

© 2026 by Yan Yang
DOI: 10.12689/jmep.2026.1401

Modeling Limits in Phenomenological Contexts: A Statement on Luo et al.'s "Improving Searcher Struggle Detection via the Reversal Theory"

Yan Yang
Independent Researcher

This paper grows out of the author's confusion during a project applying Reversal Theory (RT) to detecting users who are struggling with web search and offers a critical analysis of the resulting publication. Searcher struggle detection provides a measure of user satisfaction valuable to search engine providers. It statistically examines logged user data, that is, user behavior. But as a form of structural phenomenology, RT rejects the reduction of motivational states to overt behavior. And yet it equally resists the classical phenomenological position that experience is irreducibly unclassifiable. It articulates the internal experience through bimodal state pairs, while keeping that "structure" at the level of description rather than quantitative measurement. This paper examines modeling decisions in Luo et al. and identifies four theoretical misalignments. First, equating increased effort with increased arousal violates RT's treatment of arousal as subjective experience. Second, conflating RT's curve of the relationship between arousal and hedonic tone with a binned statistical distribution in the data is a conceptual confusion at mathematical form level. Third, like the arousal issue, inferring meta-motivational states from query diversity or topic category substitutes external features for the self-report basis RT requires. Fourth, appealing to RT's topological diagrams to justify excluding one of RT's four dimensions ("rules") from the model grounds quantitative inferences on structural descriptions without numerical meaning. Despite these misalignments, Luo et al.'s model retains predictive utility, owing largely to statistical regularities in the data discovered empirically using RT constructs. The deeper tension is twofold: Firstly, RT's structural phenomenology leaves room for formalization but stops at topological description, so any quantitative extension requires additional definition that cannot be read directly from RT. Secondly, since experiential RT constructs must be grounded in subjective self-report, models that substitute behavioral or environmental variables for this basis are best regarded as RT-inspired approximations rather than theory implementations. Future work may draw on neurophenomenology's first-/third-person dual-source method of mutual constraints, and on the theoretical grounding that embodied cognition frameworks offer, to advance formalization while preserving the self-report foundation—maintaining an explicit reflexivity about which of RT's commitments are and are not amenable to formalization. These considerations carry direct methodological implications for the growing body of work in human-computer interaction and computational psychology that seeks to formally model motivational and experiential states.

Keywords: Reversal Theory, Structural Phenomenology, Searcher Struggle Detection, Behavioral Data, Statistical Operationalization, Computational Modeling, Human-Computer Interaction, Epistemology of Modeling, Information Retrieval, Psychology

Unresolved Confusion from the Statistical Application of Reversal Theory

Back in 2019, when I was a master's student, I was asked to assist with a collaborative project that later resulted in a publication claiming to apply Reversal Theory (RT) to user search log classification (Luo et al., 2024). I initially joined to help with writing but was soon assigned to refine the statistical terms translated from RT so that they could be justified for implementation over log data. The decision to use RT and to operationalize it statistically was made by others. It was not based on my interests or decisions. Nevertheless, I took the task seriously and began reading related psychological literature.

Yan Yang, Independent Researcher. Correspondence concerning this article should be addressed to Yan Yang. E-mail: yan.yang.research@proton.me

Even with my limited background in psychology, I quickly sensed a mismatch between the theory and its practical application. RT emphasizes that individuals switch between different meta-motivational states over time (Apter, 2001a). However, in the modeling draft I reviewed, all search records from a single user were assigned to a single fixed motivational type, implying stable internal states across different sessions—an apparent oversimplification.

I suggested a technical adjustment: instead of fixing one state per user, we allowed both telic and paratelic modes to appear for the same user. My research advisor accepted this and asked me to implement it in statistical terms. While this made the model more flexible, I still felt that the core of RT was not preserved. At the time, I couldn't clearly express the problem and stepped back from further involvement. As a result, I did not take part in the later stages of the project and did not have the chance to read the full manuscript before its publication.

Recently, through further work, I have preliminarily clarified the confusion; however, I have also found it to be more nuanced than I initially thought. In line with the principle that "all models are wrong, but some are useful" (Box, 1976), my current conclusion is that RT does offer a potentially useful framework for modeling searcher struggle detection, but the strength of any author's claim to implement RT depends on how closely the operationalization satisfies RT's theoretical commitments. Where those commitments (notably, grounding state identification in subjective experience rather than behavioral proxies) cannot be met, the claim to be "applying RT" needs to be relaxed accordingly, and the resulting model treated as an RT-inspired approximation rather than a faithful implementation of the theory.

This brief article examines that gap in the case of the searcher struggle detection study and, while grounded in a specific modeling scenario, offers tentative guidelines specifically for translating structurally oriented Continental phenomenological theories (like reversal theory) into quantitative computational frameworks. Its insights apply most directly to computational human-computer interaction (HCI) and formal motivational modeling, with limited transfer to general consciousness research absent supplementary phenomenological measurement tools.

RT as a Case of Structural Phenomenology

Theoretically, RT is best understood as a form of structural phenomenology (Apter, 1981)—a hybrid position that inherits the phenomenological tradition's commitment to grounding motivational states in first-person subjective experience, yet departs from that tradition by imposing a structuralist architecture of recurring bimodal states onto that experience. The sections below trace this dual inheritance, first against the Continental phenomenological tradition and its behaviorist counterpart, then against the demands of quantitative modeling.

Classical phenomenological thought, rooted in the Continental tradition, promotes an epistemological stance centered on the phenomenological epoché—the methodological bracketing of assumptions about external reality (Stumpf & Fieser, 2014, Chapter 19). It suspends judgment on whether an objective world exists, in order to attend to the structures of lived experience as they are given to consciousness. In doing so, it detaches from third-person perspectives and resists any approach that reduces lived experience to externally observable proxies, as behaviorism does.

Historically, RT was developed through a critical engagement with Hebb's model of arousal (1955). It departs from Hebb on its uni-modality, a point we return to below. Here, following the philosophical lineage, Hebb shifted away from purely behaviorist models toward recognizing intrinsic motivation, a theme that resonates with some phenomenological ideas about the importance of subjective experience and intentionality. However, Hebb does not place subjective experience at the core of his research (Hebb, 1955). RT takes a further step by placing subjective experience and self-interpretation at the heart of theoretical construction (Apter, 1981). Apter states that "such states of mind cannot be defined through overt behavior, but must always be defined phenomenologically" (Apter, 1981).

Yet RT does not embrace this tradition in full. Classical phenomenology also resists rigid categorization, treating lived experience as something that ultimately eludes fixed classification (Stumpf & Fieser, 2014). It is here that RT parts company with it, taking a more hybrid position that fuses the phenomenological tradition with structuralism (Apter, 1981). Specifically, RT emphasizes the "structural" aspect of phenomenology by seeking to identify recurring patterns and relationships within lived experience (Apter, 1981). A clear example is its use of binary coding, such as the four pairs of opposing motivational states. This approach aligns in part with the aims of statistical modeling, which also seeks to delineate meaningful relationships.

The relationship between RT's structuralism and mathematical modeling, however, is not a simple yes-or-no situation. Both seek to identify and articulate structure, yet they do so in different registers: structuralism, in RT's context, characterizes structure qualitatively, through the opposition and ordering of states, whereas mathematical modeling typically presupposes that such structure can be rendered in quantitative, measurable terms. The two therefore converge on the aim of delineating structure while diverging on the question of quantification. This is a question on which RT takes no explicit position, neither requiring such a step nor ruling it out. It is in this gap that the alignment becomes partial rather than complete.

In such structural-phenomenological form, RT defines itself against two opposite over-commitments: on one side, the reduction of experience to externally observable behavior (as in the pure Stimulus-Response model), which denies experience any defining role; on the other, the radical subjectivism of classical phenomenology, on which experience is ultimately incommunicable and resists all classification. RT takes lived, first-person experience as the sole ground of motivational states, yet treats that experience as exhibiting an identifiable structure. The pull between the fluid and the structured is a tension RT chooses to work within rather than resolve. This tension was the source of productive ambiguity I experienced in our 2019 research.

In practice, several psychometric scales have been developed to operationalize aspects of RT, measuring either dominance types or meta-motivational states at a single point in time (Reversal Theory Society, n.d.; O'Connell & Calhoun, 2001; Apter et al., 1998). These instruments focus on self-reported subjective experience rather than observed behavior, in line with the theory's rejection of behavioral definitions of motivational states. They also acknowledge the inevitable simplification involved in quantifying experience and are designed for specific populations rather than claiming universal applicability.

These two constraints—a reliance on subjective self-report and an unwillingness to compress experience that varies richly across diverse populations into a reduced set of variables—partly explain why large-scale data analyses involving RT remain rare. The theory's methodological commitments impose limits on the scale and structure of the data it can meaningfully accommodate.

A Review of the Modeling Decisions in the Searcher Struggle Detection Study

At the core of the searcher struggle detection study is the adoption of RT's bi-modal arousal-hedonic model. Specifically, the authors mapped the telic/paratelic distinction from RT onto user behavioral data, challenging struggle detection's traditional association of "more effort" with "more struggle." The other three pairs were treated by the authors as irrelevant in this specific scenario, whether on theoretical or empirical grounds. In their statistical operationalization, several key modeling choices were made:

- Arousal level = Effort level: Arousal, a core RT construct relating to subjective excitement or tension, was treated as equivalent to observable user behaviors such as frequency of queries, clicks, etc.
- Hedonic level = "Happiness" = Session concentration: The RT "hedonic level" was equated with "happiness" and operationalized by binning sessions along a behavioral effort feature score. The behavioral effort feature score is taken to represent arousal level, and more sessions in a bin indicate higher happiness at that arousal level.
- Motivational state = Session attributes: The "telic" and "paratelic" states were assigned based on the detected session topic (e.g., career or entertainment), while "conformist" and "negativist" states were associated with a lower versus higher *ExploreScore*, respectively, which was calculated from session behavioral features.
- Modal dimension selection: As noted, of the four bi-modal dimensions in RT, three were excluded during modeling. The transaction and relationship dimensions were omitted because they require an interpersonal context that is not present in individual search tasks. The rules dimension was disregarded based on statistical tests of the *ExploreScore* and the authors' reading of Apter (2001a, p. 44, Fig. 2.2).
- "Anxiety"/"Manic" both treated as struggling: Sessions with low hedonic/happiness level at high effort/arousal were labeled "anxiety" and "manic", for telic and paratelic respectively, each falling at a different point along the arousal axis, since the peaks of the two curves are offset. Both categories are operationalized as "user struggling."

Two of these choices are well-grounded. Treating "anxiety" and "manic" as struggling is sound, since both are states RT itself identifies. Substituting "happiness" for "hedonic level" seemingly introduces ambiguity because "happiness" is a philosophically loaded, multidimensional term that resists reduction to a single variable (Aristotle, ca. 350 B.C.E./2009; Locke, 1690/1975). However, within RT this is at least partly mitigated: conditioning the hedonic level on the motivational state adds a layer of differentiation, so the substitution is defensible. However, certain modeling practices may reflect underlying theoretical misalignments, which in turn introduce challenges at both the conceptual and methodological levels.

1. Equating arousal level with effort level is a significant leap, which assumes that maximal behavioral effort reflects optimal arousal for all users and contexts. However, neither RT nor Hebb's arousal theory supports this relation. In RT, arousal is itself a layer of lived experience (with the hedonic level a second, interpretive layer built upon it) (Apter, 2001b). With arousal being a matter of experience, the mapping violates the phenomenological tradition of not defining subjective experiences in terms of overt behaviors. Even when we turn to RT's behaviorist-neuropsychological predecessor, Hebb's model of arousal, we find that Hebb used an inverted-U shape rather than a linear relationship between arousal and behavioral effectiveness (Hebb, 1955).[1]

2. The theoretical "inverted-U," a functional relationship between arousal and hedonic level, was conflated with the statistical notion of a Gaussian distribution across session count bins. The inverted-U itself marks a second point on which RT departs from Hebb (the first being its placement

[1] Hebb used the same inverted-U shape to describe this arousal-behavioral relation and the arousal-hedonic relation. He sees both as related manifestations of the same underlying process.

of subjective experience at the core, as noted earlier): rather than the single, universal optimum of optimal-arousal theory, RT posits two state-dependent optima—"two inverted U-curves instead of the single inverted U-curve of optimal arousal theory" (Apter, 2001a, p. 23).[2] What concerns us here, however, is not which curve RT endorses but the conflation of any such theoretical curve with a statistical distribution. To distinguish the two concepts and then link them, consider a stable arousal-hedonic relationship within a given meta-motivational state and assume people are motivated to maximize their hedonic experience. Reflecting a tendency toward self-regulation around the "sweet spot(s)" of experience, the Gaussian distribution emerges as a secondary artifact. It is plausible that the distribution of session counts across observed effort-level bins would peak near the optimal hedonic point, resulting in the mentioned Gaussian-like distribution. However, while this link exists, real-world data is influenced by individual factors and empirical patterns and may diverge from the theoretical curve. Most importantly, RT itself offers no statistical description of how the behavioral data should be distributed. The theoretical arousal-hedonic curve and the empirical frequency distribution remain distinct kinds of object.

3. Using behavioral features, such as those composing the *ExploreScore*, to infer meta-motivational states is conceptually problematic. The paper calculates the *ExploreScore* from behavioral metrics (e.g., query diversification and engagement with rare queries or results) and uses it to classify sessions as conformist or negativistic. Yet in RT, metamotivational states are primary: they shape how individuals interpret and enact behavior, and are not themselves read off from it (Apter, 2001b). Recovering states from observable behavior alone is thus a data-driven approximation rather than an application of the phenomenological theory; however effective empirically, it departs from RT's commitment to grounding state identification in subjective reports and lived experience. Such mappings are best regarded as secondary constructions, liable to confuse behavioral patterns with motivational structures.

4. Similarly, the paper uses the topic of a search session (such as "job" or "entertainment")—categories that were determined independently of the behavioral metrics—to assign sessions to telic or paratelic states. It overlooks the fact that topic does not always determine the motivational state: the significance of a session's topic for the individual can vary widely, and RT explicitly argues that states must be defined by subjective interpretation, not by topic category or external context (Apter, 1981). Therefore, these modeling practices trade off theoretical rigor for convenience, departing from the phenomenological roots of RT.

5. To justify excluding the "rules" dimension, the paper appeals to Apter's diagrams—reading the conformist and negativistic curves (Apter, 2001a, p. 44, Fig. 2.2) as showing that struggling occurs at the same effort level in both, and hence that the dimension is irrelevant. But these diagrams are qualitative: they specify only the presence of bimodal structures and the qualitative relationships among states (Apter, 1981), without numerical values or cross-graph comparability. A figure of this kind cannot establish that the two optima coincide, and so cannot support the conclusion that the rules dimension is irrelevant. (The exclusion may well be justified on dataset-specific grounds, but that is an empirical matter; it is not something the RT literature itself licenses.)

Nevertheless, despite these theoretical concerns, the empirical model demonstrates predictive utility. This apparent effectiveness can largely be attributed to several pragmatic aspects of the modeling process. Labeling each session as a whole according to its topic reduces granularity but is sufficient to identify broad motivational tendencies, even though it misses within-session motivational reversals and cross-context variation, where the same topic triggers different meta-motivational states across users or separate searches by the same user. The behavioral and contextual features selected, such as the *ExploreScore* and session topic classification, leverage statistical regularities in real-world data to yield useful approximations. Ultimately, these outcomes reflect the practical benefits of empirically driven feature engineering rather than strict adherence to the theoretical intentions of RT.

Reflections and Discussions

While the original paper acknowledges practical challenges such as imperfect detection of the positive class (struggling sessions), noise in interaction logs, and limited generalizability across users and platforms, these concerns largely remain internal to the statistical modeling framework. In contrast, my primary concern is structural and epistemological: translating key aspects of RT into purely mathematical or categorical models risks oversimplifying and misrepresenting the theory's original intent and nuance (Apter, 1981). This applies most directly to its grounding of motivational states in subjective interpretation, which resists behavioral proxy. It also applies to the dynamics of reversal, for which RT offers only descriptive accounts of triggering conditions, without even the topological structure that grounds the arousal-hedonic model (Apter, 2013).

Yet, this does not imply that modeling is entirely incompatible with RT. RT's structural commitments are themselves amenable to formalization: the bistable states and their qualitative relations can in principle be represented in formal terms. The deeper issue is one of data source rather than of

[2] RT's standard depiction is an X-shaped pair of "butterfly" curves (Apter, 2001a, p. 21); the two inverted-U curves arise as an extension Apter raises as possible when the arousal axis is pushed to its extremes (p. 23). Luo et al. adopt this inverted-U version.

formalization as such—the problems identified above arise not because RT's structures resist formal treatment, but because the operationalizations reviewed here substitute behavioral proxies for the subjective-experiential grounding that RT requires. Improvements in data quality and technical implementation may yield a better empirical fit, but they do not address this deeper issue. The practical upshot is that models working with behavioral proxies should explicitly acknowledge which of RT's theoretical commitments have been set aside. Models should treat their outputs as approximations rather than as implementations of the theory, with the scope and limitations of those approximations made explicit.

Statement and Future Work

This brief note outlines the conceptual mismatch uncovered during my early participation in the RT-based searcher-struggle detection project. I offer these observations for scholars working at the intersection of structural phenomenology and statistical modeling.

Future inquiry can build on this case study to resolve the core tension identified here: reconciling first-person experiential commitments of structurally oriented phenomenology with forms of mathematical and statistical modeling. Crucially, this line of work only extends to Continental frameworks that formalize recurrent experiential structures (the subset exemplified by reversal theory); it does not purport to bridge anti-formalist strands of Continental thought that reject all quantifiable operationalization outright. Concrete intermediate avenues already exist to advance this agenda. Neurophenomenology (Varela, 1996) formalizes a rigorous dual-source method that places mutual constraints between disciplined first-person subjective experiential reports and third-person behavioral and physiological recording. Meanwhile, embodied cognitive frameworks (Foglia & Wilson, 2013) provide theoretical grounding for taking both subjective report and behavior seriously. These methodological bridges will grow more vital as HCI and computational psychology increasingly aim to formalize nuanced motivational and experiential states, rather than only observable surface behaviors.

Acknowledgments

The author thanks Professor Grace Hui Yang and Dr. Jiyun Luo for the opportunity to participate in the project that ultimately gave rise to this analysis. She is grateful to Professor Richard Squier for encouraging her to write up and organize the reflections that developed after the original work was published. She also thanks the anonymous reviewers for their comments on an earlier draft, which prompted a further round of scrutiny and reconsideration, and led to substantial revisions in both the framing and the formulations of the paper.

AI Use Statement

The author used AI-assisted tools for language checking and refinement. All intellectual content, analysis, and conclusions are the author's own.

References

Apter, M. J. (1981). The possibility of a structural phenomenology: The case of reversal theory. *Journal of Phenomenological Psychology, 12*(2), 173–187.

Apter, M. J. (Ed.). (2001a). *Motivational styles in everyday life: A guide to reversal theory*. American Psychological Association. https://doi.org/10.1037/10427-000

Apter, M. J. (2001b). *Rethinking motivation and personality: An introduction to reversal theory*. PSW Science. https://pswscience.org/meeting/rethinking-motivation-and-personality-an-introduction-to-reversal-theory/

Apter, M. J. (2013). Developing reversal theory: Some suggestions for future research. *Journal of Motivation, Emotion, and Personality*, 1(1), 1–8. https://doi.org/10.12689/jmep.2013.101

Apter, M. J., Mallows, R., & Williams, S. (1998). The development of the motivational style profile. *Personality and Individual Differences, 24*(1), 7–18. https://doi.org/10.1016/S0191-8869(97)00148-7

Aristotle. (2009). *The Nicomachean ethics* (W. D. Ross, Trans.; L. Brown, Rev. ed.). Oxford University Press. https://doi.org/10.1093/actrade/9780199213610.book.1 (Original work published ca. 350 B.C.E.)

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association, 71*(356), 791–799. https://doi.org/10.1080/01621459.1976.10480949

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(3), 319–325. https://doi.org/10.1002/wcs.1226

Hebb, D. O. (1955). Drives and the C.N.S. (conceptual nervous system). *Psychological Review, 62*(4), 243–254. https://doi.org/10.1037/h0041823

Locke, J. (1975). *An essay concerning human understanding* (P. H. Nidditch, Ed.). Oxford University Press. https://doi.org/10.1093/actrade/9780198243861.book.1 (Original work published 1690)

Luo, J., Yang, Y., Nayak, V., & Yang, G. H. (2024). Improving searcher struggle detection via the reversal theory. *Discover Computing, 27*, Article 51. https://doi.org/10.1007/s10791-024-09492-z

O'Connell, K. A., & Calhoun, J. E. (2001). The telic/paratelic state instrument (T/PSI): Validating a reversal theory measure. *Personality and Individual Differences, 30*(2), 193–204. https://doi.org/10.1016/S0191-8869(00)00026-X

Reversal Theory Society. (n.d.). *Reversal theory research instruments*. https://reversaltheory.net/wp-content/uploads/RT-Instruments.pdf

Stumpf, S. E., & Fieser, J. (2014). *Philosophy: A historical survey with essential readings* (9th ed.). McGraw-Hill.

Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies, 3*(4), 330–349.