Journal of Mathematical Psychology 91 (2019) 195–200

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

State-trace analysis misinterpreted and misapplied: Reply to Stephens, Matzke, and Hayes (2019)

F. Gregory Ashby

University of California, Santa Barbara, United States

HIGHLIGHTS

- Stephens, Matzke, and Hayes (2019) misinterpreted and misapplied state-trace analysis.
- They reported no evidence either for or against single or multiple systems.
- Their conclusion that a single-system account is sufficient was preordained.
- Their use of parsimony contradicts state-trace analysis and the field of statistics.
- Their empirical contribution reinforces a point that was always universally accepted.

ARTICLE INFO

Article history:
Received 11 May 2019
Received in revised form 23 June 2019
Accepted 8 July 2019
Available online 17 July 2019

Keywords:
Categorization
State-trace analysis
COVIS

ABSTRACT

After using state-trace analysis to reanalyze results from 63 different categorization studies, Stephens, Matzke, and Hayes (2019) concluded that "the evidence for two distinct category learning systems is much more limited and inconsistent" (p. 14) than Ashby and Valentin (2017) had previously claimed. This reply shows that Stephens et al. (2019) misinterpreted and misapplied state-trace analysis. They report no evidence that favors a single learning system over multiple systems. They acknowledge that they would favor a single-system account, regardless of how their re-analyses had turned out. They justify this bias by claiming that single-system theories are more parsimonious than dual-systems theories, but they use a definition of parsimony that is inconsistent with state-trace analysis, and with the entire statistical field of model selection. By any accepted definition of parsimony, the dual-systems COVIS model is more parsimonious than the single-system model they favor in the current applications. The correct interpretation of their results is that none of the 63 studies they examined, by itself, definitively identifies the number of parameters that are varying across the conditions of that study. However, this was never an issue of contention, and was stated explicitly in prior publications.

1. Introduction

Stephens, Matzke, and Hayes (2019, hereafter referred to as SMH) used state-trace analysis to reanalyze results from many different published reasoning and category-learning studies. The category-learning database included 63 studies (reported in 28 different articles), many of which were run in my lab. This reply focuses on the SMH re-interpretation of these 63 studies.

The 28 category-learning articles were all testing predictions of the dual-systems model of category learning called COVIS (Ashby, Alfonso-Reese, Turken, & Waldron, 1998) by examining performance in rule-based (RB) and information-integration (II) categorization tasks. SMH concluded that "the troubling consequence (of their reanalysis) is that many interaction effects cited as evidence for multiple processes may have been over-interpreted" (p. 4), and that "these state-trace analyses show that the evidence for two distinct category learning systems is much more limited and inconsistent than is implied by the impressive list of dissociations presented by Ashby and Valentin (2017)" (p. 14). They also concluded that "instead of two learning systems, a single latent variable such as 'degree of learning' – or a dimension-weighting parameter as included in the Generalized Context Model (GCM) of categorization (Nosofsky, 1986) – would often be sufficient to account for the results" (p. 14).

In this reply, I show that SMH misinterpreted and misapplied state-trace analysis. Their re-analyses found no evidence favoring a single learning system over multiple systems, nor did they identify any weaknesses or mispredictions of any multiple systems theories. In fact, state-trace analysis was never designed to identify the number of underlying systems, and it is known that there are no state-trace plots that can provide any information

E-mail address: fgashby@ucsb.edu.

https://doi.org/10.1016/j.jmp.2019.07.001
0022-2496/

about this issue (Ashby, 2014; Dunn, Kalish, & Newell, 2014). Despite these facts, SMH acknowledged that they would have concluded that a single-system account should be favored over COVIS, regardless of how their re-analyses turned out. As a result, their conclusions were preordained. They justify this bias by claiming that single-system theories are more parsimonious than dual-systems theories, without ever defining what they mean by parsimony. In the third section, I show that their use of this term is inconsistent with how state-trace analysis defines parsimony, and with the entire statistical field of model selection. If any of these usual definitions of parsimony are adopted, then in the applications that SMH consider, the dual-systems COVIS model is more parsimonious than the single-system GCM.

The correct interpretation of the SMH results is that none of the 63 studies they examined, by itself, definitively identifies the number of underlying free parameters that are varying. However, this was never an issue of contention. For example, this has been known for at least 60 years and in the case of the 28 articles examined by SMH, made explicitly in prior publications. Furthermore, the number of free parameters that are varying is irrelevant to the debate about whether RB and II category learning are mediated by one or two systems, or whether COVIS or the GCM provides the better account of these data.

State-trace analysis (Bamber, 1979; Dunn & Kirsner, 1988) plots performance on two tasks across various experimental conditions against one another and examines the resulting scatterplot. On the basis of the type of scatterplot that emerges, inferences are then made about the number of underlying free parameters that are varying across the different conditions. For example, suppose the same participants complete two tasks, T_1 and T_2. Let $P(T_1)$ and $P(T_2)$ denote their performance on tasks T_1 and T_2, respectively. A state-trace analysis begins by plotting values of $P(T_1)$ and $P(T_2)$ against each other (e.g., with values of $P(T_2)$ on the ordinate and values of $P(T_1)$ on the abscissa). Although many different outcomes are possible, SMH focus on only two. In their opinion, "the crucial question is whether the state-trace is 'one-dimensional', with all data points falling on a single monotonically increasing (or decreasing) curve. If so, the data points are consistent with a single underlying latent variable" (p. 6). They describe the other possibility as follows: "if instead the state-trace is two-dimensional (i.e., some of the data points reliably depart from monotonicity), the data are inconsistent with any model based on a single latent variable. Thus, they may support a dual-process account" (p. 6).

These are both misinterpretations or mis-statements of state-trace analysis. Section 2 considers the first claim — that is, about 'one-dimensional' state-trace plots, Section 3 considers SMH's use of the word parsimony, and then Section 4 considers the claims of SMH about 'two-dimensional' plots.

2. One-dimensional state-trace plots

The claim that one-dimensional state-trace plots "are consistent with a single underlying latent variable" (p. 6) is a misleading interpretation of state-trace analysis. The correct mathematical statement is that if a state-trace plot is one-dimensional then no inferences are possible because the data could have been generated by a model that is characterized by any number of latent variables (Ashby, 2014; Bamber, 1979; Dunn & Kirsner, 1988). For example, Fig. 2A shows a one-dimensional state-trace plot that was produced by a model with 3 free parameters. So a state-trace analysis can only be used to rule out a subset of competing models if the state-trace plots are two-dimensional because these are the only types of plots that make any inferences possible about the underlying processes.

SMH reanalyzed data from 63 different published categorization studies using a state-trace approach. In most cases they concluded that the state-trace plot was one dimensional. Thus, their results indicate that a state-trace analysis can add nothing new to this literature because a one-dimensional plot neither supports nor rules out any underlying model.

So what do the SMH state-trace analyses contribute? Well, they do show that each of the 63 different data sets, when considered in isolation, does not conclusively identify the number of underlying latent variables that are varying. If this was a novel inference, then it would make a useful contribution. But in almost all cases, the 63 studies report a single dissociation, and it has been known for at least 60 years that a single dissociation, by itself, is insufficient to rule out all alternative interpretations (e.g., Teuber, 1955). Furthermore, to my knowledge, none of the articles reporting the 63 studies claimed that their study, by itself, ruled out all alternative interpretations — nor even all single-system interpretations. For example, Ashby (2014) noted that "it is also important to acknowledge that more traditional dissociation logic is also flawed, in the sense that it is rarely the case that any particular dissociation (or lack thereof) can conclusively favor or rule out either one or multiple systems" and also that "a careful examination of each dissociation in isolation would likely show that that one result, by itself, was, at best, only weakly diagnostic with respect to the single- versus multiple-systems question" (p. 943). So the SMH analyses are a solution in search of a problem. They make no new theoretical or empirical contribution. They merely reinforce a point that was always universally accepted.

So if the 63 categorization studies are not each definitive about the number of underlying processes, then what contribution do they make? Suppose you are trying to determine if a coin is fair or biased towards heads. Any one coin toss provides inconclusive evidence about this question. But if 63 tosses in a row all come up heads, then collectively all these data strongly suggest that the coin is biased and not fair. Science is a cumulative process. It is highly unlikely that any single study will prove definitive, especially in psychology. So a converging operations approach is required. As Ashby (2014) noted "as more and more data are collected, it is vital to consider what theory or model is most consistent with the entire body of available data" (p. 943). This is why Ashby and Valentin (2017) did not claim that any one of the 63 studies, by itself, was definitive, but rather that "*collectively*, these results also provide strong evidence that learning in these tasks is mediated by separate systems" (p. 175; emphasis added).

SMH reject this approach to science. They make no attempt to show that any single-system model can account for any of the data from the 63 studies they examined, nor do they identify any weaknesses or mispredictions of COVIS. Instead, they argue that until some single experiment is produced that definitively rules out all other interpretations, dual-systems models like COVIS (Ashby et al., 1998) should be rejected in favor of some unspecified single-system account — despite the fact that COVIS easily accounts for the results of all 63 studies simultaneously. In my opinion, the SMH approach to science – that is, of waiting for the perfect experiment – would prevent psychology from ever making scientific progress.

SMH interpret their results to favor single-system accounts of category learning over dual-systems — despite their failure to produce any mathematical, theoretical, logical, or empirical evidence that favors a single system. This bias is evident in their misleading interpretation that one-dimensional state-trace plots "are consistent with a single underlying latent variable" (p. 6). Note how much the tenor of the SMH article would have changed if they had instead written the following equally valid interpretation of a one-dimensional plot: a one-dimensional plot is consistent with two underlying latent variables; or if they had written that a one-dimensional plot is consistent with 10 underlying latent variables.

So despite the fact that no inferences are possible from a one-dimensional state-trace plot, they interpret this result as favoring the single-system GCM over the dual-systems COVIS. But in addition, they also argue that two-dimensional state-trace plots "may also be consistent with 'single-process' accounts that posit multiple parameters" (p. 18). More specifically, they recommend endorsing the GCM over COVIS if the state-trace plot is either one-dimensional or two-dimensional. Since, by their definition, these are the only two possible outcomes of a state-trace analysis, their conclusion that the GCM is a better model of categorization than COVIS was preordained. By their own admission, there was no possible set of outcomes that they could have observed that would have led them to a different conclusion. So why even complete the analyses?

3. Parsimony

Why are SMH biased toward a single-system account of category learning? Their article allocates only one word to explain this bias — parsimony. Specifically, SMH argue that in cases where the state-trace plot allows no inferences, then "a single-process theoretical account may be preferred on the grounds of parsimony" (p. 18). So according to SMH, parsimony trumps all, even the data from 63 separate studies. Given the extraordinary importance they assign to this principle, it is unfortunate that they did not even allocate one sentence to explain how they interpret parsimony in the present context.

Obviously, for any set of empirical results, we should seek the simplest possible theoretical account. But is a single-system model always more parsimonious than a dual-systems model? SMH seem to believe so, even in cases in which the single-system model has the same number of free parameters as the dual-systems model. For example, they write that "in category learning a two-dimensional state-trace could reflect the dimension-weighting and response-bias parameters in the (GCM) Nosofsky (1986), which would generally be interpreted as a 'single-process' account" (p. 18). So SMH define parsimony according to the number of underlying systems, not the number of underlying parameters. In fact, they explicitly state that in any test of COVIS "the latent variables are the explicit and procedural learning systems" (p. 10). Thus, they have no interest in how many free parameters are varying, but only the number of underlying systems that are presumed.

Unfortunately, the SMH definition of parsimony contradicts the definition of parsimony assumed by their chosen method of data analysis — namely, state-trace analysis. Bamber (1979) made it clear that state-trace analysis was designed only to identify cases where more than one free parameter is varying. As such, state-trace analysis knows nothing about the architecture of the underlying models. So at best, all one can conclude about a one-dimensional state-trace plot is that such data are not enough to rule out the possibility that only one free parameter is varying. But that single free parameter could be varying in a model that postulates one system, two systems, or 10 systems. According to state-trace analysis, all these accounts are equally parsimonious.

Parsimony is a well-established and highly-valued principle within the statistical field of model selection. It is universally recognized that all else being equal, model selection should favor the more parsimonious model. Parsimony is operationalized in the AIC and BIC goodness-of-fit statistics as the number of free parameters. In these methods of model selection, models pay a penalty that increases with the number of free parameters, and a model with more free parameters is selected only if its absolute fit exceeds its competitors by more than the extra penalty it pays. Neither AIC nor BIC assigns any weight to the architecture of the models. Single- and dual-systems models with the same number of free parameters pay the same penalty and therefore are considered equally parsimonious.

More sophisticated model selection statistics, such as negative free energy and minimum description length, penalize for mathematical flexibility as well as the number of free parameters. Thus, among models with the same number of free parameters, the model that is mathematically most rigid is considered most parsimonious. However, there is no reason to assume that a single-system model is more mathematically rigid than a dual-systems model with the same number of free parameters. It is easy to construct counterexamples in which the single-system model is more flexible. Therefore, in the statistical field of model selection, the number of systems the model postulates is irrelevant to parsimony.

Thus, SMH endorse a unique definition of parsimony that is inconsistent with state-trace analysis and with all current statistical methods for model selection. The SMH parsimony criterion seems to suggest that if we encounter response-time data that are fit equally well by serial and parallel models, then we should favor the serial model because it postulates that only one system is active at any one time. And because of its presumed greater parsimony, SMH seem to predict that a bias toward singleness should be evident in nature. However, such a bias is not obvious. For example, why do we have multiple sensory systems? And even within the single modality of vision, why do we have separate photopic and scotopic systems, or dorsal and ventral streams?

Unfortunately, SMH conflate systems and free parameters. State-trace analysis is and always was designed to provide information about the number of underlying free parameters. It was never proposed as a method that could say anything about the architecture that produced the state-trace plot, nor are there any conditions in which it could ever be used in this way.

Not all applications of state-trace analysis define parsimony in terms of the number of underlying systems. For example, Dunn et al. (2014) acknowledged that identifying the number of systems "is an impossible task for any statistical procedure or inferential logic, because the concept of a 'system' is itself not well defined" (p. 952). Even so, the question of how many systems a model assumes is critical to SMH because their idiosyncratic definition of parsimony depends on the ability to identify the number of underlying systems. But whether one looks at the architecture proposed by the GCM, COVIS, or any other model and sees 1, 2, or 10 different systems is irrelevant to every point made in this article. First, how one defines a system is irrelevant to state-trace analysis. The original pioneering article in which Bamber (1979) developed state-trace analysis never even mentions underlying systems. In contrast, it mentions parameters many times because the method was developed to provide information about the number of underlying free parameters, not the number of underlying systems. Second, the definition of system is irrelevant to the statistical literature on model selection, and therefore to all current statistical definitions of parsimony. In the literature on model selection, parsimony depends on the number of free parameters in a model, and sometimes also on the model's mathematical flexibility, but never on the number of underlying systems that define the model.

I typically describe COVIS as assuming two systems, but I personally have no interest in the semantics of how one defines a "system". There is no definition of "system" that would change any COVIS predictions, and therefore how one defines a "system" has no empirical consequence. So whereas the definition of a system is of great concern to SMH, it is of no concern whatsoever to state-trace analysis or to any accepted definitions of parsimony. More importantly, however, a debate about how the word "system" should be defined could never improve our understanding of human category learning.

Although SMH never define parsimony, for the remainder of this article, I will adopt the operational definition of parsimony that forms the basis of state-trace analysis, and the model selection statistics AIC and BIC — that is, that the most parsimonious model is the one with the fewest free parameters.

Even a cursory reading of the articles cited by SMH reveals that for the state-trace plots they constructed, the dual-systems model COVIS is more parsimonious than the single-system GCM favored by SMH. This is because many of the dissociations that SMH conclude are nondiagnostic were predicted by COVIS with zero free parameters. For example, the Ashby and Valentin (2017) article cited by SMH clearly states that many of the dissociations are "parameter-free a priori predictions" (p. 158) of COVIS. First, COVIS predicts that a feedback delay must interfere with II learning more than with RB learning, whereas the GCM makes no predictions about how feedback delay might affect either task.[1] Second, COVIS predicts that a dual-task that recruits working memory must interfere with RB learning more than II learning, whereas the GCM makes no predictions about how a dual task might affect either task. Third, COVIS predicts that switching the locations of the response buttons after initial learning must interfere with II performance more than RB performance, whereas the GCM makes no predictions about how a button switch might affect either task. Fourth, COVIS predicts that limiting time and attention for feedback processing must interfere with RB learning more than II learning, whereas the GCM makes no predictions about how feedback processing might affect either task. Fifth, COVIS predicts that unsupervised II learning must be worse than unsupervised RB learning, whereas the GCM makes no predictions that removing feedback should differentially affect either task. Sixth, COVIS predicts that analogical transfer must be worse in II tasks than in RB tasks, whereas the GCM makes no predictions about the differential success of analogical transfer in RB and II tasks. All of these COVIS predictions have been empirically supported in replicated experiments (for a review, see Ashby & Valentin, 2017).

This list could be extended. Currently, there are somewhere around 27 different dissociations predicted by COVIS that have been empirically confirmed (for a review of most of these, see Ashby & Valentin, 2017). The GCM may or may not be able to account for each of these dissociations *post hoc* by manipulating one or more free parameters (in most cases, this is unknown). But even if it could, the choice is between an account that predicted these results MUST occur (i.e., with zero free parameters) versus an account that made no *a priori* predictions and was only able to fit the state-trace curves *post hoc* by manipulating one or more free parameters. Which account is more parsimonious?

Furthermore, by arguing in favor of the GCM over COVIS solely on the basis of a misguided notion of parsimony, SMH also ignore an enormous amount of other relevant research. First, Ashby and Rosedahl (2017) showed that the GCM is mathematically equivalent to a special case of the procedural system of COVIS. In addition to the procedural system, COVIS also postulates an explicit rule-learning system. So when SMH hypothesize that all categorization behavior is consistent with the GCM, then they are assuming that all categorization behavior is procedural or similarity-based and therefore that humans have no special rule-learning abilities. But a rule is a set of necessary and sufficient conditions, and these conditions do not have to be similarity-based. For example, slightly rounding the corners of one of two identical squares has almost no effect on perceptual similarity, but moves the rounded square out of the rectangle category. So the SMH hypothesis that the GCM is sufficient to account for all categorization behavior is sharply inconsistent with the large literature on human rule-guided behavior (e.g., Bunge & Wallis, 2008).

Second, COVIS hypothesizes that optimal performance in RB tasks is mediated by a broad neural network that includes the prefrontal cortex, the anterior cingulate, and the hippocampus, whereas optimal performance in II tasks depends primarily on the basal ganglia, and especially the striatum. In contrast, except for Ashby and Rosedahl (2017), the GCM has no neuroscience underpinnings. Thus, rejecting COVIS in favor of the GCM on the basis of parsimony ignores results of the animal lesion studies, animal neurophysiology studies, human neuropsychology studies, and human functional neuroimaging studies that support the neuroscience predictions of COVIS.

4. Two-dimensional state-trace plots

As mentioned earlier, SMH interpret a two-dimensional state-trace plot to mean that "the data are inconsistent with any model based on a single latent variable" (p. 6). As an example of a model that postulates a single latent variable they cite applications of the GCM in which only the dimension-weighting parameter varies. Fig. 1, which is adapted from Ashby (2014), shows four different state-trace plots predicted by the GCM in applications to RB and II experiments where only one parameter is varying. In panel A, only the GCM overall attention parameter c varies. In panels (B), (C), and (D) only the dimension-weighting parameter w varies. Note that the state-trace plots in panels (B), (C), and (D) are all two-dimensional, which clearly shows that SMH misinterpreted state-trace analysis. The exact same single-latent variable model they recommend clearly falsifies their claim that two-dimensional plots "are inconsistent with any model based on a single latent variable" (p. 6). When only a single parameter varies, the GCM can produce any type of state-trace plot. Thus, when state-trace analysis is applied to data from RB and II categorization experiments, no inferences about the number of underlying free parameters is possible — regardless of whether the state-trace plot is one- or two-dimensional.

But perhaps state-trace analysis is better at identifying the number of underlying systems (i.e., when we accept the SMH definition of a system). Ashby (2014) also examined this issue. Fig. 2, which is also adapted from Ashby (2014), shows state-trace plots produced by the single-system GCM (left column), and by a simplified version of the dual-systems model COVIS (right column) in a hypothetical experiment in which RB and II categories are learned under single-task and dual-task conditions. Note that both models unambiguously predict either one- or two-dimensional state-trace plots. So neither empirical finding provides any information about the number of underlying systems.

Obviously, Figs. 1 and 2 are highly relevant to the entire premise of the SMH article, especially since all state-trace plots shown in these figures are from the same RB and II tasks used in the 63 studies re-analyzed by SMH. Even so, SMH chose not to mention (or cite) the Ashby (2014) article — despite strong evidence that they knew of its existence. For example, they cite a comment on Ashby (2014), which was written by some of the editors of the special issue in which SMH appeared. So if they read the articles they cited, then they knew of these figures.

The key issue is monotonicity. Specifically, a model with one freely varying parameter must predict a one-dimensional state-trace plot only if performance on both tasks increases (or decreases) monotonically with increases in that parameter (see e.g., Ashby, 2014, for a proof). Ashby (2014) showed that if this assumption is invalid then a model with one freely varying parameter can predict any state-trace plot (e.g., see Fig. 1). Although SMH acknowledge that their analyses depend on this monotonicity assumption, they never discuss its possible validity.

[1] Of course, there are boundary conditions for all of these predicted dissociations. For example, a feedback delay of one year would likely abolish all learning, both in RB and II tasks. So technically, in each case these predictions are weak orders, rather than strict orders.

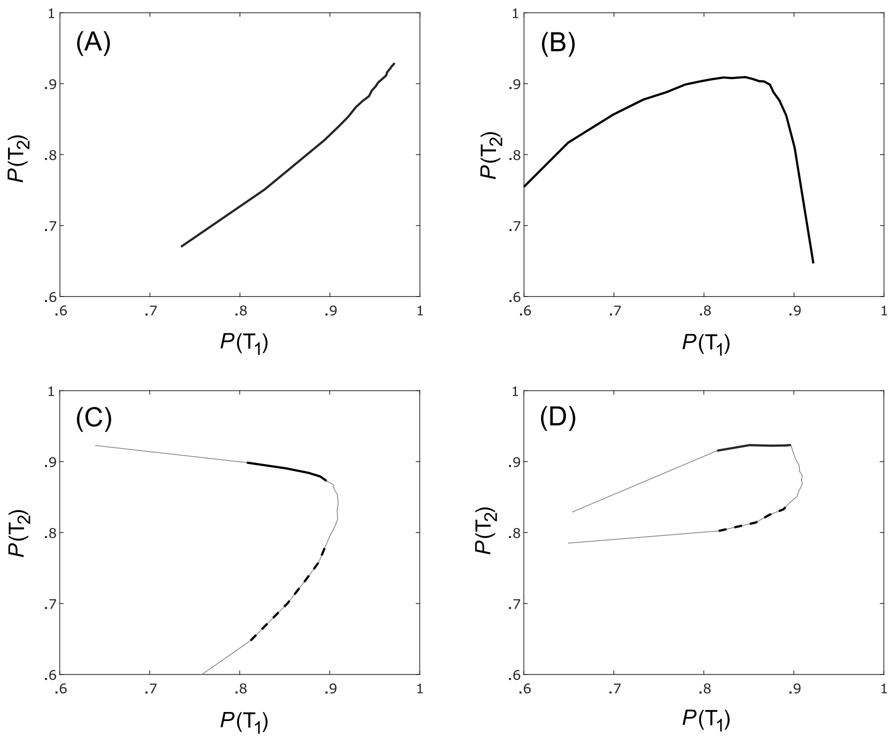

Fig. 1. Some state-trace plots produced by the GCM in applications to RB and II experiments in which only one parameter varies across tasks T_1 and T_2. $P(T_1)$ and $P(T_2)$ are both categorization accuracy. (A) Only the GCM overall attention parameter c varies. (B), (C), and (D) Only the dimension-weighting parameter w varies.
Source: Adapted from Ashby, F.G., 2014. Is state-trace analysis an appropriate tool for assessing the number of cognitive systems? *Psychonomic Bulletin & Review*, *21*, 935–946.

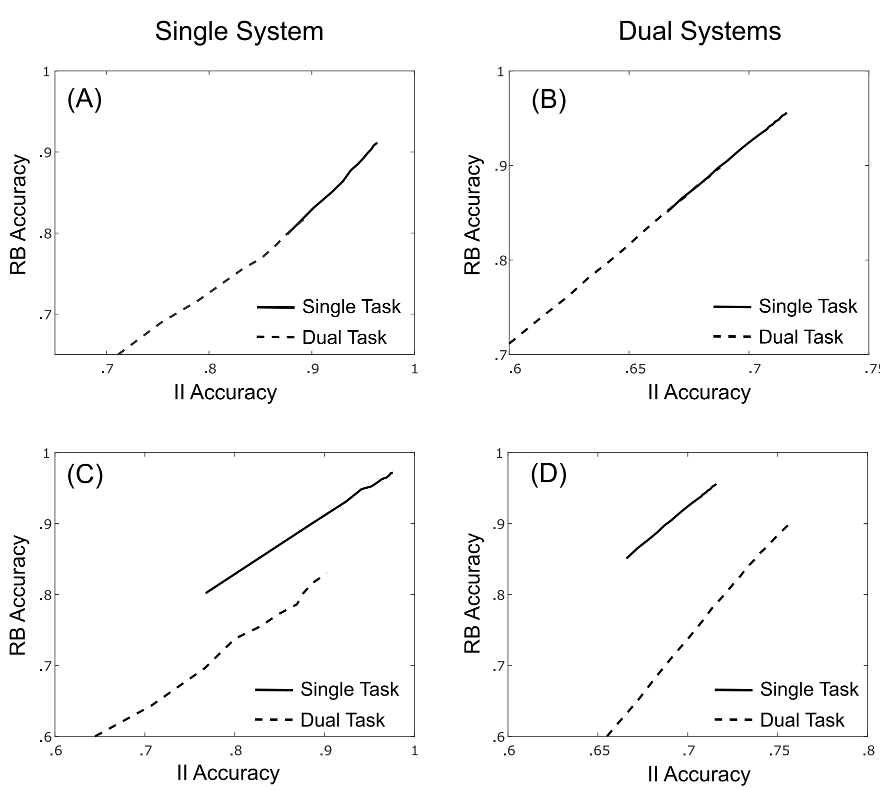

Fig. 2. State-trace plots from the GCM (left column) and a simplified version of COVIS (right column) from RB and II experiments under single-task and dual-task conditions. (A) Three parameters vary across tasks and conditions. (B) Only one parameter varies across tasks and conditions. (C) & (D) Two parameters vary across tasks and conditions.
Source: Adapted from Ashby, F.G., 2014. Is state-trace analysis an appropriate tool for assessing the number of cognitive systems? *Psychonomic Bulletin & Review*, *21*, 935–946.

There surely are empirical domains in which it is reasonable to assume that performance is monotonic with all possible parameters that might be varying across tasks and conditions. As described by Bamber (1979), state-trace analysis can make useful contributions to any such field. However, category learning is most clearly NOT such a domain because in virtually all popular categorization models, performance changes nonmonotonically with one or more parameters. In fact, I don't know of a single popular model for which this is not true.

For example, consider the SMH recommendation that "instead of two learning systems, a single latent variable such as ... (the) dimension-weighting parameter as included in the (GCM) Nosofsky (1986) – would often be sufficient to account for the results" (p. 14). The optimal value of the GCM dimension-weighting parameter (i.e., w) in the II task is $w = .5$ (equal attention allocated to both dimensions). Thus, as w increases from 0 to 1, the GCM predicts that II accuracy will increase until $w = .5$ and then decrease as w continues to rise. So, as is evident in Fig. 1, varying the single dimension-weighting parameter in the single-system GCM produces two-dimensional state-trace plots. Thus, despite the claims of SMH, the dimension-weighting parameter of the GCM is NOT sufficient to account for the many one-dimensional state-trace plots they report.

Virtually all popular categorization models include parameters that violate the SMH monotonicity assumption. First, many categorization models include a dimension-weighting parameter that operates identically as in the GCM. Included in this list are exemplar models, prototype models, and decision-bound models (e.g., see Ashby & Maddox, 1993). Second, decision bound models predict that accuracy is maximized in any categorization task when the decision bound has some specific intermediate intercept and curvature. Thus, decision bound models predict that accuracy will increase to some peak value and then decrease as the intercept increases from $-\infty$ to ∞. A similar prediction occurs for the amount the decision bound curves (e.g., from negative to positive). Third, the COVIS rule-learning model predicts that accuracy will increase with the parameter that measures dopamine levels and then decrease when these levels pass an optimal value. Fourth, all connectionist and neural network models predict that for any given amount of training, accuracy increases with the value of the learning-rate parameter up to a point, and then performance will deteriorate if the learning rate becomes too large. Thus, if any of these parameters vary across tasks or conditions, then the resulting state-trace plot should be two dimensional — even if that parameter is the only one that varies.

So to rule out the possibility that a single underlying parameter is varying from a two-dimensional state-trace plot, it is necessary to assume that no parameters of this type are varying across any of the experimental conditions or tasks. This is highly unlikely in the categorization studies re-analyzed by SMH. This is because the two tasks in these studies are RB and II categorization tasks. In many of the RB tasks, only one stimulus dimension is relevant, whereas two dimensions are equally relevant in the II task. Thus, any model with a dimension-weighting parameter will predict changes in this parameter across tasks, and as mentioned, in virtually all models with such a parameter, performance is nonmonotonic with the dimension weight. For this reason alone, applying state-trace analysis to data from RB and II categorization tasks is a misapplication of the method.

5. Conclusions

COVIS has been an extremely successful theory because it has encouraged many new studies that otherwise would not have been run. This includes virtually all of the 63 categorization studies examined by SMH. For example, no prior theory predicted that delaying feedback could have different effects for different types of category structures, so without COVIS, it is likely that we would not now know that feedback delays impair II learning more than simple RB learning. Even so, it is important to acknowledge that COVIS can be improved — as is true of all theories in psychology. Of course we know that COVIS is incomplete because it leaves out an enormous amount of detail, but it is likely wrong in more fundamental ways. Certainly the theory has evolved since it was first proposed 20 years ago, and this process is likely to continue.

Theories are tools to motivate new research, which will inevitably point to flaws in the current theories, and hopefully lead to new and better theories, at which point the cycle can begin again. This is the scientific process and the reason that I welcome alternative theories to COVIS. The lack of serious alternatives has slowed progress in the field. Unfortunately, however, SMH do not offer any alternative theories of category learning. Although they argue for a single learning system, and make off-hand suggestions that the GCM might be the best model, they fail to show that the GCM or any other single-system model can account for even one of the 63 studies they examine. Furthermore, neither do they identify any weaknesses or mispredictions of COVIS. Instead, their only empirical contribution is to point out that none of the 63 studies they examined conclusively identify the number of underlying free parameters — a point that was never in contention and that was made explicitly in previous publications. Thus, SMH make no new empirical contribution.

Even worse, however, is that their theoretical recommendations could actually harm the field. Specifically, they argue that (1) psychology should adopt a definition of parsimony that contradicts state-trace analysis and all definitions from the statistical field of model selection, and (2) no new theories should be endorsed until a single experiment is produced that definitively rules out all other accounts. In my opinion, adopting these proposals would effectively halt all scientific progress in psychology.

Acknowledgment

This work was supported by the National Institutes of Health (grant 2R01MH063760).

References

Ashby, F. G. (2014). Is state-trace analysis an appropriate tool for assessing the number of cognitive systems? *Psychonomic Bulletin & Review*, *21*, 935–946.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372–400.

Ashby, F. G., & Rosedahl, L. (2017). A neural interpretation of exemplar theory. *Psychological Review*, *124*, 472–482.

Ashby, F. G., & Valentin, V. V. (2017). Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen, & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (2nd ed.). (pp. 157–188). Amsterdam: Elsevier.

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology*, *19*, 137–181.

Bunge, S. A., & Wallis, J. D. (2008). *Neuroscience of rule-guided behavior*. Oxford University Press: New York.

Dunn, J. C., Kalish, M. L., & Newell, B. R. (2014). State-trace analysis can be an appropriate tool for assessing the number of cognitive systems: A reply to Ashby (2014). *Psychonomic Bulletin & Review*, *21*, 947–954.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*, 91–101.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Stephens, R. G., Matzke, D., & Hayes, B. K. (2019). Disappearing dissociations in experimental psychology: Using state-trace analysis to test for multiple processes. *Journal of Mathematical Psychology*, *90*, 3–22.

Teuber, H.-L. (1955). Physiological psychology. *Annual Review of Psychology*, *6*, 267–296.